UNIVERSITY OF KWAZULU-NATAL

# INVESTIGATION OF FERTILIZER USAGE IN MALAWI WITHIN THE RURAL LIVELIHOOD DIVERSIFICATION PROJECT USING GENERALIZED LINEAR MODELS AND QUANTILE REGRESSION

2013

HILDA JANET JINAZALI KABULI

# INVESTIGATION OF FERTILIZER USAGE IN MALAWI WITHIN THE RURAL LIVELIHOOD DIVERSIFICATION PROJECT USING GENERALIZED LINEAR MODELS AND QUANTILE REGRESSION

By

HILDA JANET JINAZALI KABULI

Submitted in fulfilment of the academic requirements for the degree of

MASTER OF SCIENCE
in
STATISTICS

in the
School of Statistics and Actuarial Sciences
University of KwaZulu-Natal
Pietermaritzburg

2013

# Dedication

To my beloved Husband Amon Kabuli, my Mum Violet Jinazali, and in memory of my Dad Ralph Jinazali.

# Declaration

The research work described in this thesis was conducted in the School of Statistics and Actuarial Sciences, University of KwaZulu-Natal, Pietermaritzburg by the author, under the supervision of Prof. Temesgen Zewotir and Prof. Principal Ndlovu.

I Hilda Janet Jinazali Kabuli, declare that this thesis is my own work, and has never been submitted in any form of any degree or diploma at any University. Duly acknowledgement is made where the work of others is used.

Mrs Hilda Janet Jinazali Kabuli

Date

Prof. Temesgen Zewotir

Date

# Acknowledgement

I would like to thank my supervisors Prof. Temesgen Zewotir and Prof. Principal Ndlovu for their tireless effort to make this thesis a success.

I would like to express my heartfelt gratitude to the International Center for Tropical Agriculture (CIAT) for financial support, CIAT staff especially Dr Jemimah Njuki, Dr Rob Delve, Dr Robin Buruchara and Dr Shamie Zingore for their technical support. I also extend my gratitude to my Director Dr Alfred Mtukuso and staff of the Department of Agricultural Research Services for their usual support.

My special thanks go to my one and only husband, Amon Kabuli for the love, encouragement and support whilst I was pursuing my studies. To my sister in-law Ziliro Makwakwa and my sister Chisomo Jinazali, I say '*zikomo!*' for assuming a motherly role in caring for my daughter Eunice and son Jones Joshua 'JJ' whilst I was away pursuing my studies. To my parents, Violet Jinazali and late Dad Ralph Jinazali – I say you are wonderful parents, as you are my source of strength and confidence through your words of wisdom "Education is a passport for a better future!". To my brothers Desmond, Daniel, and Alfred, and friends I say – "My achievement, is your achievement!' God bless!

# Abstract

Malawi’s economy relies heavily on agriculture which is threatened by declines in soil fertility. Measures to ensure increased crop productivity at household level include the increased use of inorganic fertilizers. To supplement the Government’s effort in ensuring food security, Rural Livelihood Diversification Project (RLDP) was implemented in Kasungu and Lilongwe Districts in Malawi. The RLDP Project was aimed at increasing accessibility and utilisation of inorganic fertilizers. We used the data collected by the International Center for Tropical Agriculture (CIAT), to investigate if there could be any significant impacts of the interventions carried out by the project. A general linear model was initially used to model the data. Terms in the model were selected using the automatic stepwise procedure in GLMSELECT procedure of SAS. Other models that were used included a transformed response general linear model, gamma model based on log link and its alternative inverse link, and quantile regression procedures were used in modelling the amount of fertilizer use per acre response given a set of fixed effect predictors where households were only sampled at baseline or impact assessment study. The general linear model failed to comply with the model assumption of normality and constant variance. The gamma model was affected by influential observations. Quantile regression model is robust to outliers and influential observations. Quantile regression provided that number of plots cultivated, timeline, household saving and irrigation interaction, and the interaction between plots and timeline significantly affected the amounts of fertilizers applied per acre amongst the 25% of the households who apply lower levels of fertilizer per acre.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Agriculture is the source of income as well as food security in most of the Sub-Saharan African countries. In Malawi it contributes about 36% of the value added gross domestic product (GDP) and 80% of foreign exchange earnings; and employs around 85% of the workforce. High levels of agricultural productivity ensure increased levels of household food security and income. On the other hand low and/or stagnant levels of agricultural productivity result in food insecurity and increased levels of poverty (Government of Malawi, 2006a and 2006b; Bationo, 2009).

Most of the African countries have been experiencing declines in agricultural productivity of which soil infertility is one of the major causes (Denning et al., 2009; Morris et al., 2007). Many studies (Denning et al., 2009; Bationo, 2009), have reported that there has been a high loss of soil fertility in most parts of Africa due to: natural resource degradation, nutrient mining by crops, increasing population density which has resulted in increased demands for land and over usage of available land, and continuous growing of maize on the same piece of land. Some of the solutions to overcome the problem of soil nutrient depletion are the use of good land management practices, application of inorganic fertilizer. Inorganic fertilizer is essential in maintaining and enhancing soil fertility for increased crop production per unit area, and also increases crop residue production  which in turn increases organic matter which improves soil nutrient capacity and structure (Heisey and Mwangi, 1996; and Kumwenda et al., 1996). Several research studies (Government of Malawi, 2007; Minde et al., 2008) have reported the application of fertilizer below the recommended rates by smallholder farmers leads to lower yields compared to potential world averages resulting into low crop production, food insecurity and increased poverty. The common barriers that lead to low uptake of fertilizers include lack of cash or credit to access fertilizers (Kumwenda et al., 1996; and Chirwa 2003). Kumwenda et al (1996) recommended ways of enhancing farmers’ adoption of improved soil fertility technologies including access to credit, integrating organic and inorganic technologies through increased use of animal manure, intercropping, legume rotation, agroforetsry, and inorganic fertilizer use, in order to increase maize crop productivity.

As one of the initiatives to end hunger in Africa, the Rural Livelihood Diversification Project (RLDP) which was funded by the United States Agency for International Development (USAID), was implemented in Malawi, Mozambique and Zambia from October 2006. The RLDP project activities included promotion and training of private sector agro-dealer networks to commercialise agricultural inputs such as fertilizer in order to improve accessibility of fertilizer to smallholder farmers; increasing awareness of the importance of fertilizer usage through on-farm demonstrations and farmer participatory research.

In line with the RLDP goal, there are several initiatives that are being conducted in Africa in order to ensure increased use of fertilizer by the smallholder farmers. The Government of Malawi is implementing the farm input subsidy programme (FISP), which started in 2005/2006 season,  aimed at increasing agricultural productivity thereby ensuring availability of enough food and increased income at household and national levels (FAO, 2009; and Government of Malawi, 2006b). In Mozambique and Zambia there are also programmes aimed at subsidising the price of fertilizers. In Zambia the fertilizer support programme (FSP) started in 2002 (Bationo, 2009; and Minde et.al., 2008). In Malawi, Mozambique and Zambia, where the Rural Livelihood Diversification Project was implemented, several nongovernmental organisations as well as international organisations also support some targeted households with agricultural inputs including fertilizer.

There were several RLDP intervention activities that were carried out in order to achieve the RLDP goals; these activities included training of agro-dealears, conducting on-farm demonstrations and farmer participatory research through on-farm experiments. Agro-dealers were trained in fertilizer product input application recommendations; safety and use of different inorganic fertilizers so as to enhance their skills in handling fertilizer and better save the communities. Demonstrations were conducted in farmers' fields with the aim of training farmers and increasing awareness in the use of improved fertilizer technologies which include improved methods of fertilizer application on given specific crops and environment. In line with demonstrations, participatory on-farm experiments were done in order to enhance the skills and capacity of the farmers in handling and use of recommended application methods of fertilizer. Participatory monitoring and evaluation (PM&E), was conducted by implementing partners throughout the project implementation in order to track changes of the project interventions.

The main objective of the thesis is to investigate the factors that affect the amount of fertilizer usage per unit area for maize production in Lilongwe and Kasungu Districts of Malawi. Based on the research survey design, method of data collection, and the quality of the data that have been collected, the following are the specific objectives of the thesis

1. To investigate the impact of RLDP on fertilizer usage for maize production.
2. To investigate the factors affecting fertilizer usage in maize production.

This research study is important as it would generate results that could be of use to the researchers, International Center for Tropical Agriculture, policy makers and other stakeholders on areas that require attention in order to design impact oriented intervention strategies that could aid in the improvement on the levels of fertilizer that smallholder farmers apply on their fields. Based on practical and analytical results from this study, the identified factors affecting the amount of fertilizer use per acre would provide a platform for researchers, government, and NGOs in making informed decisions on promoting and scaling up factors that have a significant positive impact on the amount of fertilizer use per acre. Moreover, the study is crucial for effective policy support in promotion and increased investment on inorganic fertilizer in maize. This in turn would enhance the increased utilization of fertilizer by the smallholder farmers in Malawi, which could translate into growth in agricultural productivity, food security and improved rural livelihoods.

This thesis is organised into five chapters. Chapter 2 describes the data that have been used in the modelling of the amount of fertilizer use per acre, and clarification of the categories of the explanatory variables. Chapter 3 assesses the factors that affect fertilizer usage which are analysed based on mean amount of fertilizer per acre using the general linear model, transformed response general linear model, gamma model using the log link and its alternative inverse link, and assessment of the amount of fertilizer use per acre using the quantile regression procedure. Chapter 4 presents the results and discussion of our findings. Chapter 5 presents the conclusions of the thesis.

# Chapter 2

# The Data

We used the data that were collected by the International Center for Tropical Agriculture (CIAT) in Malawi. The research in Malawi was conducted in Lilongwe and Kasungu districts where the Rural Livelihood Diversification Project was implemented in Malawi. Lilongwe and Kasungu Districts lie in the Lilongwe-Kasungu plain, which is the main agricultural area in Malawi and is characterized by red-yellow soils (latosols). Lilongwe is located at $13^{0}59^{'}$ South and $33^{0}47^{'}$ East. Kasungu is located $13^{0}2^{'}$ South, and $33^{0}29^{'}$ East.

A two stage cluster sampling technique (Scheaffer et. al., 1986) was used to select households for the baseline research study. In the first stage, community group villages were selected from the selected two districts in Malawi, Lilongwe and Kasungu districts, where the project was implemented. Ukwe group village was selected from Lilongwe district with a total of 16 small villages. Suza and Kalikwembe group villages were selected from Kasungu district with about 38 small villages. In the second stage, a simple random sample of households was selected from the selected community group villages. A total of 357 households were sampled for the baseline survey which was conducted in June 2007, of which 143 households were from the Lilongwe and 214 households were from the Kasungu district.

Impact assessment was conducted as a follow up study in July 2008 with the aim of assessing whether the project had achieved its overall goal and specific objectives. Interviews were based on a simple random sample of households from the villages which were sampled in the baseline study. A total number of 198 households were randomly selected, and 99 households were from each district some of whom were also interviewed during the baseline study. In both baseline and impact assessment studies the same structured questionnaire as given in Appendix C, was used in collecting data on: the household characteristics, access to fertilizer, amount of fertilizer usage, land availability, fertilizer and fertility perceptions. Enumerators who were trained in the objectives of the project and data collection techniques were employed to collect the data.

Data has been checked in order to identify invalid or incorrect data values, extreme values, and missing values which could have resulted from improper recording of the data on the questionnaire as well as typing errors when entering data in a computer. The identified errors were assessed and corrected, so as to improve the quality of the data and reduce bias of the results from the analysis.

## 2.1 Definition and measurement of variables

The dependent variable in the statistical models used was the amount of fertilizer use per acre. Amount fertilizer use per acre was computed from the total amount of fertilizer applied divided by the total land size (in acres) where the fertilizer was applied under maize crop in a given household. In this research work the sampling unit is the household with own farm which was selected from a population of households. In this thesis a household is defined as members of the same family who might be related or not related, who live together in one house or several houses, do agricultural operations together on the same garden and eat together the food prepared from the same pot. The explanatory variables are described below.

**District of farming household:** It is being hypothesised that the Lilongwe and Kasungu households' abilities to meet their basic needs as well as their abilities to purchase fertilizer for improving crop production are different as a result of the differences in the wealth of the two districts (Government of Malawi, 2005b). The district of the farming household has the dummy variable with the value of 1 for Lilongwe district, and 2 for Kasungu district. Kasungu district is the reference category.

**Household head:** The assumption is that households headed by women or children have a higher likelihood of using less fertilizer per unit area than those headed by men. This is because women generally earn less than men, and this is likely to cause lower fertilizer purchases amongst the female headed households than amongst male headed households (Smale and Phiri, 1998). Household head is represented by the dummy variable with 1 categorising male headed household and 2 otherwise.

**Capacity building through training:** According to Fufa and Hassan (2006), knowledge capacity building is essential in ensuring empowerment of farmers in making critical decisions in the allocation of limited resources thereby improving the levels in which farmers realise the agricultural benefits. Therefore it is being hypothesised that farmers who received some training in the use of fertilizer have a higher likelihood of using more fertilizer per unit area than those without training. Capacity building is reflected by the dummy variable with value of 1 if the household received some training on fertilizer use and 2 otherwise as the reference.

**Source of fertilizer:** Inorganic fertilizers are costly thereby unaffordable to most smallholder farmers (Kumwenda et al., 1996). Therefore in this study it is being hypothesised that households with access to cost subsidised fertilizers would apply more fertilizer per unit area. Source of fertilizer is captured by the dummy variable that takes the value of 1 for purchased fertilizer, and 2 otherwise as the reference category.

**Land size:** The expectation is that the amount of fertilizer applied per acre will be correlated with land size. Households with large land sizes produce more crops for sale which enables them to purchase more fertilizer to apply per unit area of land (Chirwa, 2003).

**Saving:** Due to uncertainty of price changes of commodities (Government of Malawi, 2005a), it is expected that households with enough savings are able to purchase more fertilizer, even when the price of fertilizer increases. Saving is captured by the dummy variable with the value of 1 if the household saves money, and the reference category value of 2 if the household does not save money.

**Irrigation:** It is being hypothesised that households who irrigate their fields have more income, through sales of irrigated crops thereby strengthening households economically to purchase fertilizer (Mangisoni, 2006). Irrigation is distinguished by the dummy variable that takes the value of 1 if the household practices irrigation, and value of 2 otherwise as the reference.

**Animal manure:** Availability of animal manure in large quantities and of appropriate quality could improve soil fertility and hence crop production (Mafongoya et al., 2006). Because of low quantities of animal manure available to farmers, the assumption is that there could be more demand for inorganic fertilizer and hence more application per unit area in order to improve soil fertility and crop productivity. Households that apply animal manure are represented by the dummy category of 1, and those that do not apply animal manure are represented by category 2.

**Number of plots cultivated:** In Malawi, smallholder farmers tend to have several spatially fragmented plots of land for cultivation of crops (Chirwa, 2003; Government of Malawi, 2002). It is being hypothesised that if the household has more plots and insufficient resources such as fertilizer, the resources are more likely to be distributed inefficiently amongst the plots.

**Uncultivated plots:** It is likely that households who leave part of their gardens uncultivated (Government of Malawi, 2002), are those who have problems in accessing fertilizer. Therefore households who leave some of their plots uncultivated are likely to apply less fertilizer per unit area of land. Uncultivated plots is characterised by the dummy variable that takes the value of 1 if the household leaves some land uncultivated, and value of 2 otherwise as the reference.

**Distance to the source of fertilizer:** It is being hypothesised that if the source of fertilizer is far away from the households, more money is spent on transport thereby affecting the quantities of fertilizer which are bought hence less applications per unit area applied (Kherallah et al., 2002).

**Number of months with enough food:** With increased use of fertilizer per unit area, there could be  increased maize yields thereby achieving adequate amounts of food and of good quality throughout the year (Heisey and Smale,1995). Hence, it is expected that households that have enough food of their own production at most times during the year would have more significant fertilizer application rates per unit area of land.

**Total income per annum:** The assumption is that households with more income per annum could apply more fertilizer per unit area (Minot et. al., 2000).

**Perceptions of impact of fertilizer on the soil:** It is being hypothesised that where fertilizer is perceived to have a positive impact on the soils as well as crop productivity, more fertilizer is applied (Government of Malawi, 2007). A dummy variable of perceptions of fertilizer is represented by a value of 1 if the fertilizers are perceived to be good for the soil and the value of 2 otherwise as the reference category.

**Perceptions on fertility of soil:** The assumption is that, with a reduction of soil fertility the demand for inorganic fertilizer could be higher per unit area (Heisey and Smale, 1995). Perception on soil fertility is reflected by the dummy variable with value of 1 if the household perceive the soil to be fertile and 2 otherwise as the reference.

**Timeline:** It is being hypothesised that more fertilizer will be applied during the impact assessment survey than during the baseline survey due to improvements and scaling up of fertilizer use interventions (Government of Malawi, 2007) by the RLDP Project, the Government and other stakeholders. Timeline is represented by the dummy variable with value of 1 if the household was interviewed at baseline and 2 at impact assessment.

Since some of the above qualitative explanatory variables had more than two levels, the qualitative variables were further categorised in order to (i) ease handling and analysis of the data; (ii) investigate how the fertilizer usage per acre would be at several levels of the explanatory variables; (iii) facilitate comparison amongst classes (Draper and Smith, 1981). Consideration on cut off points for each class was subjectively defined based on adequate representation (at least 20%) of the frequencies.

Since some of the above explanatory variables are quantitative whilst others are qualitative, Table 2.1 displays the categories of the qualitative explanatory variables and Table 2.2 displays the frequency distribution of the categorized explanatory variables excluding households interviewed at both baseline and impact assessment studies.

Table 2.1: Definition and abbreviations of key explanatory variables

| **Explanatory variable** | **Abbreviation in the thesis** | **Definition** |
|---|---|---|
| District of study | District | Dummy 1 if Lilongwe<br>Dummy 2 if Kasungu |
| Household head | Head | Dummy 1 if adult Male<br>Dummy 2 otherwise |
| Capacity building (if ever received training regarding fertilizers) | Training | Dummy 1 if yes<br>Dummy 2 if no |
| Source of fertilizer | Source | Dummy 1 if purchased<br>Dummy 2 if government subsidised cost/or nongovernmental organisations |
| Land size | Land | Land size cultivated (in acres) |
| Savings | Saving | Dummy 1 If save money<br>Dummy 2 If does not save money |
| Irrigation | Irrigation | Dummy 1 If practice irrigation<br>Dummy 2 If does not practice irrigation |
| Use of animal manure | Anlmanure | Dummy 1 if use animal manure<br>Dummy 2 If does not use animal manure |
| Number of plots cultivated | Plots | Total number of plots cultivated |
| Some land uncultivated | Nogrow | Dummy 1 If left some land uncultivated<br>Dummy 2 If did not leave some land idle |
| Distance to source of fertilizer | Distance | Total distance travelled to source fertilizer (in kilometres) |
| Months the household has enough food | Lenglast | Number of months the household has adequate food of its own production |
| Total annual income | Total income | Total amount of money (in Malawi Kwacha) the household has per annum |
| Fertilizer impacts perceptions | Fertgood | Dummy 1 If good for the soil<br>Dummy 2 If not good for the soils |
| Perceptions on soil fertility | Fert_perc | Dummy 1 If medium to high fertile<br>Dummy 2 If low fertile |
| Timeline | Timeline | Dummy 1 if baseline<br>Dummy 2 if impact assessment |

Table 2.2 shows that 55.30% of the households who were interviewed came from the Kasungu district. Most of the households were headed by males (51.52%). The majority of the households were not formally trained on fertilizer use (66.17%), did not apply animal manure (56.61%) and practiced irrigation (58.59%). Table 2.2 also shows that 57.00% of the households sourced fertilizer through the Government subsidy programme and the nongovernmental organisations. Although most households reported that their soil is of medium to high fertility, perceptions on the impact of fertilizer on the soil was generally not good with 72.26% of the households having negative perceptions.

Table 2.2: Frequency distribution of the explanatory variables

| Variable | Frequency | Percentage (%) |
|---|---|---|
| **District** | | |
| Lilongwe | 177 | 44.70 |
| Kasungu | 219 | 55.30 |
| **Household head** | | |
| Male | 203 | 51.52 |
| Others (Female, child) | 191 | 48.48 |
| **Training** | | |
| Received training | 135 | 33.83 |
| Not trained | 264 | 66.17 |
| **Source of fertilizer** | | |
| Purchased | 169 | 43.00 |
| Subsidy/NGOs | 224 | 57.00 |
| **Savings** | | |
| Save money | 184 | 60.33 |
| Does not save money | 121 | 39.67 |
| **Irrigation** | | |
| Irrigates | 208 | 58.59 |
| Does not irrigate | 147 | 41.41 |

Table 2.2: Frequency distribution of explanatory variables continued

| Variable | Frequency | Percentage (%) |
|---|---|---|
| **Animal manure** | | |
| Use | 128 | 43.39 |
| Does not use | 167 | 56.61 |
| **Some land uncultivated** | | |
| Yes | 96 | 24.74 |
| no | 292 | 75.26 |
| **Fertilizer impact perceptions** | | |
| Good | 109 | 27.74 |
| Not good | 284 | 72.26 |
| **Soil fertility perceptions** | | |
| Medium to high fertile | 210 | 52.90 |
| Low fertile | 187 | 47.10 |
| **Timeline** | | |
| Baseline | 300 | 75.19 |
| Impact assessment | 99 | 24.81 |

# Chapter 3

# Review of statistical methods

## 3.1 Generalized linear models theory

Generalized linear models introduced by Nelder and Wedderburn (1972), is an extension to traditional general linear models, as the distribution of the observations of the response variable may come from the exponential family of distributions. Generalized linear models have been applied in modelling of the mean for both discrete and continuous data. It gives a consistent way of linking together the systematic elements in the model with the random elements (Nelder and Wedderburn, 1972).

In generalized linear models, the response is not necessarily normal, and possess a probability distribution of the exponential family (Pregibon, 1980). The exponential family covers a wide range of distributions including normal (used in linear regression and analysis of variance), Poisson (used in discrete forecasting models and the log-linear model), binomial and multinomial responses (used in analyses involving proportions), as well as gamma and negative binomial distributions (O'Brien, 1983). According to McCullagh and Nelder (1989), the probability density response Y for the continuous response variables, or the probability function for discrete responses can be expressed as

$$f(y;\theta.\emptyset) = exp\left\{\frac{y\theta - b(\theta)}{a(\emptyset)} + c(y,\emptyset)\right\} \tag{3.1}$$

where a(.), b(.) and c(.) are specific functions that determine the specific distribution. The parameter θ is called the natural location parameter and $\emptyset$ is called the scale parameter. The mean and variance of the distribution are given by

$$E(\mathrm{Y}) = \frac{db(\theta)}{d\theta} = b'(\theta) = \mu\ . \tag{3.2}$$

$$Var(\mathrm{Y}) = \frac{d\mu}{d\theta} a(\emptyset) = b''(\theta) a(\emptyset) = \sigma^2. \tag{3.3}$$

In generalized linear models, consideration is on the set of model parameters $\beta_1$,$\beta_2$,..., $\beta_p$ such that a linear combination of the $\beta_p$'s is equal to some function g(.) of the expected value of the responses $\mu_i$ =E($Y_i$ ), i.e. for *i*=1,2,...,n, given by

$$\eta_i = g(\mu_i) = \boldsymbol{x}_i'\beta \tag{3.4}$$

where g(.) is the link function, a term derived from the fact that the function is the link between the mean $\mu_i$ and the linear predictor $\eta_i = \boldsymbol{x}_i'\beta$; $\boldsymbol{x}_i$ is a p × 1 vector of explanatory variables; and β= ($\beta_1$,……., βp)' is the p × 1 vector of parameters.

The choice of the appropriate link functions depends on the specific exponential family, as for each exponential family there is a general natural or canonical link function that relates the linear predictor $\eta_i = \boldsymbol{x}_i'\beta$ to the expected value $\mu_i$. For example, the link function for the normal distribution is the identity.

Having selected a particular distribution and link function of the model, it is required to estimate the model parameters, and to assess the precision of the estimates. In generalized linear models, parameters are estimated using the maximum likelihood method. The maximum likelihood estimator is consistent and unbiased with large samples, asymptotically efficient, and asymptotically normal. The maximum likelihood estimate of the vector of parameters $\beta$ is the value of $\beta$, $\hat{\beta}$, which minimizes the likelihood function of goodness of fit criterion and maximise the log likelihood (Dobson, 1983).

The maximum likelihood equations are in general nonlinear and have to be solved iteratively. The most common widely used algorithms include the Fisher scoring or iteratively reweighted least squares and the Newton-Raphson method. Maximum likelihood estimate calculations by the Fisher's scoring method is similar to iterative least squares procedure (Jorgensen, 1983). Maximum likelihood is the principal method of estimation used for all generalized linear models. For the assumed n independent observations, the likelihood function is given by

$$L(\theta; y) = \prod_{i=1}^{n} exp\{[y_i\theta_i - b(\theta_i)]/a_i(\emptyset) + c(y_i, \emptyset)\} \qquad (3.5)$$

this yields the log-likelihood of

$$l(\theta; y) = \sum_{i=1}^{n} \frac{[y_i\theta_i - b(\theta_i)]}{a_i(\emptyset)} + \sum_{i=1}^{n} c(y_i, \emptyset). \qquad (3.6)$$

The parameters of interest in the linear predictor are $\beta_1, \beta_2, ..., \beta_p$. Therefore to obtain maximum likelihood the equations are solved simultaneously

$$\frac{\partial}{\partial\beta_j} l(\theta_1, \ldots\ldots, \theta_n; \emptyset) = \sum_{i=1}^{n} \frac{\partial l_i}{\partial\beta_j} = 0, \text{ for } j\text{=1,2,…,p} \qquad (3.7)$$

where $l_i = \frac{[y_i\theta_i - b(\theta_i)]}{a_i(\emptyset)} + c(y_i, \emptyset)$. By the chain rule of differential calculus (3.7) can be expressed as

$$\frac{\partial l_i}{\partial\beta_j} = \frac{\partial l_i}{\partial\theta_i}\frac{\partial\theta_i}{\partial\mu_i}\frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta_j}. \qquad (3.8)$$

From the definition of $l_i$ above we obtain directly $\frac{\partial l_i}{\partial\theta_i} = \frac{y_i - b'(\theta_i)}{a_i(\emptyset)}$. Since $\mu = b'(\theta_i)$ it follows that $\frac{\partial l_i}{\partial\theta_i} = \frac{(y_i - \mu)}{a_i(\emptyset)}$ and that $\frac{\partial\mu_i}{\partial\theta_i} = b''(\theta_i) = V(\mu_i)$. The linear predictor $\eta_i = g(\mu_i) = \boldsymbol{x}_i'\beta$ supplies the last two terms required

$$\frac{\partial\eta_i}{\partial\beta_j} = x_{ij} \text{ and } \frac{\partial\eta_i}{\partial\mu_i} = g'(\mu_i) \text{ so } \frac{\partial\mu_i}{\partial\eta_i} = \frac{1}{g'(\mu_i)}.$$

Substituting all the above equations in (3.8), we obtain the maximum likelihood equations

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{a_i(\emptyset)V(\mu_i)g\prime(\mu_i)} = 0 \text{ for j=1,2,…,p.} \qquad (3.9)$$

whose solutions are maximum likelihood estimators of the $\beta_j s.$The equations are solved iteratively. This can be done using the Newton-Raphson method, or Fisher's scoring method. The Newton-Raphson method is the simplest numerical procedure for maximising the likelihood function (Green, 1984). The Newton-Raphson method converges quadratically, and the convergence is very fast when the initial guess is close to the solution (Everitt, 1987; and Thisted, 1988).

The commonly used automatic variable selection procedures for regression model building include forward selection, backward selection, and stepwise selection procedures. With forward selection, an effect is entered into the model singularly at each stage at a given critical p-value of which the model with the intercept only is the initial stage. Forward selection process is done until no more effects are entered into the model. Backward selection starts with the full model containing all the effects (saturated model) in its selection process, of which an effect which does not satisfy a given critical p-value is dropped from the model at each stage. The backward selection process is continued until no more effects are removed from the model. Stepwise selection procedure is a combination of backward and forward selection procedure of which a single effect leaves and another variable enters the model at each stage.

Type III sum of squares analysis is based on the calculation of the model sum of squares adjusting for other variables in the model, and does not depend on the order in which the effects are entered into the model.

Making inferences of the model involves testing the hypothesis about the parameters in the model, obtaining confidence intervals, assessing the validity of the fit of the model, and interpreting the results. After fitting the model, investigation on how well the model fits observed values is done. The overall goodness of fit tests of the GLM include the deviance, the Pearson chi-square, likelihood ratio tests which are described below.

**Deviance test:** The deviance provides a measure based on a twice log likelihood between the full model and the reduced model. For all possible exponential models, the maximum achievable log likelihood is $\boldsymbol{l(y,y)}$ in which the fitted values are equal to the observed data (McCullagh and Nelder, 1989). The fitted log likelihood is $\boldsymbol{l(\hat{\mu},y)}$. The deviance is given by

$$D^*(\boldsymbol{y},\boldsymbol{\hat{\mu}}) = 2l(\boldsymbol{y};\boldsymbol{y}) - 2l(\boldsymbol{\hat{\mu}};\boldsymbol{y}). \tag{3.10}$$

The deviance is used for model checking as well as inferential comparison of models, hence it tests the null hypothesis that the model fits the data. The deviance has exact and asymptotic $\chi^2$ distribution for normal models and non-normal models respectively when the model fits the data, with n-p degrees of freedom where n is the number of observations and p is the number of parameters in the model.

For binary data the deviance is $D = 2\sum_{i=1}^{n}\left(y_i log\frac{y_i}{\hat{\mu}_i} + (n_i - y_i)log\frac{(1-y_i)}{(1-\hat{\mu}_i)}\right)$. The null hypothesis for testing the fitness of the model is rejected in favour of alternative hypothesis at a given level of significance, if the calculated D is greater than the critical value (i.e. $D > \chi^2_{n-p,\alpha}$). Small values of deviance are obtained when the fitted model likelihood is similar to the saturated model likelihood, an indication that the fitted model is good (Collet, 2003), therefore maximising the model likelihood is equivalent to minimising the deviance. If the model fits the data perfectly, the deviance is zero. Hence it is expected that the calculated scaled deviance should not exceed the upper 100(1-$\alpha$) percent point as this may indicate a poor fit of the model to the data (Krzanowski, 1998). Sometimes the fitted model is declared not adequate if $\frac{D}{n-p} > 1$ , where D is the Deviance, *n* the number of observations and p the number of parameters (Montgomery et al., 2006).

**Akaike information criterion (AIC):** AIC is a model selection method used in comparing and selecting the best model. The best model is selected as the one with the lowest AIC value. AIC is defined by

$$AIC = -2\ln(\boldsymbol{L}) + 2p \tag{3.11}$$

where $\boldsymbol{L}$ is the maximised log likelihood, and p is the number of model parameters.

**Pearson chi-square test:** According to McCullagh and Nelder (1989), the Pearson chi-square statistic is given by

$$X^2 = \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2 / V(\hat{\mu}_i) \qquad (3.12)$$

where $V(\hat{\mu}_i)$ is the estimated variance function of the distribution under consideration, $y_i$ are responses, and $\hat{\mu}_i$ are fitted means.

The Pearson chi-square statistic (3.12) has exact chi-square distribution with n-p degrees of freedom for normal distributed models and has asymptotic chi-square distribution for non-normal models. The null hypothesis is rejected when the Pearson chi-square test statistic is greater than the critical value at a given level of significance.

**Model diagnostics:** Residuals are checked to evaluate the extent in which the data used in the model building supports the model. Analysis of residuals is essential before inferences are made about the model to be fitted to the data (McCullagh and Nelder, 1989). There are two types of residuals in the generalized linear models the Pearson residual, and the deviance residual.

Pearson residual is defined by

$$r_{pi} = \left\{ \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}} \right\} \qquad (3.13)$$

where $V(\hat{\mu}_i)$is the variance-mean function, $\hat{\mu}_i$is the fitted maximum likelihood estimate.

The deviance residual is defined by

$$r_{Di} = \left\{ \sqrt{2l(y_i; y_i) - 2l(\hat{\mu}; y_i)} \right\} \qquad (3.14)$$

where $l(\hat{\mu}; y_i)$is the fitted log likelihood function

The deviance residuals are approximately normally distributed hence making the assessment of these residuals appropriate and simple as compared to the Pearson residuals.

In standardised residual, the variance of the residual is scaled in order to diagnose outliers (extreme observations). Hence the standardised residual is given by

$$d_i = \frac{e_i}{\sqrt{MSE}} \quad i = 1,2, \dots., n \tag{3.15}$$

Where $e_i$ is the raw residual, $MSE$ is the estimate of the standard deviation of the error terms.

If the model fits the data the standardised residuals should lie within the range of ±3, as the outliers affect the parameter estimates in the model by overestimating or underestimating the model parameters. Curvature could be an indication of the importance to include a quadratic term in the explanatory variable(s) in the model (McCullagh and Nelder, 1989) when the residuals are plotted against the predictors, and a linear pattern could be an indication of omission of an important variable in the model.

It is essential to check if the specification of generalized linear model link function used is correct as misspecification of a link function could lead to biased estimates. Methods of evaluating a link function include analysis of deviance, refer to (3.10), of the assumed model, and inclusion of extra terms in the assumed model. Significant extra terms and significant reduction in deviance indicate an improper link function when extra variables are added (McCullagh and Nelder, 1989; and Pregibon, 1980). The link function could also be assessed by checking the changes in the deviance when a constructed variable is added into the model (Collet, 2003). The constructed variable is given by

$$z_i = -\{1 + \hat{p}_i^{-1} log(1 - \hat{p}_i)\} \tag{3.16}$$

where $\hat{p}_i$ is the fitted response probability of the *i*th observation.

Significant reduction of deviance, with the chi-square distribution at one degree of freedom, when a constructed variable is added into the model concludes that the hypothesised link is unsatisfactory.

The other way of checking the correct link specification is checking the results of the linear predicted value and the linear predicted value squared in the logistic model. If the model is correctly specified then the predicted value squared is insignificant whilst the predicted value specification is statistically significant at a given level of significance (Vittinghoff et al., 2005).

Observations which are outliers in the X-space of the explanatory variables are defined to have high leverage, $h_{ii}$, which is measured from the diagonal elements of the hat matrix, **H**, given by

$$\boldsymbol{H} = \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}^{\frac{1}{2}} \tag{3.17}$$

where $\boldsymbol{W}$ is the $n \times n$ diagonal matrix of weights, $\mathbf{X}$ is the $n \times p$ design matrix in (4.7), where $p$ is the number of model parameters.

The values of $h_{ii}$ are always between 0 and 1. Observations which have a $h_{ii}$ > 2p/n, (where p is the number of explanatory variables, and n is the number of observations), are considered to have high leverage and their value gets closer to 1 (Belsley et al., 2004; Hoaglin and Welsch, 1978; Puterman, 1988). At high leverage points, residuals have minimal variance hence constraining the detection of outlying observations if only residual assessment is solely done (Neter, et al., 1990). But still, for verification, it is important to assess points which are at a distance rather than the others in X-space as they could be of potential influence in the model.

An influential observation is an observation that has significant effect on the fitted model estimates, as it causes large changes on the estimated regression parameter as compared to other observations, when deleted from the data set during analysis. Assessment of influential observations is essential as it enables the location of points of influence and assesses how influential they are in the fitted model (Cook, 1977; Hinkley et al., 1991; Krzanowski, 1998). The measure of influence proposed by Cook (1977) is given by

$$C_i = \frac{(\hat{\boldsymbol{\beta}}-\hat{\beta}_{(i)})'\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}(\hat{\boldsymbol{\beta}}-\hat{\beta}_{(i)})}{pa(\emptyset)} \tag{3.18}$$

where $\hat{\beta}$ is the estimate of β with all observations in a sample, $\hat{\beta}_{(i)}$ is the estimate of β excluding observation $i$, $\boldsymbol{W}$ is the diagonal matrix of weights, $p$ is the number of explanatory variables, $a(\emptyset)$ is the dispersion parameter.

The Cook's distance could also be approximated by

$$C_i \doteq \frac{h_{ii}(r_i)^2}{p(1-h_{ii})} \tag{3.19}$$

where $r_i$ is the standardised residual, $h_{ii}$ is the measure of leverage, p is the number of model parameters.

Large $C_i$ indicates that the observation has a large influence on the parameter estimates compared to other observations in the data set. Commonly the observations with Cook's distance value close to one, needs to be scrutinised for further analysis.

## 3.2 General linear model

The general linear model is a member of the generalized linear model which assumes $Y_i$s are independent normal distributions with mean $\mu_i$ and variance $\sigma^2_i$. In the linear model the natural link is the identity link. That is,

$$\eta_i = g(\mu_i) = \mu_i = \boldsymbol{x}_i'\beta. \tag{3.20}$$

In this case estimator of $\beta$ can be found in a closed form as $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}$, where $\mathbf{W}$=diag($1/\sigma^2_i$).

When the assumptions of normality are violated, some of the options are to transform the responses to normality or use nonparametric general linear modelling procedures

(Sakia, 1992). Transformation of the response variable, which involves altering the scale of the initial measurement, is employed in order to achieve the assumptions of normality and constant variance thereby making the results of the analysis more valid (Box and Cox, 1964).

Box-Cox transformations, which are widely used involves the family of power transformations of which the response value $y_i$ is transformed to $y_i^{\lambda}$ given by

$$y_i^{\lambda} = \begin{cases} \frac{y_i^{\lambda}-1}{\lambda}; \ \lambda \neq 0 \\ log y_i; \ \lambda = 0 \end{cases} \qquad (3.21)$$

for $y_i$>0

where $\lambda$, is the unknown parameter which could be estimated from the data, $y_i^{\lambda}$ is the vector of transformed responses.

Power transformation is more appropriate for data that looks at amounts or counts (Hoaglin, et al., 1983). According to Box and Cox (1964), the optimal $\lambda$ is obtained by maximising the log likelihood function

$$Lmax(\lambda) = \frac{1}{2} n log \hat{\sigma}^2(\lambda) + log J(\lambda; y) \qquad (3.22)$$

where $\hat{\sigma}^2(\lambda) = \frac{s(\lambda)}{n}$, $s(\lambda)$ is the residual sum of squares, $J(\lambda; y)$ is the jacobian transformation parameter, n is the number of observations.

## 3.3 Quantile Regression

Koenker and Basset (1978) introduced quantile regression as a robust method for linear models. Quantile regression, which is appropriate for continuous response data (Koenker, 2005), is an alternative regression approach to regression of the mean. Quantile regression provides a complete picture of the behaviour of the data set in the model compared to modelling with the mean as it uses all data in fitting the regression quantiles and is based on least absolute value regression, of which the model is fitted to the data by minimising the sum of weighted absolute residuals. Quantile regression uses the median as a measure of central location, and it extends in measuring the relationship between the response variable and the covariates in the non central parts of the response variable.

According to Koenker and Machado (1999), quantile regression could be applied in statistical analysis of both linear and non linear response modelling, and extends in flexible application in parametric and nonparametric methods. Quantile regression has been applied in economics, environment, health and medicine (Austin et al., 2005; Koenker and Hallock, 2001).

According to Konker and Basset (1978), in ordinary quantile regression, a random variable Y is characterised by the following distribution function

$$F(y) = Prob(Y \leq y)$$

then the $\tau^{\text{th}}$ quantile of Y is defined as the inverse function

$$F^{-1}(\tau) = inf\{yF(y) \geq \tau\}$$

where $0 < \tau < 1$. The median is then $F^{-1}\left({}^{1}\!/_{2}\right)$

In this thesis, quantile regression is used with the aim of estimating conditional quantiles of fertilizer usage per acre given a set of predictor variables. In this case, we are not only interested in the median of fertilizer usage per acre, but also conditional groupings of fertilizer application per acre, to distinguish between the performance of those people who apply lower levels of fertilizer per acre compared to those who apply more fertilizer per acre.

According to Koenker (2005), the linear model for the $\tau^{th}$ quantile is given by

$$y_i = \boldsymbol{x}_i' \beta_\tau + e_i \quad i=1,\ldots,n \tag{3.23}$$

where the $\tau^{th}$ quantile of $e_i$ is zero, Y is n × 1 vector of dependent responses, $\beta$ is p × 1 vector of known regression parameters which depend on $\tau$, $\boldsymbol{x}_i$ is the p × 1 vector of explanatory variables, and e is n × 1 vector of independent and identically distributed random errors, having a zero median, also the errors are independent of the regressors, at a given quantile, but quantile regression can also accommodate heterogeneous errors.

Parameters of the conditional quantile regression function can be estimated by minimising the objective function, also called the check function or loss function (Koenker, 2005), given by

$$\hat{\beta}_\tau = argmin_{\beta_\tau \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau \left(y_i - \boldsymbol{x}_i{}' \beta_\tau\right) \tag{3.24}$$

where $y_i$ is the dependent variable, $\boldsymbol{x}_i$ is the p × 1 vector of explanatory variables, $\rho_\tau$ is the loss function which is solved by

$$\rho_\tau(u) = u\big(\tau - I(u < 0)\big) \text{ for some } \tau \in (0,1)$$

where $u$ is the difference between observed value and estimated value, $I(.)$ is the indicator function

Therefore the conditional median $\tau = 0.5$ can be calculated by

$$\hat{\beta}_\tau = argmin_{\beta_{0.5} \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau \left(y_i - \boldsymbol{x}_i{}' \beta_{0.5}\right) \tag{3.25}$$

which cannot be solved explicitly as the check function is not initially differentiable, but it could be solved by a modified simplex algorithm in order to get conditional median estimates.

Information criteria are one of the bases for model selection. They provide a powerful tool for choosing a model amongst models that best fits the data. The information criteria includes the Akaike's Information Criteria (AIC), and the Bayesian Information Criteria (BIC) also referred to as Schwarz Information Criteria (SIC) which is closely related to AIC. AIC is a test statistic for measuring goodness of fit of the selected model. A model with the lowest AIC is considered the best model. The AIC score (Koenker, 2005) could be obtained by

$$AIC(j) = \log(\hat{\sigma}_j) + p_j \tag{3.26}$$

where $\hat{\sigma}_j$ is objective function i.e. the function to be minimized and $p_j$ is the number of model parameters.

According to Hampel (1986), likelihood ratio test (also referred to as *p* test (Koenker and Machado, 1999)) is equivalent to testing with F-test. Let

$$\hat{V}_\tau = argmin_{\beta_\tau \in \mathbb{R}^p} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i{}'\beta_\tau\right) \tag{3.27}$$

denote the value of objective function of the unrestricted minimiser $\hat{\beta}_\tau$ , and let

$$\tilde{V}_\tau = argmin_{\beta_\tau \in \mathbb{R}^p | R\beta_{\tau=r}} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i{}'\beta_\tau\right) \tag{3.28}$$

refer to objective function under restricted estimator $\tilde{\beta}_\tau$.
under i.i.d. error assumption, the test statistic (Koenker and Machado, 1999),

$$L_\tau = \frac{2(\tilde{V}_\tau - \hat{V}_\tau)}{\tau(1-\tau)s(\tau)}$$

is asymptotically chi-squared distributed under the null hypothesis with r degrees of freedom, where $s(\tau) = {}^{1}/_{f}\left(F^{-1}(\tau)\right)$

Methods for estimating the confidence interval in quantile regression analysis include direct method (Zhou and Portnoy, 1996), rank score method, and resampling method. Direct method is based on the estimates which ought to be asymptotically normal, whilst the resampling method uses bootstrap techniques which assess accuracy of the sample quantile. The rank score tests constructs the confidence interval based on the inversion of rank score tests which are not asymmetric but centred around zero in order to generate sequential but fixed length confidence intervals. Rank score test is an order statistic, which performs better for small samples, it is a robust measure to model assumptions as it ought to be less sensitive to heterogeneous error distributions (Koenker, 2005), and does not require estimation of sparsity function, also referred to as a nuisance parameter. According to Koenker and Machado (1999), the rank score test statistic is given by

$$\hat{a}_n(\tau) = argmax\{y'a|X_1'a = (1-\tau)X_1'a, a \in [0,1]^{\,n}\} \qquad (3.29)$$

where e denotes an n-vector of 1's, and X has been partitioned as $[X_1 \vdots X_2]$

The quantile regression conditional goodness of fit test for a given quantile is obtained by use of coefficient of determination (pseudo R$^2$) (Koenker and Machado, 1999) given by

$$R^1{}_{(\tau)} = 1 - \frac{\hat{V}_{(\tau)}}{\tilde{V}_{(\tau)}} \qquad (3.30)$$

where $\hat{V}_{(\tau)}$ as specified in eqn (3.31), is the sum of the weighted absolute deviation of a given model, and $\tilde{V}_{(\tau)}$ is the weighted sum minimised for reduced model with only the intercept

The coefficient of determination in modelling the mean $(R^2)$, cannot be compared to the coefficient of determination in modelling the quantiles $(R^1)$, because of their differences in nature of obtaining the values as the mean coefficient of determination works on the global/ entire distribution whilst the quantile coefficient of determination is based on the local measure of a given quantile.

The Mahalanobis distance is a measure for detecting multivariate outliers and it is given by

$$MD_i = \sqrt{\left((\boldsymbol{x}_i - \boldsymbol{t})'\boldsymbol{C}^{-1}(\boldsymbol{x}_i - \boldsymbol{t})\right)} \tag{3.31}$$

where **t** is the estimated multivariate location estimator, **C** is estimated covariance matrix of explanatory variables, $\boldsymbol{x}_i$ is the ith row vector of matrix **X**.

Large values of MD is an indication of outliers, generally the value that exceeds the cutoff point of $\sqrt{\chi^2_{p,0.975}}$ is an outlier. The relationship between the Mahalanobis distance (MD$_i$) and the hat matrix is given by

$$h_{ii} = \frac{1}{n-1} MD_i^2 + \frac{1}{n} \tag{3.32}$$

where n is the number of observations

Diagonal elements of the hat matrix are used to detect leverage points, but this hat matrix is affected by the masking effect of the residuals (Hubert et al., 2008), whereby the effect of multiple outliers could have an influence on the estimates. In order for the Mahalanobis distance to be robust to outliers, robust estimates for covariance matrix, such as minimum covariance determinant estimator (MCD), is used (Rousseeuw and Driessen, 1999). The use of robust estimators of location leads to robust distance (RD) which measures the robust distance of an observed value and robust estimated value, given by

$$RD_i = \sqrt{\left(\boldsymbol{x}_i - T(A)\right)' C^{-1}\left(\boldsymbol{x}_i - T(A)\right)} \tag{3.33}$$

where T(A) is the robust multivariate location, C is the scale estimate generated by the minimum covariance determinant (MCD).

## 3.4 Generalized linear model versus quantile regression

Generalized linear models have the advantage of modelling the response variable which is not normally distributed. The challenge in generalized linear models lies in the prior identification of the link function as well as the variance structure. Model misspecification could also be introduced when a wrong distribution of the response variable is assumed. As a result of model misspecification, there is bias of the regression parameter estimates and mean estimates of the response variable making the results unreliable.

The generalized linear model estimates the mean of the response variable given a set of covariates, but different parts of the response variable could be affected differently by the independent variables therefore quantile regression goes further in modelling the effects of covariates at different levels of the response variable.

Quantile regression is advantageous over the regression of the mean as it uses the median as a measure of central tendency rather than the mean, therefore a quantile regression technique is robust against outliers compared to the mean regression which is affected by the presence and masking effect of the outliers as they make the mean regression results not to be efficient and meaningful. Quantile regression results are based on the estimates of the conditional groupings of the response variable given a set of predictor variables. Quantile regression is more resistant to extreme values than the generalized linear models, because the outlying observations significantly affect the model estimates whilst quantile regression tends to be robust in the presence of outliers, heavy tailed distributions, and heterogeneity (Koenker, 2005). In cases where we have a weak relationship between the response and explanatory variables, quantile regression might be employed whereby the conditional quantile estimates of the response variable may be modelled to obtain efficient results as in modelling the mean (Cade and Noon, 2003). In quantile regression, the distribution assumptions are relaxed, as it does not assume the normality of the response variable and normality of the errors (Koenker and Machado, 1999). Quantile regression provides flexibility in the analysis of the data as there is no involvement of the link function as is done in analysis of the generalized linear models, as well as specification of the variance link to the mean. For distributions which are skewed and have heavy tails, there are differences of the results of the mean and median thereby there is loss of precision based on the mean regression than the median which has robustness properties.

# Chapter 4

# RESULTS AND DISCUSSION

## 4.1 General linear model results

The factors in the model were the variables given in Table 2.1, with the response variable as the fertilizer usage per acre. The full model for the amount fertilizer used per acre consisted of the main effects of all the explanatory variables and the two way interaction of these variables.

Selection of the influential variables/factors for the final model was done using the automatic stepwise procedure in Proc GLMSELECT of SAS (Cohen, 2006). The cut off point of significance for an effect to stay or enter in the model was 0.10. The reduced model had the following interaction effects in the model land by training, land by saving interaction, training by saving, saving by irrigation, distance travelled to source fertilizer by total annual income, land size by timeline, and number of plots cultivated by timeline.

Analysis of residuals for the selected model does show some departure from model assumptions of constant variance and normality as shown in Figure 4.1. In Figure 4.2, observations number 77, 340, and 271 were identified to have high residuals. Figure 4.3 and Figure 4.4 shows that observation number 345 has high leverage and high influence on the model.

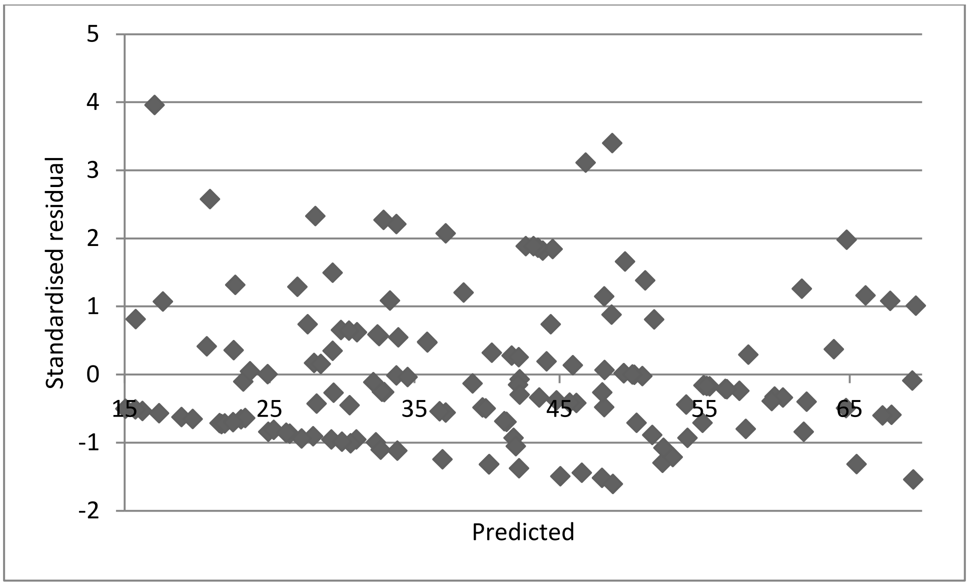

Figure 4.1: Plot of predicted amount of fertilizer per acre against standardised residuals in a general linear model

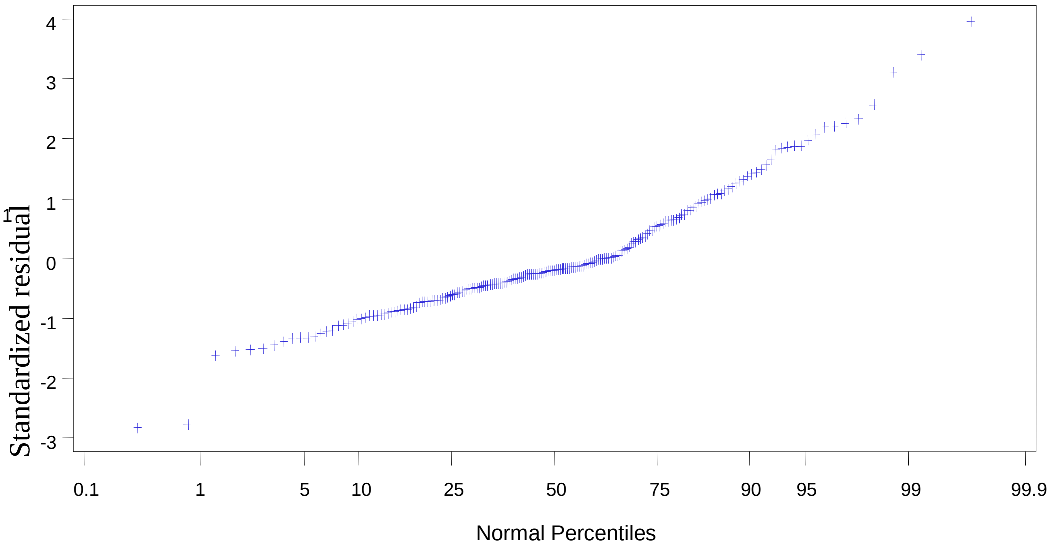

Figure 4.2: Normal probability plot in a general linear model

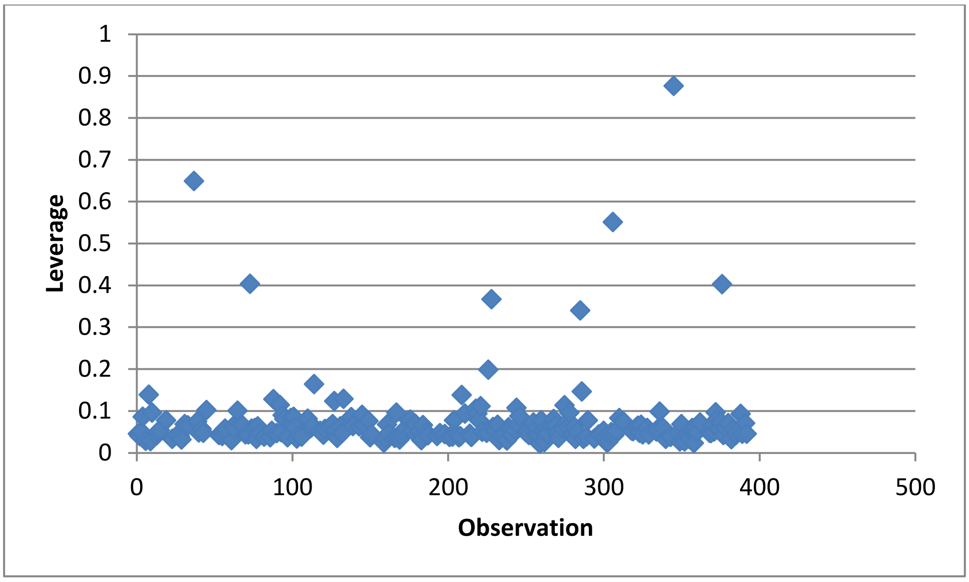

Figure 4.3: Index plot of leverage values

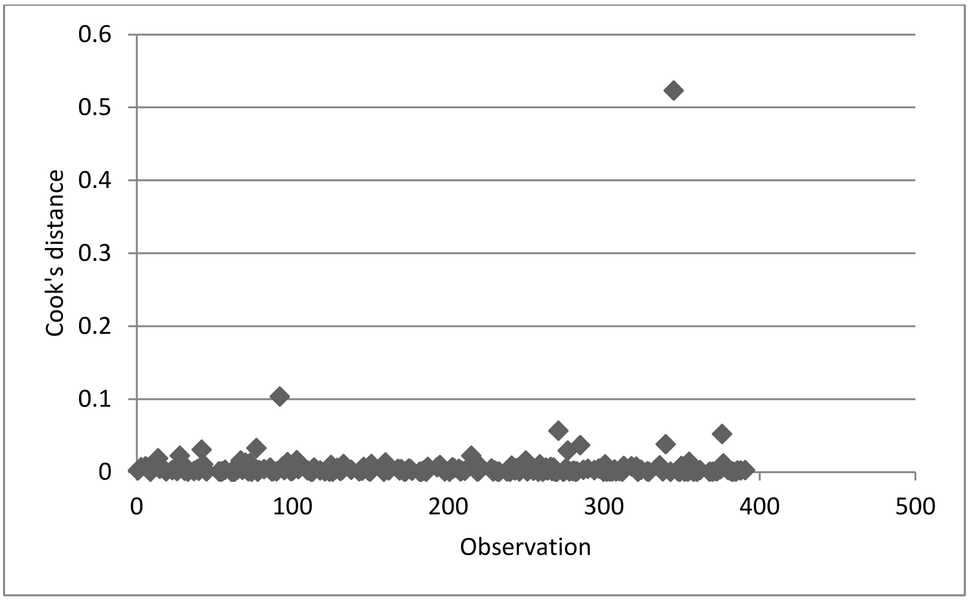

Figure 4.4: Index plot of Cook’s distance in a general linear model

Due to the presence of points of high residual and high influence, raw data were rechecked and no anomalies in the data were detected to have caused the high values in standardised residual. Dropping of the extreme observations one at a time indicated not much effect of the model conclusions as shown in Table 4.1 hence all the extreme observations were retained in the final model.

Table 4.1: A general linear model test analysis for full and deleted observation

| Source | Full data | | Observation 345 deleted | | Observation 77 deleted | | Observation 340 deleted | | Observation 271 deleted | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Type III SS | P-value | Type III SS | P-value | Type III SS | P-value | Type III SS | P-value | Type III SS | P-value |
| District | 1821.9613 | 0.1833 | 2357.85333 | 0.1288 | 1509.19920 | 0.2147 | 2087.20770 | 0.1429 | 1673.77394 | 0.1845 |
| Head | 1198.6454 | 0.2801 | 1221.67389 | 0.2735 | 1657.28853 | 0.1936 | 668.67653 | 0.4060 | 1642.69620 | 0.1886 |
| Training | 406.5593 | 0.5288 | 815.78941 | 0.3707 | 266.87979 | 0.6012 | 251.92952 | 0.6098 | 441.10646 | 0.4949 |
| Source | 35.0307 | 0.8533 | 19.21897 | 0.8906 | 0.19098 | 0.9888 | 240.32909 | 0.6181 | 2.47344 | 0.9592 |
| Saving | 2951.2218 | 0.0909 | 839.12680 | 0.3639 | 2439.92039 | 0.1152 | 2456.95280 | 0.1121 | 3823.60806 | 0.0456 |
| Irrigation | 4126.1133 | 0.0459 | 3832.15019 | 0.0533 | 3080.52656 | 0.0770 | 3522.05324 | 0.0575 | 2619.89525 | 0.0973 |
| Animal manure | 21.3598 | 0.8852 | 98.28385 | 0.7558 | 0.00939 | 0.9975 | 0.48169 | 0.9822 | 1.84627 | 0.9648 |
| Plots | 686.8041 | 0.4132 | 600.61240 | 0.4423 | 796.05985 | 0.3670 | 669.44581 | 0.4057 | 836.80918 | 0.3475 |
| Nogrow | 1016.2500 | 0.3198 | 1377.62678 | 0.2450 | 624.76158 | 0.4241 | 2174.18439 | 0.1349 | 1886.27240 | 0.1590 |
| Fertgood | 538.0691 | 0.4688 | 459.50646 | 0.5014 | 825.61516 | 0.3583 | 932.17666 | 0.3267 | 951.71224 | 0.3164 |
| Fert_perc | 1514.5954 | 0.2249 | 2022.39362 | 0.1594 | 652.38105 | 0.4141 | 1917.85518 | 0.1601 | 428.64430 | 0.5010 |
| Timeline | 58.3184 | 0.8114 | 8.44242 | 0.9274 | 3.80898 | 0.9502 | 0.05156 | 0.9942 | 1.80769 | 0.9651 |
| Land | 2787.9988 | 0.1002 | 57.07333 | 0.8126 | 2973.76685 | 0.0822 | 2795.54281 | 0.0903 | 3004.70465 | 0.0760 |
| Lenglast | 1304.5886 | 0.2599 | 1906.40768 | 0.1718 | 943.37347 | 0.3262 | 928.21734 | 0.3277 | 540.28736 | 0.4501 |
| Distance | 288.5486 | 0.5957 | 898.35935 | 0.3475 | 471.26387 | 0.4875 | 250.90397 | 0.6105 | 795.68524 | 0.3596 |
| Total_income | 128.0080 | 0.7237 | 356.68867 | 0.5536 | 132.61209 | 0.7125 | 120.09597 | 0.7245 | 202.16951 | 0.6439 |
| Land*Training | 4388.3518 | 0.0396 | 5493.70065 | 0.0210 | 4086.00150 | 0.0420 | 3945.21770 | 0.0445 | 4770.67859 | 0.0257 |
| Land*Saving | 6966.8426 | 0.0098 | 926.08528 | 0.3402 | 6062.06931 | 0.0135 | 7015.47071 | 0.0076 | 7144.56689 | 0.0065 |
| Training*Saving | 5110.4786 | 0.0265 | 5680.92705 | 0.0189 | 4421.27291 | 0.0345 | 4039.48766 | 0.0421 | 5598.82485 | 0.0158 |
| Saving*Irrigation | 2892.7486 | 0.0941 | 3235.86406 | 0.0756 | 2285.69173 | 0.1272 | 2269.38548 | 0.1267 | 3950.91678 | 0.0421 |
| Distance*Total_income | 642.8865 | 0.4286 | 434.84516 | 0.5131 | 607.51957 | 0.4306 | 875.18405 | 0.3419 | 548.58514 | 0.4466 |
| Land*Timeline | 26747.3520 | <.0001 | 23739.65189 | <.0001 | 27947.64639 | <.0001 | 26971.50587 | <.0001 | 27041.65670 | <.0001 |
| Plots*Timeline | 15228.2172 | 0.0002 | 16431.76321 | <.0001 | 15130.35633 | 0.0001 | 13762.80369 | 0.0002 | 15008.05990 | <.0001 |

The original general linear model with all the observations included indicates that the overall model fit is significant (P<0.0001) as shown in Table 4.2.

Table 4.2: Analysis of variance in a general linear model

| Source | DF | Sum of Squares | Mean Square | F Value | P-value |
|---|---|---|---|---|---|
| Model | 23 | 104768.1788 | 4555.1382 | 4.46 | <.0001 |
| Error | 177 | 180724.8872 | 1021.0446 | | |
| Corrected Total | 200 | 285493.0660 | | | |

Since the general linear model assumptions seems not to be fulfilled due to violation of the normality and constant variance assumption violation, fitting of another model with Box-Cox transformation is proposed.

## 4.2 General linear model transformed results

In order to achieve normality and constant variance, Box-Cox transformation in TRANSREG procedure of SAS (SAS, 2004) was used to obtain the appropriate $\lambda$ for the response power transformation of the data. The estimation procedure of optimal $\lambda$ was done using the maximum likelihood method (Draper and Smith, 1981). Some observations had the response value of zero, hence the TRANSREG procedure in SAS failed to run the analysis, therefore a constant value (c=1) was added to the response before log transformation procedure. Plots for the standardised residuals against the predicted values, and normal probability plot are presented below in Figures 4.5 and 4.6 for the transformed response.

As shown in Figure 4.5 and Figure 4.6, data transformation based on Box-Cox transformation failed to meet the constant variance assumption, this entails that the generated results from the transformed response variable, as shown in Table 4.3 and Table 4.4, provides the biased estimates.

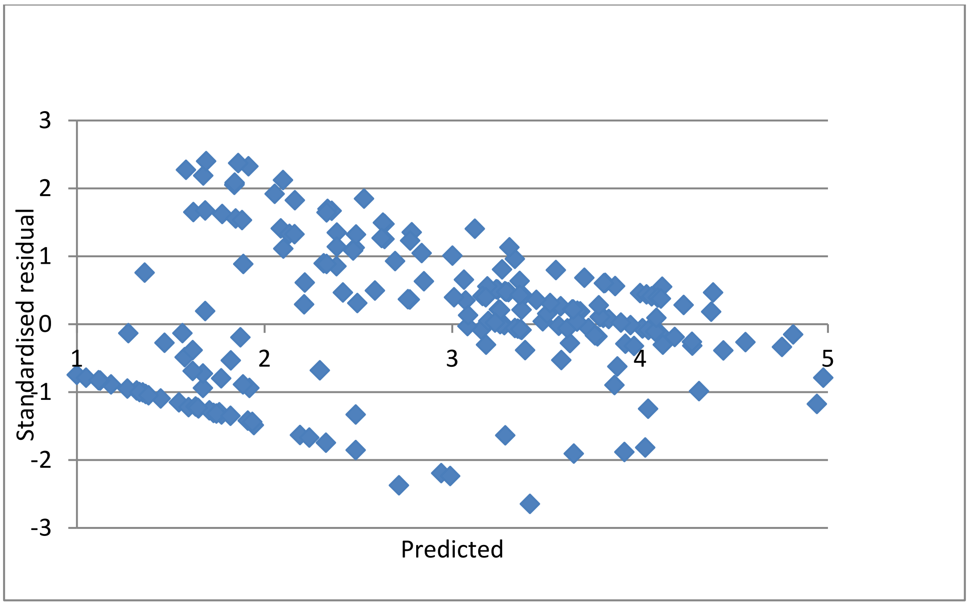

Figure 4.5: Plot of predicted amount of fertilizer per acre against standardised residuals in a transformed general linear model

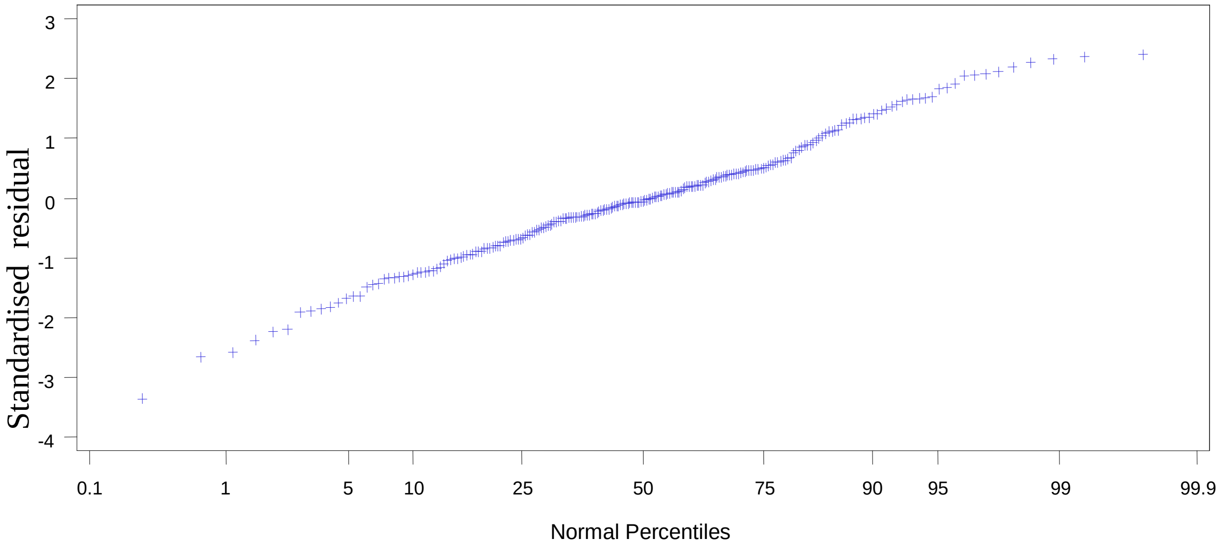

Figure 4.6: Normal probability plot in a transformed general linear model

Similar to the results got in Table 4.2, the analysis of variance in transformed general linear model indicates that the overall fit if the model is highly significant (P<0.0001). But unlike earlier results, the transformed general linear model has R-square statistic of 0.51 which is higher than the original general linear model which had R-square statistic of 0.37.

Table 4.3: Analysis of variance in a transformed general linear model

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 23 | 349.3401433 | 15.1887019 | 8.06 | <.0001 |
| Error | 177 | 333.4292485 | 1.8837811 | | |
| Corrected Total | 200 | 682.7693917 | | | |

Table 4.4: A general linear transformed model parameter estimates

| Source | Type III SS | P-value |
|---|---|---|
| District | 1.1746 | 0.4308 |
| Head | 6.6435 | 0.0620 |
| Training | 4.3591 | 0.1300 |
| Source | 0.21822 | 0.7340 |
| Saving | 7.2023 | 0.0521 |
| Irrigation | 2.2297 | 0.2781 |
| Animal manure | 0.6647 | 0.5533 |
| Plots | 0.4285 | 0.6340 |
| Nogrow | 0.7899 | 0.5181 |
| Fertgood | 0.0539 | 0.8658 |
| Fert_perc | 1.1310 | 0.4395 |
| Timeline | 10.7728 | 0.0178 |
| Land | 34.3920 | <.0001 |
| Lenglast | 0.3411 | 0.6710 |
| Distance | 0.8140 | 0.5118 |
| Total_income | 0.1925 | 0.7496 |
| Land*Training | 15.6314 | 0.0045 |
| Land*Saving | 11.0840 | 0.0163 |
| Training*Saving | 9.3101 | 0.0275 |
| Saving*Irrigation | 3.1483 | 0.1978 |
| Distance*Total_income | 0.0815 | 0.8355 |
| Land*Timeline | 84.2469 | <.0001 |
| Plots*Timeline | 20.6151 | 0.0011 |

Comparing results in Table 4.1 and Table 4.4 shows that irrigation is significant in the transformed model as well as the interaction between saving and irrigation at 10% level of significance, these terms were not significant in the original general linear model.

## 4.3 Gamma model results

As an alternative to modelling the data based on generalized linear models for skewed data through transformation, analysis based on changing the link function assuming the gamma distribution model approach is proposed. Analysis was done assuming a gamma distribution with non canonical log link, and then canonical inverse link in the GENMOD procedure of SAS. A positive, 1, constant was added to the response before analysis so as to cater for the zero responses.

A plot of standardised residuals against the predicted values as shown in Figure 4.7 and Figure 4.8 indicate that the residuals do comply with the model assumption under the inverse link function as compared to the log link.

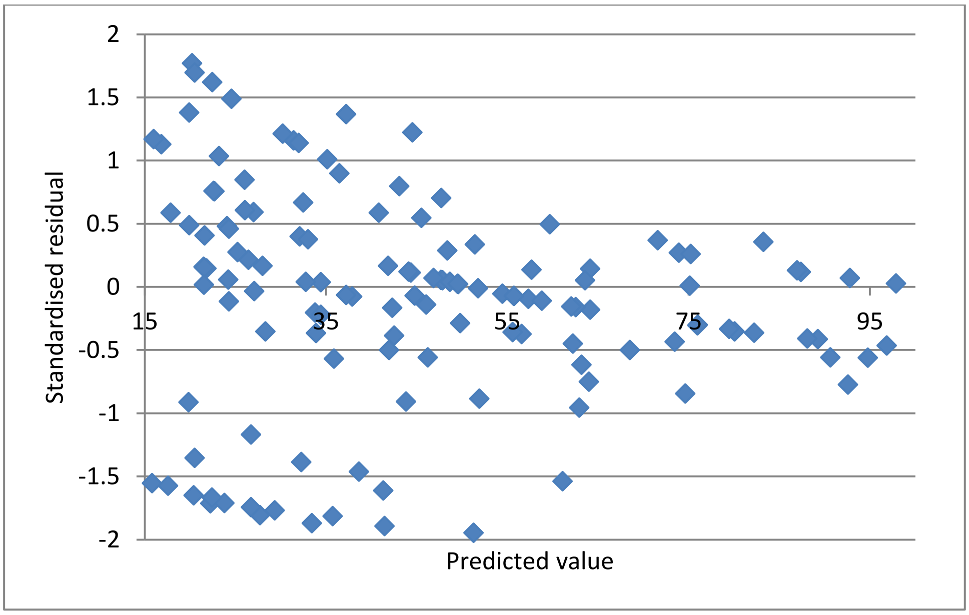

Figure 4.7: Plot of predicted amount of fertilizer per acre against standardised residuals in gamma model with inverse link

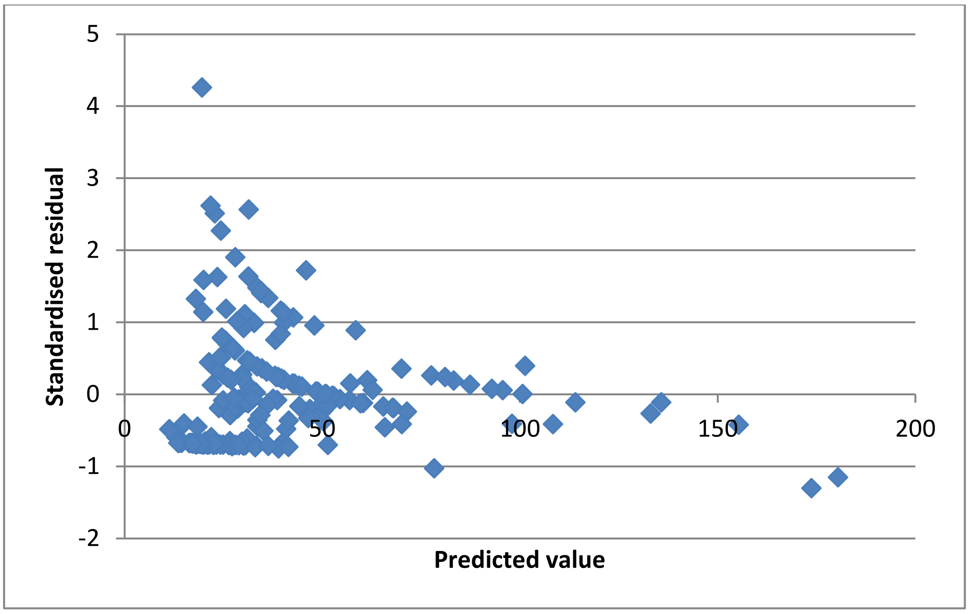

Figure 4.8: Plot of predicted amount of fertilizer per acre against standardised residuals in gamma model with a log link

Diagnosis of influential observations was assessed in the two gamma models. The results from the assessment indicates that observation number 28, 67, 125, 160, 271, and 345 had high influence on the gamma model with a log link as shown in Figure 4.9. Figure 4.10 shows that observation number 376 had high influence in the inverse link gamma model.

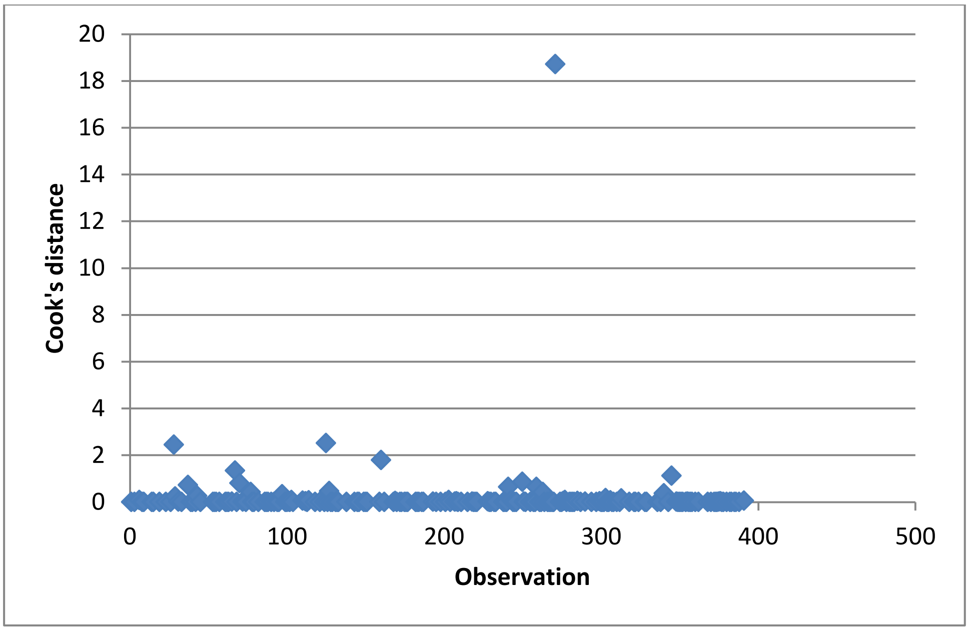

Figure 4.9: Index plot of Cook’s distance in a gamma model with a log link

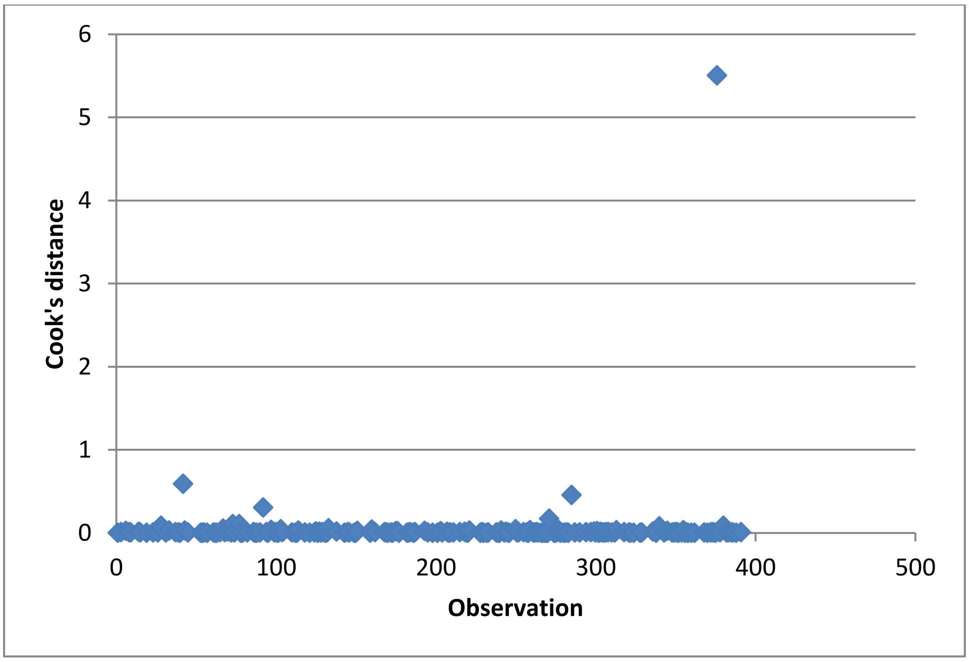

Figure 4.10: Index plot of Cook’s distance in a gamma model with inverse link

Another comparison between the log gamma link and the inverse gamma link shows that the standard errors are lower for inverse link than log link as shown in Table 4.5. Further assessment on the proper link function was done by comparing changes in the deviance when an extra term, land*land, was added in the model. As shown in Table

4.6, highly significant extra term, and significant reduction in the deviance indicate improper log link function when an extra variable was added in the model. The results indicate that the gamma fit with inverse link outperform the gamma fit with log link.

Table 4.5: Parameters for the gamma model log link and inverse link

| | Log link | | | Inverse link | | |
|---|---|---|---|---|---|---|
| Parameter | Estimate | Standard Error | P-value | Estimate | Standard Error | P-value |
| Intercept | 3.5054 | 0.7354 | <.0001 | 0.0314 | 0.0146 | 0.0312 |
| District | -0.1945 | 0.2947 | 0.5093 | 0.0030 | 0.0067 | 0.6590 |
| Head | 0.1983 | 0.2021 | 0.3264 | -0.0021 | 0.0055 | 0.6997 |
| Training | -0.2464 | 0.4769 | 0.6055 | 0.0114 | 0.0111 | 0.3024 |
| Source | 0.0304 | 0.2172 | 0.8887 | 0.0009 | 0.0051 | 0.8589 |
| Saving | 0.6940 | 0.4759 | 0.1447 | -0.0032 | 0.0113 | 0.7759 |
| Irrigation | -0.1227 | 0.3490 | 0.7252 | 0.0004 | 0.0102 | 0.9716 |
| Animal manure | -0.2222 | 0.2302 | 0.3344 | 0.0018 | 0.0052 | 0.7356 |
| Plots | 0.1508 | 0.1689 | 0.3717 | -0.0038 | 0.0027 | 0.1604 |
| Nogrow | -0.1672 | 0.2209 | 0.4492 | 0.0009 | 0.0057 | 0.8680 |
| Fertgood | 0.0120 | 0.2397 | 0.9601 | -0.0016 | 0.0054 | 0.7684 |
| Fert_perc | -0.6869 | 0.3190 | 0.0313 | 0.0093 | 0.0106 | 0.3805 |
| Timeline | -0.6731 | 0.6185 | 0.2765 | -0.0048 | 0.0179 | 0.7882 |
| Land | 0.0388 | 0.2983 | 0.8965 | 0.0011 | 0.0057 | 0.8415 |
| Lenglast | 0.0360 | 0.0356 | 0.3122 | -0.0008 | 0.0011 | 0.4462 |
| Distance | 0.0075 | 0.0136 | 0.5808 | -0.0002 | 0.0003 | 0.5859 |
| Total_income | 0.0000 | 0.0000 | 0.4599 | -0.0000 | 0.0000 | 0.3920 |
| Land*Training | -0.1378 | 0.2766 | 0.6184 | 0.0034 | 0.0037 | 0.3692 |
| Land*Saving | -0.5148 | 0.1884 | 0.0063 | 0.0048 | 0.0041 | 0.2412 |
| Training*Saving | 0.6668 | 0.4503 | 0.1387 | -0.0197 | 0.0109 | 0.0707 |
| Saving*Irrigation | -0.4216 | 0.4370 | 0.3347 | 0.0061 | 0.0114 | 0.5930 |
| Distance*Total_income | 0.0000 | 0.0000 | 0.7334 | 0.0000 | 0.0000 | 0.6618 |
| Land*Timeline | 1.1085 | 0.2372 | <.0001 | -0.0127 | 0.0049 | 0.0089 |
| Plots*Timeline | -0.6657 | 0.2640 | 0.0117 | 0.0173 | 0.0086 | 0.0450 |
| Scale | 0.5911 | 0.0000 | | 0.5112 | 0.0000 | |

Table 4.6: Assessing the link

| Criteria | Log link | | Inverse link | |
|---|---|---|---|---|
| | Selected model | With extra term | Selected model | With extra term |
| Deviance | 299.4536 | 279.4537 | 346.2617 | 337.5285 |
| Log likelihood | -883.7016 | -875.3400 | -901.977 | 898.9523 |
| P-value | | 0.0004 | | 0.0328 |

In modelling the mean of amount of fertilizer use per acre, three models were used to model the data which included regression based on assumption that the response distribution is normal, Box-Cox transformation of the response amount of fertilizer use per acre, and the gamma model assuming a log link and its alternative inverse link. In all the models that were assessed, statistically significant results were got in the interaction between land and timeline as well as the interaction of plots and timeline.

General linear model assumes independency, homoscedasticity and normality of the observations, but the results failed to comply with the model assumptions making the results unreliable. Transformation was conducted using logarithmic transformation of the response amount of fertilizer use per acre, but reliability of the results is still questionable as the assumption of constant variance was not met.

Modelling the mean fertilizer per acre based on the gamma model indicated that the gamma model with inverse link outperformed the gamma model with log link but both the models revealed some observations of high undue influence on the models making the estimates of the model parameters unreliable and the validity of the results could be questionable. Removing the observations of high influence in the model could result in loosing some of the essential information and biased results.

## 4.4 Quantile regression model results

In the analysis, the fertilizer applied per acre will be subdivided into several quantiles so as to investigate the effect of the predictors on several levels of the amount of fertilizer use per acre. Quantile regression will be used to estimate the conditional quantiles of fertilizer usage per acre to complement the results obtained through generalized linear modelling. In this case the interest will not only be on the mean of fertilizer applied per acre, but also on the conditional groupings of fertilizer applied per acre to distinguish between the factors of the households who applied lower levels of fertilizer per acre compared to those who applied more fertilizer per acre.

Different quantiles (25th , 50th, and 75th percentile) were estimated using a simplex algorithm (Barrodale and Roberts, 1973). In order to reduce computation burden, a direct method was used to generate confidence intervals across the quantiles in the QUANTREG experimental procedure in SAS. All the terms which were used in linear modelling were used in the quantile regression model so as to ease comparison between the two models. No point was detected to have high leverage as shown in Figure 4.11.

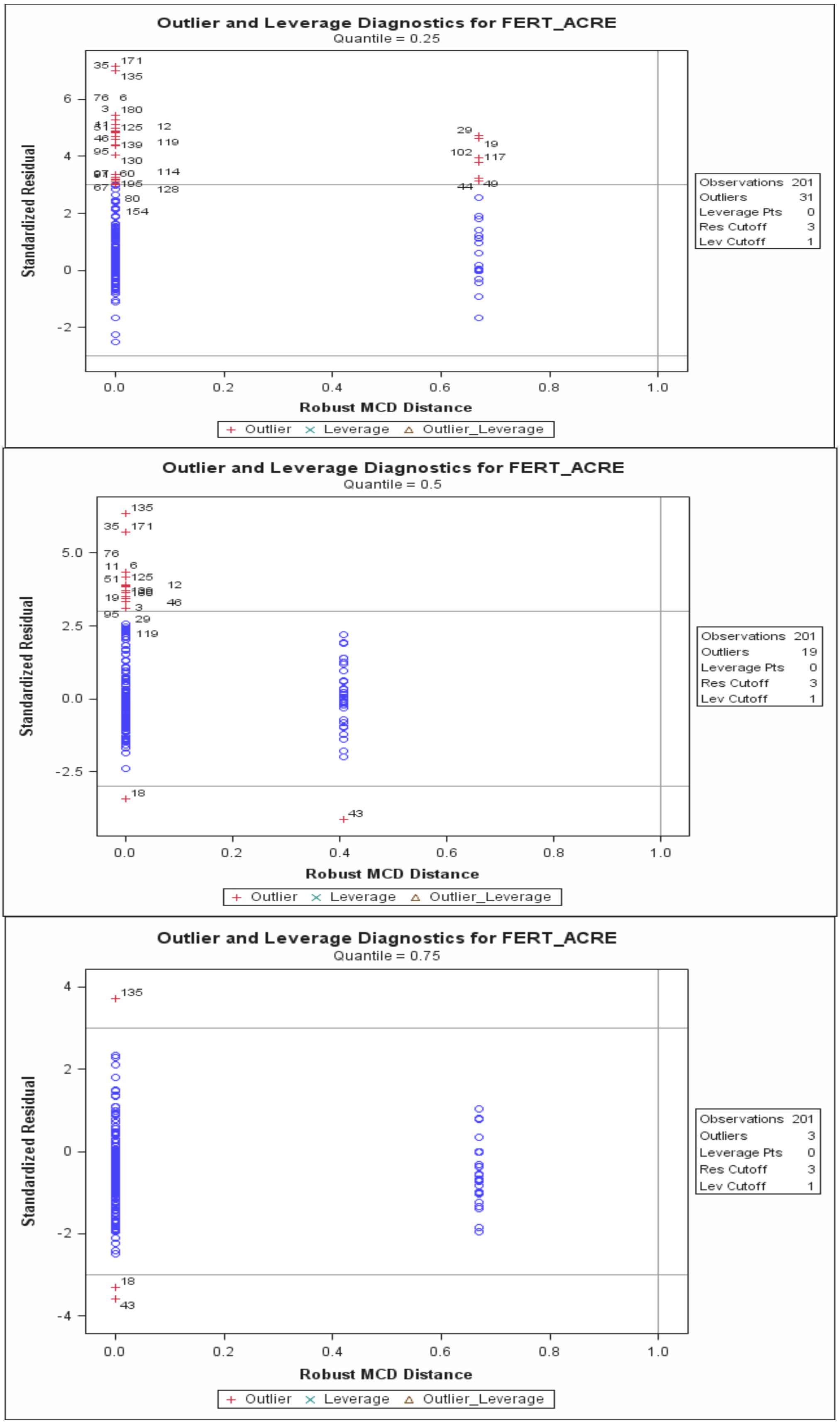

Figure 4.11: Plot of standardized residual versus robust Mahalanobis distance

The robust parameter estimates for the final selected model are presented in Table 4.5. The calculated coefficient of determination (pseudo $R^2$) varied across the quantiles which indicates the differences in the amount of fertilizer use per acre explained by the model at different quantiles. The median has the highest coefficient of determination of 0.6753, followed by the 0.75th quantile (pseudo $R^2$=0.6616), and 0.25th quantile has coefficient determination of 0.6280. The calculated AIC at the median is 2092.8660,  which is higher than the values at the other quantiles of which the 25th percentile reported the lowest AIC value of 1418.7303, and the 0.75th quantile had AIC of 1932.4203.

Table 4.7 and Figure 4.12 shows that there is variation of the results generated across the different quantile levels. There is a decreasing trend of the significant coefficients 5% level of significance on the number of plots per household at as we move from the lower quantile level to the median, but there is an increasing trend as we move from the median to the high quantile level which indicates that smallholder farmers at 0.25th quantile of the amount fertilizer use per acre distribution have additional 14.30 increase in fertilizer usage per acre whilst those at the 0.75th quantile have additional 12.49 increase and the small holder farmers at the median have additional 9.72 increase in amount of fertilizer use per acre. The increase in fertilizer use results amongst households who have more plots contradicts earlier findings by Chirwa (2003). Households who leave some of their plots uncultivated shows a decreasing slope in the levels of fertilizer that they apply per acre, but the estimates are significant at the 0.75th quantile at 10% level of significance which suggests that leaving some plots uncultivated has a decreasing effect on the levels of fertilizer application amongst the 25% of the small holder farmers who apply more levels of fertilizer controlling for other explanatory variables in the model. The results on low or no fertilizer usage amongst small holder farmers who leave some uncultivated land confirms earlier findings as established in the Malawi National land policy (Government of Malawi, 2002). Timeline has insignificant negative slopes in the first quantile and third quantile, but the estimates are significant at 10% level at the median.  The interaction between irrigation and household saving shows a significant decreasing slope from the 0.25th quantile to the 0.75th quantile but the interaction effect is stronger amongst households who apply higher levels of fertilizer per acre, which implies that the joint effect of irrigation and household saving significantly reduces fertilizer applied per acre where farmers save their money and practice irrigation a possible explanation could be that there could be no or low household savings where farmers irrigate their fertilized fields because irrigation requires higher amounts of fertilizer application due to leaching of the nutrients. There is an increasing trend of the coefficients of the land size and timeline we move from the lower quantile level to the high quantile level, but the estimates are significant at the 0.50th quantile at 5% level of significance suggesting that land size and timeline interaction effect has a contribution to variations in amount of fertilizer use per acre. The interaction between the number of plots and the timeline show significant reduction in the levels of the amount of fertilizer applied per acre.

Table 4.7: Quantile regression summary of results (p-values in parentheses)

| Variable | Parameter estimates | | |
|---|---|---|---|
| | 25th percentile | Median | 75th percentile |
| Intercept | 14.0871 (0.5882) | 26.0503 (0.0333) | 20.0896 (0.4662) |
| District | -10.2736 (0.1108) | -3.6636 (0.5482) | -7.1929 (0.4442) |
| Head | 0.8285 (0.7631) | 3.2359 (0.4809) | 9.1333 (0.2249) |
| Land | -0.8348 (0.9739) | 4.3232 (0.5962) | 0.8075 (0.9643) |
| Training | -0.9884 (0.6210) | -4.4776 (0.6075) | -14.0116 (0.4772) |
| Source | -0.8230 (0.7139) | 1.3004 (0.6981) | 3.8158 (0.6132) |
| Saving | -1.0401 (0.8781) | 10.7812 (0.4233) | 28.0760 (0.2490) |
| Irrigation | 0.6943 (0.8006) | 4.1189 (0.5910) | 15.3133 (0.1035) |
| Anlmanure | 0.5288 (0.8126) | 0.1188 (0.9749) | -1.1701 (0.8600) |
| Plots | 14.3027 (0.0033) | 9.7231 (0.0223) | 12.4887 (0.0017) |
| Nogrow | 0.9653 (0.7112) | 0.1579 (0.9703) | -13.0140 (0.0806) |
| Distance | -0.0390 (0.7183) | -0.0039 (0.9909) | 0.3206 (0.5213) |
| Lenglast | -0.0291 (0.9325) | -0.0050 (0.9938) | 1.6894 (0.1742) |
| Total income | 0.0000 (0.8777) | -0.0000 (0.9152) | 0.0000 (0.8225) |
| Fertgood | -0.8096 (0.7376) | 0.1038 (0.9834) | 10.3087 (0.1757) |
| Fert_perc | -0.9403 (0.8494) | -1.5925 (0.7620) | -0.9710 (0.9352) |
| Timeline | -9.3624 (0.7152) | -15.2230 (0.0893) | -4.1555 (0.8424) |
| Land*Training | -6.6972 (0.2422) | -8.2525 (0.2648) | -4.9642 (0.7184) |
| Land*Saving | -1.5925 (0.8062) | -8.6163 (0.2289) | -16.6847 (0.1185) |
| Training*Saving | 11.0767 (0.8062) | 16.5126 (0.1612) | 28.0778 (0.1643) |
| Saving*Irrigation | -10.5505 (0.0656) | -21.0327 (0.0376) | -42.5716 (0.0059) |
| Distance*Total income | 0.0000 (0.3196) | 0.0000 (0.1954) | 0.0000 (0.5165) |
| Land*Timeline | 14.7244 (0.5430) | 18.2812 (0.0124) | 22.8157 (0.2264) |
| Plots*Timeline | -16.9254 (0.0014) | -15.8796 (0.0004) | -22.8438 (0.0182) |

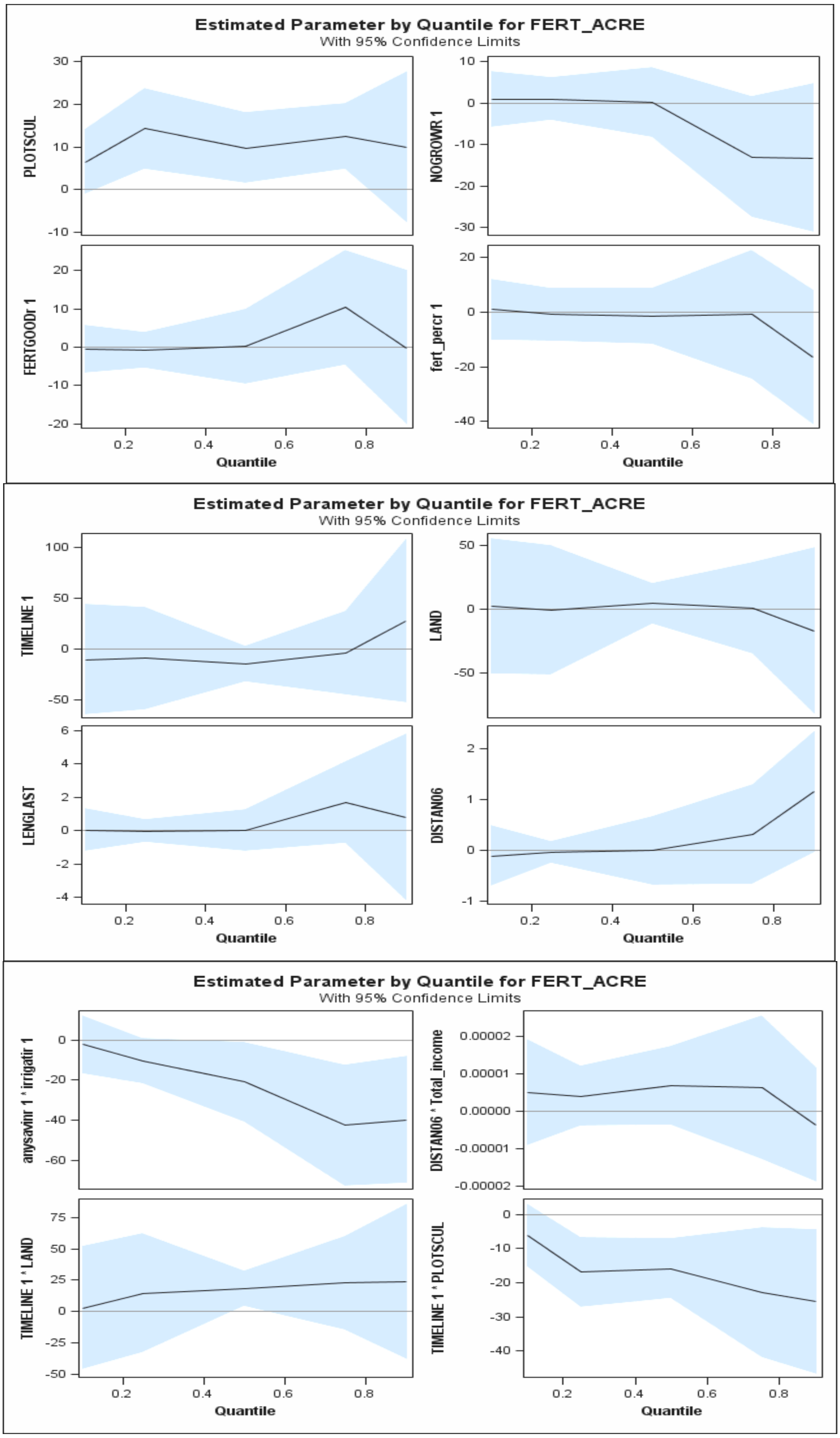

Figure 4.12: Quantile plots for the estimated parameters in the model with their 95% confidence bands

# Chapter 5

# CONCLUSION

The main objective of the thesis was to investigate and model factors that affect the amount of fertilizer usage per acre under maize cultivation in Lilongwe and Kasungu Districts of Malawi. These two districts benefited from the Rural Livelihood Diversification project (RLDP) which was aimed at improvement of fertilizer management strategies, accessibility and proper utilisation of the fertilizer by the smallholder farmers. In line with the thesis objective several statistical procedures were conducted which included modelling the mean amount of fertilizer use per acre with first order interaction of the explanatory variables based on general linear model; transformed response general linear model; gamma model; and modelling amount of fertilizer use per acre based on quantile regression. The recognized explanatory variables would aid the government, researchers, non governmental organisations, the private sector and other stakeholders on key areas that need more focus and support in order to improve the levels of fertilizer that small holder farmers apply in their fields in order to boost agricultural production at household level as well as national level.

A two stage cluster sampling technique was used to sample households. Since some sampled households were interviewed at baseline research study in June 2007 and impact assessment study which was conducted in July 2008, the assumption was that there could be a correlation of the results and therefore these households were excluded in the analysis. Before analysis, data were checked for incorrect data values and extreme values. Exploratory data analysis was conducted and the results indicated that most of the households that were sampled came from Kasungu district. Most of the sampled households were not trained in fertilizer management and use, do not apply animal manure in their gardens and had negative perceptions on the use of fertilizer in their gardens.

A general linear model with continuous amount of fertilizer use per acre as the response, and a set of fixed explanatory variables was used to model the data. There were many explanatory variables for the model building therefore; automatic variable subset selection procedures using the GLMSELECT procedure in SAS were done in order to reduce the number of the explanatory variables in the model. Analysis of the residuals and influential observations of the given models was done in order to check if model assumptions were not violated and validate if the models were adequate by use of various diagnostic tools. As discussed earlier, the distribution of the residuals under the general linear model was not normal and there was non constant variance, therefore log transformation of the response variable was explored but still there were violations of the model assumptions of constant variance and normality. Due to failure of the general linear model to model the data because of violation of normality assumption, generalized linear modelling assuming a gamma distribution of the response variable was further employed on the data using the log link and its alternative inverse link. The results indicate that gamma modelling assuming inverse link performed better than the log link option, but both the models were affected by the influential observation making the models invalid.

As a possible robust alternative of mean regression to model a continuous response amount of fertilizer use per acre given a set of predictor variable, quantile regression model was considered using the QUANTREG procedure in SAS. Quantile regression was used to assess the changes in the distribution of fertilizer per acre given a set of predictors at the 0.25th, 0.50th and 0.75th quantiles. Quantile regression modelling is essential because it is robust against outliers and violations of the distributional assumptions. Since quantile regression is not robust to high leverage points, analysis of leverage points was done and no high leverage points were detected to affect the model.

The differences in the results of the mean regression and quantile regression could be an indication that the distribution of the response variable given a set of explanatory variables was asymmetric hence relying on the results from the mean regression could be questionable. The application of quantile regression in modelling amount of fertilizer use per acre resulted in making richer inferences as there was analysis based at different levels of amount of fertilizer use per acre distribution rather than the median, which enabled the investigator to capture the trend on the results of the lower 25th percentile and upper 75th percentile. Quantile regression provided a vessel of flexibility amongst the fertilizer usage determinants, as they had different impacts at different levels of the response amount of fertilizer use per acre. This is shown in having the variations of the results at different quantile levels.

The quantile regression results indicates that the effect of household saving on the amount of fertilizer use per acre varies at different levels of irrigation and the results indicate significant high reduction on the amount of fertilizer applied amongst the households who apply the highest amount of fertilizer per acre. These results imply that intervention aimed at increasing the amount of fertilizer applied per acre in irrigated fields should also consider improving on the amounts of household savings, which confirms our hypothesis that farmers who save their money have a higher likelihood of purchasing more fertilizer even when there are price changes of the fertilizer commodity. Therefore there is a need to strengthen interventions that could aim at improving smallholder savings either through formal banks or informal village savings banks. Policies that aim at diversifying smallholder farmers' sources of income for them to save such as income generating activities, would ensure adequate amount of money amongst the smallholder farmers for purchasing inputs such as fertilizer.

The significant coefficient of the land size and timeline interaction suggests that increase in land size is critical in influencing high amount of fertilizer application. The results confirms with (Chirwa, 2003), that increase in land size significantly increases the land sizes. Policies aimed at increasing access to cultivable land should be strengthened in order to increase maize productivity there by ensuring food security and increase in household incomes. There is need to intensify rehabilitation of degraded cultivable land as one of the initiatives to increase access of agricultural land.

The results reveal that the increase in number of plots significantly reduces the amount of fertilizer use per acre towards the end of the project during impact assessment study. This could be the case because increase in number of different fragmented plots that smallholder farmers cultivate on could have different soil types and the fertility could decline with time, requiring critical decision on the levels of fertilizer to be applied in order to achieve more crop productivity. The findings of our study also suggest that, there is a need to do more evaluation of the soils in different plots where smallholder farmers cultivate their crops, as this would enable researchers to come up with critical informed recommendations for the suitable fertilizer types that ought to be applied to specific soil types. The soil analysis results from specific areas would enable farmers apply appropriate amounts of fertilizer to the soil which would in turn ensure availability of adequate amounts of essential nutrients for maize crop growth and development.

Lower levels of fertilizer application could have negative consequences amongst the smallholder farmers who could be at risk of low maize productivity, thereby understanding the factors affecting low fertilizer application could assist researchers, policy makers to develop interventions based on evidence from research results and inequalities in terms of households applying lower levels of amount of fertilizer per acre and those that apply higher amounts of fertilizer per acre. Analysis of our data based on quantile regression implies that policies and programmes should consider factors that affect the farmers at risk, especially the farmers that apply lower levels of fertilizer. Application of other soil fertility practices such as intercropping of maize with leguminous crops, conservation agriculture practices, and agroforestry, would enhance availability of fertilizer to the crops thereby enhancing crop nutrient uptake as well as crop productivity where inorganic fertilizer use is limited.

As observed in modelling amount of fertilizer use per acre using various models, it is critical to know the behaviour of the response variable whether it is continuous or not before choosing any model. Also consideration should be made on whether the results are correlated or not, as observed in this study that some households were interviewed at both baseline and impact assessment studies. It is also essential to carry out model diagnostics before conclusions can be drawn using a given model to avoid violating model assumptions which could lender the model inappropriate.

Several weaknesses were discovered in the sample selection procedures, data collection and handling. In data collection it was found that there was no clear definition of the interviewee in a given household. Therefore it is necessary in future research that there should be a clear definition on the mode of selection of the interviewee in a sampled household, whether the household head is interviewed or the spouse or the children, as failure to define the interviewee could lead to, or introduce sampling bias as there could be an overrepresentation or under representation of a specific gender being interviewed. It is also recommended that before data collection, there should be a clear definition of the research hypothesis so that analytical procedures are identified. The other assumption in the project was that all the sampled households would participate in the project; therefore there is a need to clarify whether or not there were dropouts during project implementation. This could also assist in selection of a representative sample of households who were involved in the project.

In this study we used data collected on the households who were involved in the Rural Livelihood Diversification Project, it is tricky to draw conclusions on whether the project was successful or not due to the short period which elapsed between June 2007 and July 2008 when the baseline data and impact assessment data was collected respectively. The other challenge to define the success of the Rural Livelihood Diversification Project is that the baseline study was conducted during the implementation of the project, it could be proper to conduct baseline research survey before the project was implemented.

In order to have reliable data, there is a need to check the questionnaires soon after field collection in order to minimise recording errors, missing and extreme observations. Missing data could have a severe effect on the model estimates as the cases with missing values were dropped from the analysis, resulting in a reduction of the data set that was used for modelling leading to bias in the parameter estimates that were generated. It is a requirement therefore to take precautions when collecting and handling data so as to get accurate and reliable estimates for better interpretation and generalization of the results to the population from which the sample was selected. The shortfall of household interviews on amount of fertilizer use per acre data collection was that most of the data collected relied on recalling sampled households' memories which could compromise the results as improper or wrong records could be collected and used. Improvements should be made in collecting such kinds of data by having data collected from the household's field as well, such as physical measurements of the garden where fertilizers were applied, as in most cases fertilizer is inadequate and could not be applied to the whole garden. Improvements could also be made by having data recording sheets for each of the sampled households for record keeping. In order to assess the goal of the project on whether the intervention was a success or not, there was a need to repeat the survey at impact assessment using the original sample (Yates, 1981), or by sampling from the original sample. This would enable the investigator to obtain accurate changes on the amount of fertilizer use per acre thereby enabling the investigator to assess whether the project had achieved its goals or not.

The analysis of the data was not exhaustive as there was no clear definition on whether the household used the basal fertilizer or top dressing fertilizer, hence analysis based on aggregated total amount of fertilizer use per acre may be prone to errors and therefore be problematic and unsatisfactory. The analysis only considered a maize crop because the data provided adequate information to analyse amount of fertilizer use per acre based on a maize crop. Future research should consider modelling of the resultant maize crop yields in order to justify increased levels of the amount of fertilizer use per acre.

It is advantageous to use classical and robust methods in the analysis of the amount of fertilizer use per acre data to ensure results that are meaningful and reliable, as there could be model failure due to violations of some model assumptions. Future research should consider time series model analysis to model the trend in amount of fertilizer use per acre given data were collected at several times across the project implementation.

The results from this study employed modelling the amount of fertilizer use amongst farming households given a set of predictors, using quantile regression modelling and this enabled the researcher to capture more on the extremes i.e. factors that affect fertilizer application among households that apply lower levels of fertilizer as well as those that apply higher levels of fertilizers rather than relying on the factors affecting the average households. The results of the study also indicate that there is a need for an integrated approach in handling issues and determinants aimed at increasing levels of fertilizer application per unit area that small holder farmers apply to their fields than in tackling the determining factors singly. The findings from this study have positive implications in suggesting to the government, researchers, non governmental organisations, private sector and other stakeholders that there is need for an integrated approach, rather than tackling the factors singly, on planning and implementation of the interventions aimed at increasing the amount of fertilizer application among small holder farmers in order to enhance the levels of amount of fertilizer application amongst the smallholder farmers.

# References

Austin, P. C., Tu, J.V., Daly, P.A., A. Alter, D.A. (2005). The use of quantile regression in health care research: a case study examining gender differences in the timeliness of thrombolytic therapy. *Statistics in Medicine* **24,** 791-916.

Barrodale, I., and Roberts, F. D. K. (1973). An Improved Algorithm for Discrete l1 Linear Approximation. *SIAM Journal on Numerical Analysis* **10,** 839-848.

Bationo, A. (2009). *Constraints and new opportunities for achieving a green revolution in Sub-Saharan Africa through integrated Soil Fertility Management.* UC Davis: The proceedings of the International plant nutrition Colloquium XVI. Retrieved from http://escholarship.org/uc/item/7hr282j2.

Belsley, D. A., Kuh, E., and Welsch R.E. (2004). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley and Sons.

Box, G. E. P., and Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* **26,** 211-252.

Cade, B. S., and Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment* **1,** 412-420.

Chirwa, E. W. (2003). Fertilizer and hybrid seed adoption among small holder maize farmers in Southern Malawi. *Wadonda consult working paper* **WC/02/03**.

Cohen, R. A. (2006). Introducing the GLMSELECT PROCEDURE for Model Selection. *SAS Users Group International (SUGI031).*

Collet, D. (2003). *Modelling binary data*. London: Chapman and Hall.

Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. *Technometrics* **19,** 15-18.

Denning, G., Kabambe, P., Sanchez, P., Malik A., Flor R., Harawa, R., Nkhoma, P., Zamba,C., Banda, C., Magombo, C., Keating, M., Wangila, J., Sachs, J. (2009). Input subsidies to improve smallholder maize productivity in Malawi Toward an African Green Revolution. . *PLoS Biology* **7(1)**. Available at www.plosbiology.org/article/info:doi/10.1371/journal.pbio.1000023

Dobson, A. J. (1983). *Introduction to statistical Modelling*. London: Chapman and Hall.

Draper and Smith, N. R., Smith, H. . (1981). *Applied Regression Analysis*. New York: Wiley.

Everitt, B. S. (1987). *Introduction to optimization methods and their application in statistics*. London: Chapman and hall.

FAO. (2009). Pathways to Success: Success Stories in Agricultural Production and Food security. Rome: Food and Agriculture Organization.

Fufa, B., and Hassan, R. M. (2006). Determinants of fertilizer use on maize in Eastern Ethiopia: A weighted indogenous sampling analysis of the extent and intensity of adoption. *Agrekon* **45,** 38-49.

Government of Malawi. (2002). Malawi National land policy. Lilongwe, Malawi.

Government of Malawi. (2005a). *Integrated household survey 2004-2005*. Zomba: National Statistics Office.

Government of Malawi. (2005b). Malawi Demographic and Health Survey 2004 In National Statistics Office (Ed.). Zomba, Malawi.

Government of Malawi. (2006a). Food security policy. Lilongwe, Malawi.

Government of Malawi. (2006b). Malawi Growth and Development Strategy 2006-2011. Lilongwe. Malawi.

Government of Malawi. (2007). *National Fertilizer Policy*. Lilongwe, Malawi: Ministry of Agriculture and Food Security.

Green, P. J. (1984). Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives. *Journal of the Royal Statistical Society. Series B (Methodological)* **46,** 149-192.

Hampel, F. R. (1986). *Robust statistics the approach based on influence functions*. New York: John Wiley and Sons.

Heisey, P. W., and Mwangi, W. (1996.). Fertilizer Use and Maize Production in Sub-SaharanAfrica, *CIMMYT Economics Working Paper 96-01*. Mexico, D.F: CIMMYT.

Heisey, P. W., and Smale, M. (1995). *Maize Technology in Malawi A Green Revolution in the Making?* CIMMYT Research Report. Mexico, D.F: CIMMYT.

Hinkley, D. V., Reid, N., and Snell, E.J. eds. (1991). *Statistical theory and modelling. In honour of Sir David Cox, FRS.* London: Chapman and Hall.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W., (eds). (1983). *Understanding robust and exploratory data analysis*. Newyork: John Wiley and Sons.

Hoaglin, D. C., and Welsch, R. E. (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician* **32,** 17-22.

Hubert M., Rousseeuw, P.J., Aelst, S.V. (2008). High-Breakdown Robust Multivariate Methods. *Statistical Science* **23** 92-119.

Jorgensen, B. (1983). Maximum Likelihood Estimation and Large-Sample Inference for Generalized Linear and Nonlinear Regression Models. *Biometrika* **70,** 19-28.

Kherallah, M., Minot, N., Kachule, R., Soule, B., Berry, B. (2002). Impact of Agricultural Market Reforms on Smallholder Farmers in Benin and Malawi, Washington, DC: International Food Policy Research Institute.

Koenker, R. (2005). *Quantile Regression*. New York: Cambridge University Press.

Konker, R., and Basset, G. (1978). Regression quantiles. *Econometrica* **46,** 33-50.

Koenker, R., and Bassett, G. (1982). Tests of Linear Hypotheses and l"1 Estimation. *Econometrica* **50,** 1577-1583.

Koenker, R., and Hallock, K. F. (2001). Quantile Regression. *The Journal of Economic Perspectives* **15,** 143-156.

Koenker, R., and Machado, J. A. F. (1999). Goodness of fit and related inference process for quantile regression. *Journal of American Statistical Association* **94,** 1296-1310.

Konker, R., and Basset, G. (1978). Regression quantiles. *Econometrica* **46,** 33-50.

Krzanowski, W. J. (1998). *An introduction to statistical modelling.* London: Arnold.

Kumwenda, J. D. T., Waddington, S.R., Snapp, S.S., Jones, R.B., and Blackie, M.J. (1996). Soil fertility management research for the maize cropping systems by smallholders in Southern Africa, *A review. NRG paper 96-02.* Mexico D.F: *CIMMYT.*

Mafongoya, P. L., Bationo, A., Kihara J., Waswa, B. (2006). Appropriate technologies to replenish soil fertility in southern Africa. *Nutr Cycl Agroecosyst* **76,** 137-151.

Mangisoni, J. H. (2006). Impact of Treadle Pump Irrigation Technology on Smallholder Poverty and Food Security in Malawi: A Case Study of Blantyre and Mchinji Districts. Lilongwe, Malawi: University of Malawi.

McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models* (2 ed.). London: Chapman and Hall.

Minde, I. Jayne, T., Crawford, E., Ariga, J., Govereh, J. (2008). Promoting Fertilizer Use in Africa: Current Issues and Empirical Evidence from Malawi, Zambia and Kenya. *Working paper for ReSAKSS/Southern Africa.* Pretoria. South Africa.

Minot, N., Kherallah, M., and Berry, P. (2000). *Fertilizer market reform and the determinants of fertilizer use in Benin and Malawi*. MSSD Discussion Paper No. 40. Washington, DC: IFPRI.

Montgomery, D. C., Peck, E.A., Vinning, G.G. (2006). *Introduction to Linear Regression Analysis.* New York: John Wiley and Sons.

Morris, M., Kelly, V.A., Kopicki, R.J. and D. Byerlee (2007). Fertilizer Use in African Agriculture: Lessons Learned and Good Practice Guidelines. Washington DC: World Bank.

Myers, R. H., Montgomery, D. C., and Vining, G. G. (2002). *Generalized Linear Models With Applications in Engineering and the Sciences*. New York: John Wiley and Sons.

Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society. Series A (General)* **135,** 370-384.

Neter, J., Wasserman, W., and Kutner, M. H. (1990). *Applied Linear Statistical Models Regression, Analysis of Variance, and Experimental Designs* (3rd. ed.). Burr Ridge, USA: IRWIN.

O'Brien, L. G. (1983). Generalised Linear Modelling Using the GLIM System. *Area* **15,** 327-336.

Pregibon, D. (1980). Goodness of link test for generalized linear models. *Applied Statistics* **29,** 15-24.

Puterman, M. L. (1988). Leverage and Influence in Autocorrelated Regression Models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **37,** 76-86.

Rousseeuw P.J. and Driessen K.V. (1999). A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics* **41** 212-223.

Sakia, R. M. (1992). The Box-Cox Transformation Technique: A Review. *The statistician* **41,** 169-178.

SAS. (2004). SAS 9.1.3. U.S.A.: Cary, NC.

Scheaffer, R. L., Mendenhall, W., and Ott, L. (1986). *Elementary survey sampling.* Boston: PWS-Kent.

Smale, M., and Phiri, A. with contributions from Chikafa, G.A., Heisey, P. W., Mahatta, F., Msowoya, M.N.S., Mwanyongo, E.B.K., Sagawa, H.G., and

Selemani, H.A.C. (1998). *Institutional Change and Discontinuities in Farmers' Use of Hybrid Maize Seed and Fertilizer in Malawi Findings from the 1996-97 CIMMYT/MoALD Survey*. Economics Working Paper 98-01. Mexico, D.F.: CIMMYT.

Thisted, R. A. (1988). *Elements of Statistical Computing: Numerical Computation*. London: Chapman and hall.

Vittinghoff, E., Glidden, D.V., Shiboski S.C., and McCulloch C.E. (2005). *Regression methods in Biostatistics Linear, Logistic, Survival, and Repeated Measures Models*. New York: Springer.

Yates, F. (1981). *Sampling Methods for Censuses and Surveys*. London: Griffin.

Yohai, V. J. (1987). High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics* **15,** 642-656.

Zhou, K. Q., and Portnoy, S. L. (1996). Direct Use of Regression Quantiles to Construct Confidence Sets in Linear Models. *The Annals of Statistics* **24,** 287-306.

# Appendices

## Appendix A

### GLMSELECT procedure

```
ods html;
ods graphics on;
proc glmselect data=nocomparisonconti plots=all;
class district head training source saving irrigation anlmanure nogrow fertgood fert_perc timeline;
model fert_acre=district|head|land|training|source|anysaving|irrigation|anlmanure|plots|nogrow|distance|lenglast|total_income|fertgood|fert_perc|timeline@2/
selection=stepwise(select=sL SLE=.1 SLS=.1 choose=adjrsq)   showpvalues stats=all;
output out=errors resid=resid predicted=predicted;
run;
ods graphics off;
ods html close;
proc gplot data=errors;
  plot resid*predicted;
  run;
proc univariate data=errors;
var resid;
histogram;
probplot;
run;
```

# Appendix B

## Calculating pseudo R-squared at Quantile=0.25

```
proc quantreg data=nocomparisonconti;
MODEL FERT_ACRE= /quantile=0.25; *intercept only model;
output out=intmodel res=resint;
run;
proc quantreg data=intmodel; *Bring in saved data to run full model;
class district head training source saving irrigation anlmanure nogrow fertgood
fert_perc timeline;
MODEL FERT_ACRE=district head training source saving irrigation anlmanure
plots nogrowr fertgood fert_perc timeline land lenglast distance total_income
land*training land*saving training*saving saving*irrigation distance*total_income
land*timeline plots*timeline/quantile=0.25; *full model;
output out=fullmodel res=resfull; *save these residuals too;
run;
data _null_;
set fullmodel end=lastrow; *Bring in the results of the previous run;
sresfull+abs(resfull); *sum up the absolute values of both residuals;
sresint+abs(resint);
if lastrow then do;
pseudoR2=(sresint-sresfull)/sresint;
put pseudoR2=;
end;
run;
```

# Appendix C

## Questionnaire for data collection

**Baseline Household Survey Questionnaire**

Household's name: ______________________ Sex: 1=Female 2=Male
Relationship to household head: 1=Wife 2=Husband
District: 1. Lilongwe 2. Kasungu
Village__________________________ _

**NB Do not leave any blank spaces. Indicate No, None where appropriate and NA where the answer is not applicable. Circle the appropriate responses**

**CROP AND LIVESTOCK PRODUCTION**

1a) How much land do you have…………………………….. Acres

1b) How many gardens of land do you have? ………………………

1c). How many gardens did you cultivate in the past growing season (2006/2007)?

| Garden | Area (Acres) | How far is it from the house in km? | Location ***1=Upland 2=Dambo*** | What is the type of soil ***1=Clay, 2=Sandy, 3=Loam, 4=Black cotton*** | What is your perception of the fertility of the soil in the plot? ***1=Low, 2=Medium, 3=High*** | What shows that the fertility is like that? (what is the indicators for the level of fertility) | What do you think is the cause of the fertility status? |
|---|---|---|---|---|---|---|---|
| 1 | | | | | | | |
| 2 | | | | | | | |
| 3 | | | | | | | |
| 4 | | | | | | | |
| 5 | | | | | | | |
| | | | | | | | |
| | | | | | | | |

1e) Did you have land which you did not use in the last growing season? …………………..**1=Yes; 0=No**

1f) If yes, why didn't you cultivate the land? ........................

***1=not enough seed, 2=not enough other input, 3=not enough labour, 4=Left fallow land, 5=lack of funds***
***6=others (specify***

2a. What are the most common varieties of crops you have been growing in the last 2 years?

| Crop | Varieties grown | How long have you used the variety? **(months/yrs)** | Source of seed/planting material ***(see codes below***) | Area under variety **(2006/2007) acres** | Amount harvested (**2006/2007) (number of 50kg bags**) | Area under variety **(2005/2006) acres** | Amount harvested (2005/2006) **(number of 50kg bags**) |
|---|---|---|---|---|---|---|---|
| Cassava | 1. | | | | | | |
| | 2. | | | | | | |
| | 3. | | | | | | |
| Groundnuts | 1. | | | | | | |
| | 2. | | | | | | |
| | 3. | | | | | | |
| Beans | 1. | | | | | | |
| | 2. | | | | | | |
| | 3. | | | | | | |
| Pigeon Peas | 1 | | | | | | |
| | 2 | | | | | | |
| | 3 | | | | | | |
| Soya beans | 1 | | | | | | |
| | 2 | | | | | | |
| | 3 | | | | | | |
| Bananas | 1 | | | | | | |
| | 2 | | | | | | |
| | 3 | | | | | | |
| Maize | 1 | | | | | | |
| | 2 | | | | | | |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |  |
|  | 3 |  |  |  |  |  |  |
| Tomatoes | 1 |  |  |  |  |  |  |
|  | 2 |  |  |  |  |  |  |
|  | 3 |  |  |  |  |  |  |
| Onions | 1 |  |  |  |  |  |  |
|  | 2 |  |  |  |  |  |  |
|  | 3 |  |  |  |  |  |  |
| Paprika | 1 |  |  |  |  |  |  |
|  | 2 |  |  |  |  |  |  |
|  | 3 |  |  |  |  |  |  |

***Codes for source of planting material 1=home saved; 2=bought from other farmers; 3=bought from market; 4=bought from trained seed producers; 5=borrowed/exchanged/given; 6=government extension; 7=NGOs; 8=research; 9=purchased from stockists; 10=others (specify)…***

2b. If seed/ planting material was purchased from stockists, how far are the stockists from your home?

| Type of seed/ planting material | Stockists where purchased<br>***1= local agro-dealers based in village or trading center 2=others (Chipiku, Seed Co., big companies 3= Other (please specify)*** | Distance from home (***KM)*** |
|---|---|---|
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

3a) Are you currently growing any crops under irrigation? **1=Yes 0=No**

| Crop | Area Planted in 2005 (Acres) | Type of irrigation (**see codes below)** | Amount produced 2005(kg) | Main use ***1=Mainly Food , 2=Mainly cash, 3=Both food and cash*** | If, sold, amount of money made 2005 MK | Area planted 2006(Acres) | Amount produced 2006(kg) | Main use ***1=Mainly Food , 2=Mainly cash, 3=Both food and cas*h** | If, sold, amount of money made 2006 MK |
|---|---|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  |  |  |  |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

**Codes for type of irrigation 1 =Treadle pump, 2 = Engine pump, 3 = Drip irrigation 4 = Gravity 5= Watering can**

3b) If you have irrigation equipment, what was the source?
***1=Purchase 2=Given by NGO 3=Given by government***

3c) If purchased? How much did you purchase it for? ..............................................Mk

3d) What other initial investments did you make on the irrigation system?

| Investments | Amount MK | If used family labour, how many Persondays |
|---|---|---|
| | | |
| | | |
| | | |
| | | |

4). Who is involved in the management activities of the crops cultivated? ***For all columns use codes below table***

| Crop *(include other crops from list above)* | Land preparation | Planting | Weeding | Spraying | Harvesting | Post harvest handling | Transporting to markets | Actual marketing | |
|---|---|---|---|---|---|---|---|---|---|
| Maize | | | | | | | | | |
| Soya beans | | | | | | | | | |
| Beans | | | | | | | | | |
| Groundnuts | | | | | | | | | |
| Cassava | | | | | | | | | |
| Bananas | | | | | | | | | |
| Tomatoes | | | | | | | | | |
| Onions | | | | | | | | | |
| Paprika | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |

***Codes 1=Husband only; 2=Wife only; 3=Husband mostly; 4=Wife mostly; 5=Husband and wife equally; 6=Children; 7=Hired labour; 8=Other (specify)……***

5). Who makes the following decisions? **For all columns use codes below table**

| Crop | Where to plant? | What area to plant? | What inputs to apply? | How to use inputs? | How much to sell? | When and where to sell? | How to use money from sale? |
|---|---|---|---|---|---|---|---|
| Maize | | | | | | | |
| Soya beans | | | | | | | |
| Beans | | | | | | | |
| Groundnuts | | | | | | | |
| Cassava | | | | | | | |
| Bananas | | | | | | | |
| Tomatoes | | | | | | | |
| Onions | | | | | | | |
| Paprika | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

***Codes 1=Husband only; 2=Wife only; 3=Husband mostly; 4=Wife mostly; 5=Husband and wife equally; 6=Children; 7=Hired labour; 8=Other (specify)……***

6a).During the last growing season, did you hire or pay laborers to work on your farm? **1=Yes; 0=N0**

6b)If yes, what activities did they carry out and how much did you spend?

| Crop | No. of labourers | Activities | Amount paid per person per day MK | Amount paid in kind (state the MK value of things given) |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |

7. Do you own any of the following livestock?

| Livestock | Number | | Who does the following? **Use codes below** | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Local | Improved | Feeding/ Grazing | House / khola construction | Watering | Selling | Keeping money from sale | Making decisions on use of money |
| Cattle | | | | | | | | |
| Poultry | | | | | | | | |
| Goats | | | | | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Pigs | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |

**Codes for who does the following 1 = Male adult, 2 = Female Adult, 3 = Male Child, 4 = Female Child, 5= Hired labour**

**KNOWLEDGE AND USE OF SOIL FERTILITY MANAGEMENT PRACTICES**

8). How often do you use the following soil fertility improvement measures and on what crops? Fill in table below

| Soil fertility measure | Do you know it? **1= Yes 0= No** | Have you ever used? **1= Yes 0= No** | If yes, How often do you use it? **1 = Every season, 2=Once very two seasons 3= Infrequently** | On what crops? | From where did you get information / knowledge on the technology<br>***See codes below*** | What problems /challenges do you encounter when you use? |
|---|---|---|---|---|---|---|
| Crop rotation | | | | | | |
| Incorporate crop residue | | | | | | |
| Animal manure | | | | | | |
| Plant agro forestry trees or shrubs | | | | | | |
| Resting land (fallow) | | | | | | |
| Use farm yard manure | | | | | | |
| Controlling soil erosion | | | | | | |
| Early ploughing | | | | | | |
| Botanicals | | | | | | |
| Cover crops | | | | | | |
| Others (specify)… | | | | | | |
| | | | | | | |

**Codes for source of information 1 = Extension worker 2=From radio 3= From Newspapers/brochures 4=Other (Specify)**

8b) If you have used Manure or other organic fertilizers on your crops, state the amounts

| Organic soil Fertility Measure | Crops used on | Amounts applied |
|---|---|---|
| Animal Manure | 1 | 1 |
| | 2 | 2 |
| | 3 | 3 |
| Backyard Manure | 1 | 1 |
| | 2 | 2 |
| | 3 | 3 |
| Compost Manure | 1 | 1 |
| | 2 | 2 |
| | 3 | 3 |
| Other (specify) | 1 | 1 |
| | 2 | 2 |
| | 3 | 3 |

9a) Farmer knowledge of fertilizers and fertilizer use
What do you think of the following statements about fertilizers? Are they true or false?-*Tick appropriate box*

| Statement on Fertilizers | True | False | I do not know / No opinion |
|---|---|---|---|
| Fertilizers are not good for the soil as they destroy the soil | | | |
| A fertilizer for tobacco is also good for maize | | | |
| The same fertilizer for planting should be used for top dressing | | | |
| All crops should be applied the same rate of fertilizers | | | |
| If I have a little fertilizer and a large area, it is better to spread it all over rather than to concentrate on a small area | | | |
| Only maize and tobacco should be applied fertilizer, other crops do not require fertilizer | | | |
| Different fertilizers contain different nutrients and should be used for different crops and for different purposes | | | |

9b). How would you rate yourself in the following aspects –*Tick appropriate box*

| Knowledge or practice | Good | Average | Poor |
|---|---|---|---|
| Knowledge of what fertilizer to use for different crops | | | |
| Knowledge how much fertilizer to apply for different crops | | | |
| Knowledge of which fertilizer to use for planting, and for top dressing | | | |

9c). Please give the following information for Fertilizer use for the 2006/07 season

| Crop | Did you use fertilizer on this crop? *1-Yes 0=No* | Area under crop on which fertilizer was applied (acre) | Type of fertilizer applied | When applied *1=Before planting 2=At planting 3=Top dressing* | Amount applied (Quantity in kg | How much did you pay for it? MK | Source of fertilizer (see codes) | What is the distance to source? KM |
|---|---|---|---|---|---|---|---|---|
| Maize | | | 1. | | | | | |
| | | | 2. | | | | | |
| | | | 3 | | | | | |
| Groundnuts | | | 1. | | | | | |
| | | | 2. | | | | | |
| | | | 3. | | | | | |
| Cassava | | | 1. | | | | | |
| | | | 2. | | | | | |
| | | | 3. | | | | | |
| Soya beans | | | 1. | | | | | |
| | | | 2. | | | | | |
| | | | 3. | | | | | |
| Beans | | | 1 | | | | | |
| | | | 2 | | | | | |
| | | | 3 | | | | | |
| Tomatoes | | | 1 | | | | | |
| | | | 2 | | | | | |
| | | | 3 | | | | | |
| Onions | | | 1 | | | | | |
| | | | 2 | | | | | |
| | | | 3 | | | | | |

***Source of fertilizers 1=purchased from market; 2=purchased from stockists; 3=purchased from other farmers; 4purchased subsidized from government (coupons); 5=received from NGOs; 6=others (specify)…***

9d. Fertilizer use for the previous growing season 2005/06 season

| Crop | Did you use fertilizer on this crop? *1-Yes 0=No* | Area under crop on which fertilizer was applied (acre) | Type of fertilizer applied | When applied *1=Before planting 2=At planting 3=Top dressing* | Amount applied | How much did you pay for it? **MK** | Where did you get fertilizer (see codes below) | What is the distance to source? **KM** | Who made decision to purchase (***1-Husband, 2=Wife, 3=Both)*** |
|---|---|---|---|---|---|---|---|---|---|
| Maize | | | 1. | | | | | | |
| | | | 2. | | | | | | |
| | | | 3 | | | | | | |
| Groundnuts | | | 1. | | | | | | |
| | | | 2. | | | | | | |
| | | | 3. | | | | | | |
| Cassava | | | 1. | | | | | | |
| | | | 2. | | | | | | |
| | | | 3. | | | | | | |
| Soya beans | | | 1. | | | | | | |
| | | | 2. | | | | | | |
| | | | 3. | | | | | | |
| Beans | | | 1 | | | | | | |
| | | | 2 | | | | | | |
| | | | 3 | | | | | | |
| Tomatoes | | | 1 | | | | | | |
| | | | 2 | | | | | | |
| | | | 3 | | | | | | |
| Onions | | | | | | | | | |

***Source of fertilizers 1=purchased from market; 2=purchased from stockists; 3=purchased from other farmers; 4purchased subsidized from government (coupons); 5=received from NGOs; 6=others (specify)…***

9e. What constraints do you face in accessing fertilizer in order of priority and what do you think should be done to address these constraints?

| Constraints in accessing fertilizers | Proposed interventions to address constraint |
|---|---|
| 1. | |
| 2. | |
| 3. | |
| 4. | |

9f). What constraints do you face in using / utilizing fertilizers in order of priority and what do you think should be done to address these constraints?

| Constraints in using / utilizing fertilizers | Proposed interventions to address constraint |
|---|---|
| 1. | |
| 2. | |
| 3. | |
| 4. | |

**MARKETS AND ENTERPRISES**

10. For what purpose did you grow the following crops in the past growing season (2005/06)?

| Crop | Purpose **1=Mainly food 2=Mainly cash 3=Both** | Total harvest (kg) | Quantity consumed (kg) | Quantity sold (kg) | Where sold? | Price per unit (no of 50kg bags) | Total amount earned MK | Who makes decisions on the use of the money? **1= husband only, 2= wife only , 3= both wife and husband** | Rank the most important in terms of income generation (1 most important) |
|---|---|---|---|---|---|---|---|---|---|
| Maize | | | | | | | | | |
| Beans | | | | | | | | | |
| Soya beans | | | | | | | | | |
| Groundnuts | | | | | | | | | |
| Bananas | | | | | | | | | |
| Cassava | | | | | | | | | |
| Tomatoes | | | | | | | | | |
| Onions | | | | | | | | | |
| Paprika | | | | | | | | | |

11. During which months of the year do you sell the following crops products?

| Crop/livestock produce | Months sold | Who do you mostly sell to (see codes below) | Where do you usually sell the crop (see codes below) | Distance to market | How often do you sell your produce?<br>*1=Daily*<br>*2=Once every week*<br>*3=Once very month*<br>*4= Once a year* |
|---|---|---|---|---|---|
| Cassava | | | | | |
| S.potatoes | | | | | |
| Beans | | | | | |
| Gnuts | | | | | |
| Maize | | | | | |
| Green maize | | | | | |
| Mangoes | | | | | |
| Rice | | | | | |
| Onions | | | | | |

***Codes for buyer 1= local trader; 2=long distance trader; other farmers, others (specify)…***
***Codes for place of sale 1=on farm; 2=Roadside near village; 3=local market; 4=district town; 5=distant market; 7=others (specify)…***

12. How do you access information on market and price?________________
***1=Radio, 2=extension office, 3=Fellow farmers, 4=neighbour, 5=group members, 6=new papers, 7=others specify***

13). Have you ever organized yourself with other farmers to sell in groups? **1=Yes 0=No**,

14). if yes, what crop/enterprise, with whom, how many times, what markets/where and what was the difference??

| Enterprise sold together | With whom | How many times | What markets? | What was the difference |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |

**INCOME**

15. Rank your sources of income in order of importance to your household

| Income source | Importance<br>***0=None; 1=Negligible; 2=moderate; 3=high; 4=very high*** | Rank | What is the average annual income from this source |
|---|---|---|---|
| Poultry | | | |
| Crops | | | |
| Animals/Livestock | | | |
| Running business | | | |
| Salary | | | |
| Food for work | | | |
| Trees | | | |
| Fruits | | | |
| Remittances | | | |
| Casual labour (ganyu) | | | |
| Others (specify) | | | |

|  |  |  |  |
|---|---|---|---|
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

16. In what months of the year does your household have the most income? In what months does your household have the least income?

| Income | Months | Amount MK (Range) | What do you sell during that period? |
|---|---|---|---|
| Most income | 1. |  |  |
|  | 2. |  |  |
|  | 3. |  |  |
| Least income | 1. |  |  |
|  | 2. |  |  |
|  | 3. |  |  |

17). At any time last year (last 12 months), did you or anyone in the household do any day labor for income?
**1=Yes; 0=No**

18). If your answer is yes, indicate how many people ___________; and which months____________

18b) How much where  you paid on a daily basis?
______________________________MK

19). Do you have savings? **1=Yes, 0=No**.
If yes, how often do you save money? ***0=Never; 1=occasionally; 2=regularly; 4=Always***

20). Where are your individual savings kept?
***1=at home; 2=with another person; 3= personal bank account; 4=group account; 5=others specify***

21). What are the priority uses for the money?
1………….2……………3………………
***1=education; 2=health; 3=loan payment; 4=agricultural input purchase; 5=housing; 6=consumption 7=celebrations; 8=others (specify)***

22. If your household income were to double, what would you do with the extra money?

| Decision | Rank the 5 most important decisions **(1 first priority)** |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

**FOOD SECURITY AND DIETARY DIVERSITY**

23. In 2005/6, how long did your harvest last?

| Crop | How long did the harvest last? (no. of months) | How long do you think your harvest will last this season? (no. of months) |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |

24). If you faced any food shortage in the past two seasons, how did you get additional food?
***0=no shortage; 1=buy food from the market; 2=buy from other farmers in the village; 3=borrow or beg for that food; 4=work for that food; 5=sell property to buy that food; 6=gather wild food; 7=eat other foods; 8=others (specify)***

25). If production is not sufficient year round, please specify the main reasons *(by circling all that apply)*
***1=not enough land, 2=Drought, 3=Poor soils, 4=Lack of fertilizers, 5=Lack of planting material, 6=Pest and diseases, 7=others (specify)***

26. What food do you normally eat during the following months?

| Months | What are the main foods that your household consumes? | How many meals per day do you eat during that period? | Source of food **see codes below** | How often do you eat eggs, meat and fish in a month during that period? **See codes below** |
|---|---|---|---|---|
| January-March | | | | |
| April-June | | | | |
| July-September | | | | |
| October-December | | | | |

***Codes for source of food; 1=home production; 2=purchase 3=Exchanging items for food; 4=Food aid; 5=Gathering wild fruits and vegetables ; 6=Food for Work 7=Other (specify)***

***Codes for no. of times of eating meat and fish; 0=less than once a week; 1=at least once a week; 2=about twice a week; 3=about three times a week; 4=almost everyday***

26b)  Please give the amounts of the main food commodities that your household consumes on a monthly basis

| Month Food is consumed | Commodity 1 Maize<br>Amount Consumed (kg) | Commodity 2______________<br>Amount consumed (kg) | Commodity 3________________<br>Amount consumed (kg) |
|---|---|---|---|
| January | | | |
| February | | | |
| March | | | |
| April | | | |
| May | | | |
| June | | | |
| July | | | |
| August | | | |
| September | | | |
| October | | | |
| November | | | |
| December | | | |

**SOCIAL CAPITAL AND CONFLICTS**

27 a. Are you or your spouse a member of any farmers' group or organization? 1= Yes   0 = No

27 b.  If yes, give please fill in the table below

| Name of group or organization (include local institution) | Composition of group<br>***1=Mixed***<br>***2=Women only***<br>***3=Men only*** | Your position in the group<br>***1=committee member***<br>***2=ordinary member*** | How long have you been a member of this group? (*months/years-specify)* | Does your wife or husband belong to the same group with you ***1=Yes, 0=No*** |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |

28. How often in the past six months have you or members of your household joined with other members of the community  to work collectively?

| Type of activity or occasion | How many times did this take place in the past six months | Estimate number of people who participated | |
|---|---|---|---|
| | | Male | Female |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

29. How would you assess this village on the following aspects? Tick where appropriate

| Aspect | **0=Never happens** | **1=poor** | **2=Average** | **3=Good** |
|---|---|---|---|---|
| 1. Participation in community activities | | | | |
| 2. Extent of trust among people | | | | |
| 3. Cooperation among people | | | | |
| 4. Extent of giving or exchanging gifts | | | | |
| 5. Extent of financial contribution for community activities or collective problems | | | | |
| 6. Extent of financial contribution for farmer  group/organization activities | | | | |
| 7. Spirit of helping others especially the poor | | | | |
| 8. Extent of settling conflicts or disputes among people | | | | |
| 9. Extent of abiding by the norms and byelaws | | | | |
| 10. Women confidence to speak in public | | | | |
| 11. Men’s respect and consideration of women | | | | |

30. What are the three biggest areas, which lead to misunderstanding and disputes between men and women in your community? Rank in terms of importance (1 MOST IMPORTANT)

1. ________________________________

2. ________________________________

3. ________________________________

**HUMAN CAPITAL DEVELOPMENT**

31a). Have you ever received any training or made a study tour to a research station or other farmers on crop or livestock management?
***0=None, 1=Training, 2=Study Tour , 3=Training and Study tour***

31b). If yes, please tell me the type of training or visit, number of times of training, who organized it, where and when.

| Type of Training or visit | No. of times of training or study tour | Where did the training / study tour take place? *(see codes below)* | When ? | Who organized it? |
|---|---|---|---|---|
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

***Codes for where training / study tout took place 1=Within the village/section 2= In another village within the district 3=Another district 4=At research station***

32). How have you used the knowledge and skills acquired from the trainings and study tours?

1. ______________________________________________________________

2. ______________________________________________________________

3. ______________________________________________________________

4. ______________________________________________________________

33). If you have trained other farmers, indicate the number of people trained and the knowledge/skill passed on

| Type of training | No of people trained | |
|---|---|---|
| | Male | Female |
| | | |
| | | |
| | | |
| | | |
| | | |

34. Is there anyone within this community or outside who helps you to solve your agricultural problems? 1=Yes; 0=No.

34 b. If yes, mention *(circle all that apply)*
***1=Farmers in this village, 2=Government extension worker, 3=NGO extension worker, 4=Group members,***
***5=Researcher, 6. Veterinary officer, 7=Others (specify)***

35. How would you assess your ability to do the following

| Do you think you can | ***0=No, not yet***<br>***1=poor***<br>***2=Average***<br>***3=Excellent*** | Who can do it better<br>***1=group members***<br>***2=Committee member***<br>***3=Local Leader(e.g chief),***<br>***4=Others*** |
|---|---|---|
| Address a group of visitors from outside you village? | | |
| Help other farmers to solve their problems? | | |
| Do your own experiments on agriculture? | | |
| Bargain with middle men | | |
| Sell your products | | |
| Explain your group activities/plans to visitors | | |

**ASSESTS**

36. Please indicate how many of these assets you have in your household.

| Asset | No. of assets | Ownership | | |
|---|---|---|---|---|
| | | Husband | Wife | Joint ownership |
| Bicycles | | | | |
| Motor cycle | | | | |
| Ngolo / ox cart | | | | |
| Granaries with food | | | | |
| Radios | | | | |
| Beds | | | | |
| Blankets | | | | |
| Mattresses | | | | |
| Chairs | | | | |
| Mats | | | | |
| Agricultural small tools spade<br>Hoe | | | | |
| Mobile phones | | | | |
| Television | | | | |
| Sofa chairs | | | | |

37. What are the decisions that men and women can take independently or jointly on the following crops, livestock and household activities? *(Please tick)*

| Type of decision | Decisions that men take independently, without consulting their wives | Decisions that men and women consult and take together | Decisions that women can take independently, without consulting their husbands |
|---|---|---|---|
| Decisions over what to grow on your land | | | |
| Decision of which crop varieties to grow | | | |
| Decision on whether to use fertilizers, which types and on which crops | | | |
| Decisions on whether to sell maize | | | |
| Decisions on whether to sell other crops | | | |
| Decision on going to markets to sell crop products | | | |
| Decisions on whether to sell livestock | | | |
| Decision on going to markets to sell livestock products | | | |
| Decision on keeping money | | | |
| Decision to borrow money | | | |
| Decisions on what to cook | | | |
| Decision on who will go for trainings or study tours | | | |
| Community decisions | | | |

**FARMERS' HOUSEHOLD CHARACTERISTICS**

| **Questions** | **Response** |
|---|---|
| 38. Sex of household head<br>***1=Female; 0=Male*** | |
| 39. Age in number of years of;<br>**a*) household head*** | |

| | |
|---|---|
| ***b) spouse*** | |
| 40. Marital status<br>***1=Married; 2=single; 3=Divorced; 4=Widowed<br>5=others (specify)*** | |
| 41. Level of education of head of household?<br>***0=no formal education; 1=primary education (Std1-Std 8); 2=secondary education (F1-F4); 3=completed MSCE 4=Certificate 5= diploma 6=degrees; 7=Postgraduate 8= Adult Literacy<br>9 =others (specify)…..*** | |
| 42. Type of residential main house (housing material) | |
| a) Wall<br>***1=Mud, 2=burnt bricks 3=Unburnt bricks 4=Cement*** | |
| b) Roof<br>***1=Thatch 2=Iron sheets)*** | |
| c) Floor<br>***1=Mud 2=Cement*** | |
| 43. Number of rooms in the house | |
| 44. How many people are currently living with you?<br>***Adult (F+M) aged 60+*** | |
| ***Adult females (18-59)*** | |
| ***Adult males (18-59)*** | |
| ***Children (7-17)*** | |
| ***Young children below 6 years*** | |
| 45. Where does the head of household reside?<br>***1=within village; 2=other village; 3=town/city*** | |
| 46. Do you have any other occupation other than farming?<br>0=No 1=Yes | |
| 47. If yes, which one?<br>***1=Teacher; 2=Agriculture officer; 3=Business 4 = Other (please specify)______*** | |
| 48. Have you ever lived outside this village?<br>***0=No; 1=in another village in the District; 2=village outside the District; 3=town; 4=City, NA = non applicable*** | |
| 49) What would be your assessment of your household well being?<br>***1=Poor 2=Medium 3=Rich*** | |
| 50) What are the reasons for your perceptions? | |
| 51). Out of your 10 neighbors, how many do you think are better off than you in terms of wealth? | |

52. Information on school going children. Please fill the table below

| Sex | Number of school age | How many attend primary schools? | How many attend secondary schools? | How many have dropped out of school? | Reasons for not being in school |
|---|---|---|---|---|---|
| Girls | | | | | |
| Boys | | | | | |

53. Do you think there are  people living with HIV/AIDS in this community?
**0=No, 1= Yes**, **a few people; 2=Yes, many people; 3=HIV/AIDS is now common**

54. Has there been any HIV/AIDS related death in this village in the last 2-3 years?
**0=No; 1=Yes, a few people; 2=yes, many people; 3=death is now common**

55. What are the most five important changes would you like to occur in your household in the next three years?

________________________________________________________________________
________________________________________________________________________
____________________________________________________
________________________________________________________________________
___________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
______

57. Would you like to make any comments or ask questions? 1=Yes 0=No, if yes, what are the comments or questions?

*Thank you very much*

*****************************************************************************************************

**Interviewed by……………………………………………..…….Date……………………… …………………………………**
**Time taken to complete interview……………………**
**Observations**
**……………………………………………………………………………………**
**……………………………………………………………………………………**
**……………………………………………………………………………………**

**Checked by………………………………………………………..….Date……………………
……………………………………….**

**Comments**