College of Natural and Computational Sciences

Department of Statistics

Statistical Analysis of non-climatic factors associated with Ethiopian cereal Crops Yield

By: Moyibon Oli

A thesis submitted to the Department of Statistics in partial fulfillment of the requirements for the degree of Master of Science in Statistics (Applied Statistics)

Advisor: Mekonnen Tadesse (Assoc. Prof)

ADDIS ABABA UNIVERSITY

GRADUATE PROGRAMS

This is to certify that the thesis prepared by Moyibon Oli, titled: Statistical analysis of non-climatic factors associated with Ethiopian cereal crops yield and submitted in partial fulfillment of the requirements for the Degree of Master of Science in Statistics (Applied Statistics) complies with the regulations of the university and meets the accepted standards with respect to originality and quality.

Signed by the examining committee:

Advisor: **Mekonnen Tadesse (Assoc. Prof)** **Signature** __________Date _______________

Examiner 1: **Eshetu Wencheko (Professor)** Signature_____________Date _______________

Examiner 2: **Habte Tadesse (Dr.)** Signature____________Date _______________

____

Chair of Department or Graduate Program Coordinator

DECLARATION

I, Moyibon Oli, hereby declare that this thesis titled "*Statistical analysis of non-climatic factors associated with Ethiopian cereal crops yield*" has been composed solely by myself and that it has not been submitted, in whole or part, in any previous application for a degree or professional qualification. Except where stated otherwise by reference or acknowledgement, the work presented is entirely my own. I confirm that appropriate credit has been given within this thesis where references has been made to the work of others.

Name: Moyibon Oli

Signature: ___________________

Place of submission: Department of Statistics, College of Natural and Computational Sciences, Addis Ababa University, Ethiopia.

Date of submission: August, 2021

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

Advisor: Mekonnen Tadesse (Assoc. Prof) ____________________

Signature

Acknowledgment

I would like to pay my special thanks to the almighty God. And I wish to show my gratitude to my Advisor Mr. Mekonnen Tadesse for what he tirelessly fed into my paper writing, and his commitment to guide me to do my best to come to this point.

I also would like to express my special thanks to all people around me, especially Kedebe Abu, my academic teachers, and my organization for its flexibility to cope up with my classes.

At last but not least, I would like to extend my thankfulness to Ethiopian Central Statistical Agency employees who are highly cooperating in providing me with the data I was intended to use.

ABSTRACT

It has been demonstrated that an increase in crop yields significantly reduces poverty. Agriculture in Ethiopia is the foundation of the country's economy, accounting for more than half of the Gross Domestic Product (GDP). The main objective of the study was to identify non-climatic factors significantly contributing to cereal crop yield in Ethiopia using the data obtained from the 2019/2020 agricultural survey provided by the Central Statistical Agency. In this study, Analysis of Covariance (ANCOVA) and dummy variable regression have been employed to identify non-climatic factors significantly contributing to cereal crop yield in Ethiopia and determine if there is a statistically significant difference between the factor levels of the categorical predictor variables on the mean response variable (average yield of cereal crops measured in kilo gram). The study results showed that seed type used, DAP, area size, and the interaction of DAP and area size were significantly contributing to the change in the mean amount of cereal crop yields in Ethiopia, even though there are regional variations in the factors/variables. Policymakers, the Ethiopian Ministry of Agriculture, and/or all concerned stakeholders should facilitate the accessibility and availability of improved seed and DAP to all farmers so that crop yield is increased and the demands for agricultural products of the country are met.

Table of Contents

CHAPTER 1 .... **1**

1. Introduction **1**

CHAPTER 2 .... **4**

2. Review of Related Literature **4**

CHAPTER 3 .... **8**

3.4 Data Analysis software **13**

3.5 Goodness of fit tests **13**

3.6 Assumptions **14**

3.6.1 ANCOVA assumptions **14**

3.7 Model Diagnostics **15**

CHAPTER 4 .... **16**

4. Results **16**

4.1 Descriptive Results **16**

Table 4.1. Regional average cereal crops yield 17

4.1.2 Cereal Crops Yield Trend **17**

4.2 Assumptions **18**

4.2.1 Linear relationship between the dependent variable and covariates **18**

4.2.2 Multicollinearity **19**

Table 4.2.Variance Inflation Factor corresponding to the continuous independent variables 19

4.3 Modelling **19**

4.3.1 Analysis of Covariance (ANCOVA) **19**

Table 4.3: ANCOVA results with all main effects included 19

Table 4.4: ANCOVA model fit results with only significant main effects 20

Table 4.5: ANCOVA model fit results involving interaction effect 20

4.3.2 Dummy Variable Multiple Linear Regression Analysis **21**

4.3.2.1 Intercept-only model **21**

Table 4.6. Intercept-only model fit results 21

4.3.2.2 Full/final model **21**

Table 4.7: Multiple linear regression model fit results 22

4.3.2.3 Cereal crops yield regression modeling for each Region **23**
4.4 Model Diagnostics **25**
4.4.1 Homogeneity of covariate regression coefficients **26**
Table 4.8: Area size regression slope test 26
Table 4.9: Quantity of DAP regression slope test 26
4.4.2 Constant error variance **27**
Table 4.10: Levene's test of homoscedasticity 27
4.4.3 Normality **28**
Table 4.11: Z-value computed from skewness and kurtosis 28
5. CHAPTER 5 .... **30**
5.1 Discussions **30**
6. Conclusions and Recommendations **32**
6.1 Conclusions **32**
6.2 Recommendations **32**
References 33
Appendices **37**

Acronyms

ANCOVA	Analysis of Covariance
CSA	Central Statistical Agency
CIMMYT	International Wheat and Maize Improvement Center
EA	Enumeration Areas
E.C	Ethiopian Calendar
FAO	Food and Agricultural Organization
GDP	Gross Domestic Product
Ha	Hectares
IFAD	International Fund for Agricultural Development
Kg	Kilograms
Mha	Million hectares
PSU	Primary Sampling Units
SSU	Secondary Sampling Units
SSNP	Southern Nations Nationalities and Peoples
UN	United Nations
VIF	Variance Inflation Factor

CHAPTER 1

1. Introduction

1.1 Background of the study

It is well known that crop yields are much lower in sub-Saharan Africa compared to elsewhere in the world and this has led to the need for the continent to import a large share of its food needs (UN, 2015). The weakness in the productivity of crops across Sub-Saharan Africa is not only related to the poor soils in many countries (CIMMYT, 2015) but also due to the limited use of essential inputs that are needed to raise the productivity level. These inputs include the use of improved seeds, fertilizers, irrigation, and pesticides. The hypothesis is that the use of these inputs would boost the productivity of crops.

Crop production in Ethiopia is used for food consumption, production for export, and raw materials for the industry. Agriculture is the backbone of the Ethiopian economy. It constitutes over 50 percent of the Gross Domestic Product (GDP), accounts for over 85 percent of the labour force, and earns over 90 percent of the foreign exchange (Zerihun et al., 2002). And according to the report by Agricultural growth Ethiopia (2014), Ethiopia's economy remains highly agrarian. Agriculture accounts for approximately 40 percent of GDP, 80 percent of exports, and employs an estimated 75 percent of the workforce.

Agriculture in the 21st century will continue to face multiple, interconnected challenges all over the world. Some countries must produce more food to feed a growing population characterized by changing consumption patterns, and dietary and nutritional preferences. Food and Agriculture Organization (FAO, 2017) conservatively projected that if the global population reaches 9.1 billion by 2050, world production will need to rise by 70%, and food production especially in the developing world will need to double. In Sub-Saharan Africa countries, there exists a huge difference between the projected yield and actual yields because often production is at subsistence levels. It has been demonstrated that every 1% increase in agricultural yield translates into a 0.6% to 1.2% decrease in the number of absolute poor households in the world (Thirtle et al., 2001). And according to World Bank (2014), 10 percent increase in agricultural productivity decrease the likelihood of being poor by 2.5 to 3 percent.

The pressure to increase crop production in many countries has resulted in the expansion of land area dedicated to agriculture and the intensification of cropland management through practices such as irrigation, use of large quantities of inputs like inorganic fertilizers and synthetic chemicals for pest and weed control (Oldfield et al, 2019). These practices have resulted in degradation of soil properties and water quality, acceleration of soil erosion, contamination of groundwater, and decline of food quality. This has prompted sustainable intensification initiatives to increase yields on existing farmland while decreasing the environmental impact of agriculture (Wang, 2014).

According to Global Yield Gap Atlas (2108), major Ethiopian staple crops are classified as cereal crops, pulse crops, oilseed crops, roots and tubers, vegetables, and coffee. According to the Ethiopian Central Statistical Agency report (CSA, 2016), grain crops (cereal, pulse, and oil crops) are cultivated on 12.5Mha with an annual production of 26.7million metric tons. These crops constituted about 80, 13, and 7% of the cultivated area respectively.

The country's five main cereal crops (teff, maize, sorghum, barley, and wheat) constitute three-quarters of the total cultivated (Alemayehu, et al, 2011).

Beyond expanding the harvesting area, it was very reasonable to improve the quality and amount of crop yield in the existing harvested areas. Expanding harvested areas may cause the degradation of soil due to the removal of trees from the area, heavy erosion, and etc.

The crop yield deficits recorded in most Sub-Saharan African countries are accentuated by climatic and non-climatic factors, including limited access to sufficient farm inputs, limited use of agro-ecology-related inputs, and climate variability (Negin et al., 2009; Denning et al., 2009). Yet in Ethiopia, it remains unclear which non-climatic factors are most important in determining yields of crops.

The main aim of this study is to identify and analyze the major non-climatic factors that may influence the amount of cereal crop yield in the harvested areas of Ethiopia. These non-climatic factors have been categorized as natural factors (field orientation and land degradation), human factors (irrigation use, field prevention from erosion and previous state of the field), fertilizer and seedling type, and crop damage due to both human and natural factors. This study also describes the trends in cereal crop yield over the last three years using agricultural sample survey data obtained from CSA.

1.2 Statement of the problem

Increasing global population, changes in consumption patterns and dietary needs, and rising demand for green energies have triggered a global need for increased food production. It is estimated that one in seven people lack access to food or are faced with malnutrition caused by poverty and rising food prices (Epule. *et al*., 2018).

Beyond its contribution to Ethiopia's largest GDP, agriculture is an irreplaceable source of feeding and industrial input in Ethiopia. As its demand for food is high, this study attempts to identify important non-climatic factors affecting the yields of cereal crops in harvested areas of Ethiopia.

1.3 General objective of the study

The main objectives of this study is to identify non-climatic factors significantly contributing to cereal crops yield in Ethiopia.

1.3.1 Specific objective of the study

This study specifically focuses on identifying if there is a statistically significant difference between the factor levels of a categorical predictor variable on the mean response variable (average yield of cereal crops measured in kg) controlling the effect of covariates in the model.

It also identifies the effect of fertilizers and area size on the change in the amount of cereal crops yield in Ethiopia.

1.4 Scope of the study

This study covers all regions of Ethiopia except Afar and Somali, using Central Statistical Agency's post-harvest agricultural data. The study has also undertaken a comparative analysis which will show a trend of average cereal crop yield of three consecutive years.

1.5 Significance of the study

Agriculture in Ethiopia is the foundation of the country's economy, accounting for about 40 percent of the gross domestic product (GDP), 80 percent of exports, and 75 percent of total employment.

This study is expected to alert policymakers about factors influencing cereal crops yield so that reforms are made to accelerate growth in cereal crop production, distribution, and to achieve growth in fertilizer use and improve fertilizer distribution.

Moreover, the results of this study may provide information on the underlying factors associated with cereal crop yield which may be fundamental for designing effective programs/policies to address the issue. In addition to this, the result of this study may be used as a basis for further studies in this area.

1.6 Limitations of the study

Even though it is very worthy to study the combined effect of climatic, soil structure, and non-climatic factors that will affect the yields of crops in Ethiopia, the study has focused only on the non-climatic factors that affect yields of crops due to the unavailability of complete information on climatic and soil structure factors.

CHAPTER 2

2. Review of Related Literature

Several factors pose a significant risk to farms leading to yield reduction when they are not correctly monitored and well managed. These factors can be categorized into three; which are technological, biological, and environmental (Tandzi and Mutengwa, 2019).

There is wide geographic variation in crop and livestock productivity, even across regions that experience similar climates. The difference between realized productivity and the best that can be achieved using current genetic material and available technologies and management is termed as the "yield gap." The best yields that can be obtained locally depend on the capacity of farmers to access and use, among other things, seeds, water, nutrients, pest management, soils, biodiversity, and knowledge. Substantially more food, as well as the income to purchase food, could be produced with current crops and livestock if methods were found to close the yield gaps. Low yields occur because of technical constraints that prevent local food producers from increasing productivity or for economic reasons arising from market conditions. For example, farmers may not have access to the technical knowledge and skills required to increase production, the finances required to invest in higher production (e.g., irrigation, fertilizer, machinery, crop- protection products, and soil-conservation measures), or the crop and livestock varieties that maximize yields (H. Charles J. et al. 2010).

After harvest or slaughter, they may not be able to store the products or have access to the infrastructure to transport the produce to consumer markets. Farmers may also choose not to invest in improving agricultural productivity because the returns do not compare well with other uses of capital and labor. Exactly how best to facilitate increased food production is highly site-specific. In the most extreme cases of failed states and non-functioning markets, the solution lies completely outside the food system. Where a functioning state exists, there is a balance to be struck between investing in overall economic growth as a spur to agriculture and focusing on investing in agriculture as a spur to economic growth, though the two are obviously linked in regions, such as sub-Saharan Africa, where agriculture typically makes up 20 to 40% gross domestic product. In some situations, such as low-income food-importing countries, investing purely in generating widespread income growth to allow food purchases from regions and countries with better production capabilities may be the best choice (H. Charles J. et al. 2010).

Due to environmental degradation and loss of ecosystem components, there would be a reduced yield of food crops. Unsustainable practices in irrigation and production may lead to increased salinization of soil, depletion of nutrients, and erosion. This, in turn, causes lower yield. The productivity of some lands has declined by 50% due to soil erosion and desertification. Africa is considered to be the continent most severely affected by land degradation. Global climate change can also affect food production; by changing overall growing conditions (rainfall distribution, temperature regime); by inducing more extreme weather such as floods, storms, and drought; and by increasing extent, type, and frequency of infestations, including that of invasive alien species. All this would adversely affect yield.

The five major cereal crops in Ethiopia are teff, wheat, maize, sorghum, and barley. Teff accounts for 28 percent of the total cereal area, while maize stands for 27 percent of total annual cereal production. After cereals, the second most important crop group (in terms of acreage) pulses. In 2004/05-2007/08 6.4 million holders grew pulses on 12.4 percent of the total area cultivated. Total pulse production averaged 1.5 million tons per year, which is 8.5 percent of total crop production. Oilseeds form the third most important crop group. It is cultivated on 6.9 percent of the total area cultivated, by 3.1 million holders. They produce an average of 0.5 million tons of oilseeds yearly, i.e. 3 percent of total annual production. Coffee is a major cash crop, accounting for 3.8 percent of GDP and 19 and 35 percent of the quantity and value of exports respectively in that period, but occupying only 2.7 percent of the total area cultivated (i.e. 306 thousand hectares). Chat, another stimulant crop, is cultivated by 2 million farmers (Alemayehu et al, 2011).

Studies indicate that non-climatic factors that affect crop yield in Uganda were cattle stock, arable, irrigation, fertilizers, and population. The same study indicates that an increase in cattle stock imposes a reduction in crop yield in Uganda because herds are required for land grazing and in some cases herds of cattle affect yields by taking up land, eating crops. Arable: the more this variable increases, the higher the yields in the short term and the lower the yields per Ha if agriculture is not mechanized in the long term. Irrigation: in regions that are faced with problems of recurrent droughts as is the case of Uganda, irrigation facilities serve as safety nets that do sustain crop yield during periods of dryness. Population: Normally when population growth is high the demand of the country is likely to be high, and this creates food insecurity challenges if the population growth fails to be matched by increased food production (Epule et al, 2018).

The same study indicated that fertilizers constitute a very important input into the agricultural system. When farmers have adequate access to fertilizers, their crop yield is likely to increase. However, in most sub-Saharan Africa including Uganda, access to inorganic fertilizers is often limited by purchasing power because the farmers are poor. Organic fertilizers which are often free and agrological and more sustainable have not been sufficiently valorized to levels at which they can sustain yields without inorganic fertilizers. Organic fertilizers include manure, potassium, phosphorus, and magnesium fertilizers inter alia.

Crops like corn and milo are very responsive in yield to the previous crop. While soybeans and wheat tend to be less responsive, Crop rotation is good, but we need to increase our diversity of crop selection in the future. The advance of biotech traits in corn and soybeans is continuing to push us to be less diverse. Cover crops help us to add diversity in the future (Paul C Hay, 2016).

The study conducted by (Kassa, 2015) using linear regression modeling shows that fertilizers and improved seed utilization, irrigation, plot soil quality, and crop rotation were significant factors that determine crop yield amounts in the Tigray region of Ethiopia.

There is a strong positive correlation between crop yield and fertilizer indicating that an increase in fertilizer access to farmers will lead to a corresponding increase in yield. And also other variables such as labor, farm size, and expenditure have had a positive correlation with yield (Zare et al, 2013).

Crop rotation has the advantage of reducing the influence of residual soil-borne disease which was adapted to the genetic trait of the crops previously cultivated on the same land. Crop rotations have historically been driven by the need to build sufficient fertility through the crop sequence and to maintain control of antagonists (e.g. weeds, pests, and disease) (Knox et al, 2011).
Rice and wheat rotation was practiced for many years and expected during the green revolution in south Asia (Yadaw et al., 2002).

Land degradation is a complex phenomenon influenced by natural and socio-economic factors. In many economic analyses, there was a tendency to attribute soil fertility decline only to soil erosion. Erosion was treated as the sole contributing factor to soil/land degradation and yield declines, as the impacts of nutrient depletion on crop yields were underestimated or completely neglected (Kerr and Pender, 2005). Land degradation was a major cause of the country's low and declining agricultural productivity, persistent food insecurity, and abject rural poverty (FAO, 2010).

The major cause of land degradation that influences the livelihoods of rural people economies like soil erosion by heavy rain which facilitated by farming sloppy, clearing forest land for cultivation due to raising of the human population, low level of education resisting to accept the new packages of soil and water conservation measure and sustainable management of natural resources (Feyera and Tsetadirgachew, 2015).

Since the start of national agricultural statistics in the 1960s in Ethiopia, teff has always accounted for the largest share of cereal area cultivated. However, over the past five decades, the share teff has declined gradually (a decrease by 5.8 percentage points from the 1960s to the first decade of the 2000s), while the share of maize has increased by 7.8 percentage points. Compared to teff, the share of other cereals stayed relatively stable over time. During the first decade of the 2000s, production of all the major cereals increased, with teff (8.9 percent), sorghum (8.6 percent), and wheat (8.3 percent) having the fastest annual growth rates. Growth in maize production, 6.8 percent per year in the 1980s and 5.5 percent per year in the 1990s, slowed to 4.2 percent in the early 2000s. This reduced growth figure was caused in part by a collapse in domestic maize prices inducing slow adoption, or even disadoption, of hybrid maize technology (Alemayehu et al, 2011).

The Ethiopian agricultural production system is largely characterized by subsistence orientation, low-level inputs, dependency on rainfall, and limited integration to the market. According to IFAD, 2012, about 12.7 million smallholders produce 95 percent of agricultural GDP, depicting the trend of production in Ethiopia. Accordingly, it is highly fluctuating and even declining in some existence. Climate is the major cause for this. For example, the yield sharply dropped in 1984/85 due to the great drought (Reta, 2014).

Comparing Ethiopia's cereal yield between 1999 and 2009, there seems to be a modest increase in the use of modern inputs such as seeds and chemical fertilizer. It was also reported that the use of improved seeds significantly affects yields in both years (Getachew, 2011).

The two major inorganic fertilizers that Ethiopian farmers are using currently are nitrogen and phosphorus in the form of UREA and DAP respectively. Inorganic fertilizers can easily be available for plants to fulfill nutrient requirements of crops while it is applied in recommended time, plant growth stage, frequency of application, depth of fertilizer placement in soil and the method of application should coincide with the package of production made by researchers. (Merga and Haji, 2019).

In Ethiopia, farmers are applying fertilizers haphazardly which could be below or above the recommendation of the nutrient required by specific cereal crops to nourish the total biomass of the plant. The government has a policy of fertilizer delivery to most farmers cultivating crops regardless of water availability on the farm size of many areas in the country which has a negative impact on production (Merga and Haji, 2019).

Improved seeds have the traits like high yield, early maturity to escape from the harsh climatic condition, resistance or tolerant to disease that reduces the cost of production, rich in nutrients to reduce the mortality of mother and child by providing strength for elders with vitamins, proteins and essential nutrients for human health from like bio-fortified crops. Moreover, it can provide multi-climatic conditions related to technical inputs; including improved seed availability, quality of improved varieties, fertilizers availability, and affordability, (Mesfin and Zemedu, 2015).

CHAPTER 3

3. Methodology and Data

3.1 Data source

Data for this study were obtained from the records available at the Central Statistical Agency (CSA). The data obtained from CSA for this study is agricultural survey data conducted for meher season of the year 2019/2020.

3.1.1 Sampling design

In order to select the sample, a stratified two-stage cluster sample design was implemented. Enumeration areas (EAs) were taken to be the primary sampling units (PSUs) and the secondary sampling units (SSUs) were agricultural households. The sample size for the 2019/20 (2012 E.C.) agricultural sample survey was determined by taking into account both the required level of precision for the most important estimates within each domain and the amount of resources allocated to the survey. In order to reduce non-sampling errors, the manageability of the survey in terms of quality and operational control was also considered.

All regions were taken to be the domain of estimation for which major findings of the survey are reported (CSA, 2020).

3.2 Modeling and Data acquisition

3.2.1 Study Variables

The *dependent/response variable* of the study is cereal crops yield measured in kilograms.

There are a variety of factors associated with cereal crop yield. The *independent/explanatory variables* considered in this study are categorical and continuous. The categorical independent variables assumed to influence crop yield are field orientation, land degradation, seed/seedling type, crop damage, irrigation, natural fertilizer, both chemical and natural fertilizer type, crop field land, and field prevention from erosion, while the continuous independent variables are quantity of UREA used measured in kg, quantity of DAP used measured in kg, and area size measured in hectares.

3.3 Method of Data Analysis

The analysis procedures employed in this study are the analysis of covariance (ANCOVA) and dummy variable regression.

ANCOVA is a statistical method that combines the methods of the analysis of variance (ANOVA) and regression analysis. This analysis method typically includes two types of independent variables: (i) class or dummy variables whose levels are used to identify different treatments; and (ii) continuous variables, called covariates, that are directly measured.

In this study, ANCOVA has been employed to test if there is a statistically significant difference between the factor levels of the categorical predictor variables on the mean response variable (average yield of cereal crops measured in kg). This method helps to evaluate whether the means of the response variable i.e. the average cereal crop yield measured in kilograms are the same across the levels of the categorical independent variables, by controlling the effects of the covariates. While covariates are continuous independent variables, factors are categorical independent variables.

ANCOVA decomposes the variance in the response variable into variance explained by covariates, variance explained by the factors, and residual variance.

In order to measure the influence of each independent variable and their interaction effect on the average cereal crop yields, it is important to run multiple linear regression. The multiple linear regression for the aforementioned outcome and independent variables will be computed after recoding factor variables using dummy coding.

In addition to those above defined variables the effect of the predictor variable on the cereal crop yields may also be non-additive, which will be tested by plotting residual versus the covariate variable. And these non-additive variables can be evaluated by constructing additional variables that are used to measure what may be referred to as multiplicative or interaction effects. The interaction variables are defined as the product of the existing indicator and/or covariate variables. This interaction term can be two ways when two variables are considered together and higher-order interaction effect when more than two variables are considered in the model.

In the current study, only two-way interaction terms are considered. The covariates are, quantity of UREA, quantity of DAP in measure in kilograms, and area size measured in hectares are denoted by $x_1, x_2, and\ x_3$ respectively.

In order to use linear regression and ANCOVA, the factors/categorical explanatory variables were re-coded into a set of separate binary variables using a recording scheme called "dummy coding" as follows.

The general ANCOVA equation is given by:

$$Y_{ij} = \mu + \tau_i + \beta(X_{ij} - \bar{X}) + e_{ij}$$

Where, Y_{ij} *is the* j^{th} *observation under the* i^{th} *category*

X_{ij} *is* is the j^{th} observation of the covariate under the i^{th} group

μ *is grand mean*

$\bar{X}$ *is the gloabl mean for covariate X*

τ_i *is the effect of* ith *level of the independent variable*

β *is the slope of the line*

e_{ij} *is the residual term related with the* i^{th} *group and* j^{th} *observation*

3.3.1 Variables coding scheme

Factor 1: Field orientation (γ).

$$\gamma_{1i} = \begin{cases} 1, & \text{if orientation of the land for the } i^{th} \text{ subject is sloppy} \\ 0, & \text{otherwise} \end{cases}$$

$$\gamma_{2i} = \begin{cases} 1, & \text{if orientation of the land for the } i^{th} \text{ subject is somewhat sloppy} \\ 0, & \text{otherwise} \end{cases}$$

Factor 2: Land degradation (δ).

$$\delta_{1i} = \begin{cases} 1, & \text{if the } i^{th} \text{ subject responded that the land is prevented from degradation} \\ 0, & \text{otherwise} \end{cases}$$

Factor 3: Seedling type (θ).

$$\theta_{1i} = \begin{cases} 1, & \text{if the } i^{th} \text{ subject used improved seed} \\ 0, & \text{otherwise} \end{cases}$$

Factor 4: Crop damage (α).

$$\alpha_{1i} = \begin{cases} 1, & \text{if the } i^{th} \text{ respondent reported crop experienced damage} \\ 0, & \text{otherwise} \end{cases}$$

Factor 5: Irrigation (λ).

$$\lambda_{1i} = \begin{cases} 1, & if\ the\ i^{th}\ subject\ responded\ that\ field\ is\ irrigated \\ 0, & otherwise \end{cases}$$

Factor 6: Fertilizer use (ρ).

$$\rho_{1i} = \begin{cases} 1, & if\ the\ i^{th}\ subject\ used\ fertilizer \\ 0, & otherwise \end{cases}$$

Factor 7: Type of fertilizer used (τ).

$$\tau_{1i} = \begin{cases} 1, & if\ the\ i^{th}\ subject\ used\ natural\ fertilizer \\ 0, & otherwise \end{cases}$$

$$\tau_{2i} = \begin{cases} 1, & if\ the\ i^{th}\ subject\ used\ both\ chemical\ and\ natural\ fertilizer \\ 0, & otherwise \end{cases}$$

Factor 8: Previous state of the field (μ).

$$\mu_{1i} = \begin{cases} 1, & if\ the\ i^{th}\ subject\ reported\ that\ land\ is\ rented\ in\ crop\ field \\ 0, & otherwise \end{cases}$$

$$\mu_{2i} = \begin{cases} 1, & if\ the\ i^{th}\ subject\ reported\ that\ is\ other\ type \\ 0, & otherwise \end{cases}$$

Factor 9: Field prevention from erosion (σ).

$$\sigma_{1i} = \begin{cases} 1, & if\ the\ i^{th}\ subject\ reported\ that\ field\ is\ prevented\ from\ erosion \\ 0, & otherwise \end{cases}$$

The following table, Table 3.1, displays the number of agricultural households at each level of the categorical explanatory variables included in our sample.

Table 3.1 Agricultural households included in the sample

Factor/ variables	Categories	Coding scheme	Frequency	Percentage
Field orientation	Flat	Defined in the Coding scheme above	15,422	52.76%
	Sloppy	>>	10,498	35.92%
	somewhat sloppy	>>	3,309	11.32%
Land degradation	Prevented from land degradation	>>	23,009	78.72%
	Not prevented from land degradation	>>	6,220	21.28%
Seedling type	Improved seed	>>	4,997	17.10%
	Non-improved seed	>>	24,232	82.91%
Crop damage	Crop experienced damage	>>	29,229	100.00%
	Crop not experienced damage	>>	0	0%
Irrigation	Field irrigated	>>	517	1.77%
	Field not irrigated	>>	28,712	98.23%
Fertilizer use	Fertilizer used	>>	21,618	73.96%
	Fertilizer not used	>>	7,611	26.04%
Fertilizer type	Natural	>>	6,989	32.31%
	Chemical	>>	11,197	51.76%
	Both	>>	3,448	15.94%
Previous state of the field	Crop field	>>	28,433	97.28%
	Other fields	>>	796	2.72%
Field prevention from erosion	Prevented from erosion	>>	23,797	81.42%
	Not prevented from erosion	>>	5,432	18.58%

The factor variable 'crop damage' will be excluded from the analysis as all cereal crops experienced damage.

The general additive model is given by:

$$Y_i = \beta_0 + \beta_1\gamma_{1i} + \beta_2\gamma_{2i} + \beta_3\delta_{1i} + \beta_4\theta_{1i} + \beta_5\alpha_{1i} + \beta_6\lambda_{1i} + \beta_7\rho_{1i} + \beta_8\tau_{1i} + \beta_9\tau_{2i} + \beta_{10}\mu_{1i} + \beta_{11}\mu_{2i} + \beta_{12}\mu_{3i} + \beta_{13}\mu_{4i} + \beta_{14}\sigma_{1i} + + \beta_{15}X_{1i} + \beta_{16}X_{2i} + \beta_{17}X_{3i} + \beta_{18}X_{4i} + \beta_{19}X_{5i} + \varepsilon_i \quad \ldots\ldots (1)$$

The general additive and interaction effect model is given by:

$$
\begin{aligned}
Y_i = \; & \beta_0 + \beta_1\gamma_{1i} + \beta_2\gamma_{2i} + \beta_3\delta_{1i} + \beta_4\theta_{1i} + \beta_5\alpha_{1i} + \beta_6\lambda_{1i} + \beta_7\rho_{1i} + \beta_8\tau_{1i} + \beta_9\tau_{2i} \\
& + \beta_{10}\mu_{1i} + \beta_{11}\mu_{2i} + \beta_{12}\mu_{3i} + \beta_{13}\mu_{4i} + \beta_{14}\sigma_{1i} + \beta_{15}x_{1i} + \cdots + \beta_{19}X_{5i} + \beta_{20}\delta_{1i}\gamma_{1i} \\
& + \beta_{21}\delta_{1i}\gamma_{2i} + \beta_{22}\theta_{1i}\alpha_{1i} + \beta_{23}\theta_{1i}\rho_{1i} + \beta_{24}\theta_{1i}\tau_{1i} + \beta_{25}\theta_{1i}\mu_{1i} + \beta_{26}\theta_{1i}\mu_{2i} + \beta_{27}\theta_{1i}\mu_{3i} \\
& + \beta_{28}\theta_{1i}\mu_{4i} + \beta_{29}\sigma_{1i}\delta_{1i} + \beta_{30}\sigma_{1i}\gamma_{1i} + \cdots \\
& + \beta_k(\gamma_{ji}\delta_{1i}\theta_{1i}\alpha_{1i}\lambda_{1i}\rho_{1i}\tau_{ij}\mu_{ij}\sigma_i X_{1i}X_{2i}X_{3i}X_{4i}X_{5i}) + \varepsilon_i \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2) \\
& \quad where \;\; i = is\ the\ i^{th}\ observation\ and\ j\ is\ the\ j^{th}\ category
\end{aligned}
$$

Only two way interaction effect is considered in the model.

3.4 Data Analysis software

The data were analyzed using excel, Stata, R studio, and SPSS software.

3.5 Goodness of fit tests

F-test: The F-test of significance is used to test each main and interaction effect, for the case of a single interval dependent and multiple (>2) groups formed by a categorical independent. F is the between-groups variance divided by the within-groups variance. If the computed p-value is small, then significant relationships exist. Adjusted means are usually part of ANCOVA output and are examined if the F-test demonstrates significant relationships exist. Comparison of the original and adjusted group means can provide insight into the role of the covariates.

For the k groups formed by categories of the categorical independents and measured on the dependent variable, the adjustment shows how these k-means were altered to control for the covariates. Typically, this adjustment is one of linear regression of the type, $Y_{adj} = \overline{\mathrm{Y}} - \beta(\overline{\mathrm{X}}_\iota - \overline{\mathrm{X}})$ where Y is the interval dependent, X is the covariate, i is one of the k groups, and β is the regression coefficient. There is no constant when Y is standardized. For multiple covariates, of course, there are additional similar X terms in the equation.

t-test: A test of significance of the difference in the means of a single interval dependent, for the case of two groups formed by a categorical independent.

In the current study, ANCOVA will be modeled using regression because dummy variables are used for the categorical independents. When creating dummy variables, we have used one less category than there are values of each independent. For full ANCOVA one would also add the interaction cross-product terms for each pair of independents included in the equation, including the dummies. Then one computes multiple regression. The resulting F tests will be the same as in classical ANCOVA. F ratio can also be computed through the extra sum of squares using the Full-Reduced Model approach.

3.6 Assumptions

3.6.1 ANCOVA assumptions

- At least one categorical and at least one interval independent. The independent variable(s) may be categorical, except at least one must be a covariate (interval level).
- Interval dependent. The dependent variable is continuous and interval level.
- Low measurement error of the covariate. The covariate variables are continuous and interval level and are assumed to be measured without error.
- Covariate linearly or in a known relationship to the dependent. The form of the relationship between the covariate and the dependent must be known and most computer programs assume this relationship is linear, adjusting the dependent mean based on linear regression. Scatter plots of the covariate and the dependent for each of the k groups formed by the independents are one way to assess violations of this assumption.
- Homogeneity of covariate regression coefficients; i.e. "parallel lines model". The covariate coefficients (the slopes of the regression lines) are the same for each group formed by the categorical variables and measured on the dependent. The more this assumption is violated, the more conservative ANCOVA becomes (the more likely it is to make Type I errors - accepting a false null hypothesis). There is a statistical test of the assumption of homogeneity of regression coefficients.

- Additivity. The values of the dependent are an additive combination of its overall mean, the effect of the categorical independents, the covariate effect, and an error term. ANCOVA is robust against violations of additivity but in severe violations, the researcher may transform the data, by using a logarithmic transformation to change a multiplicative model into an additive model. Note, however, that ANCOVA automatically handles interaction effects and thus is not an additive procedure in the sense of regression models without interaction terms.
- Independence of the error term. The error term is independent of the covariates and the categorical independents. Randomization in experimental designs assures this assumption will be met.
- Independent variables are orthogonal to covariates. The independents are orthogonal. If the covariate is influenced by the categorical independents, then the control adjustment ANCOVA makes on the dependent variable before assessing the effects of the categorical independents will be biased since some indirect effects of the independents will be removed from the dependent variable.
- Homogeneity of variances. There is a homogeneity of variances in the cells formed by the independent categorical variable Heteroscedasticity is the lack of homogeneity of variances, in violation of this assumption.
- Multivariate normality. For purposes of significance testing, variables follow multivariate normal distributions.
- Compound sphericity. The groups display sphericity (the variance of the difference between the estimated means for any two different groups is the same) a more restrictive assumption,

called compound symmetry, and is that the correlations between any two different groups are the same value. If compound symmetry exists, sphericity exists. Tests or adjustments for lack of sphericity are usually actually based on possible lack of compound symmetry.

3.7 Model Diagnostics

For model diagnostics, we will examine the data and residuals and check the three standard assumptions, check the same-slope assumption (plots, interaction term) and look for influential outliers.

The variance inflation factor is used to detect if there is multicollinearity among the covariates and ANOVA among factor variables. While AIC is used to select the best fitting model among competing models.

CHAPTER 4

4. Results

4.1 Descriptive Results

In this study, data obtained from the 2019/20 agricultural sample survey conducted by the Central Statistical Agency of Ethiopia were used to assess non-climatic factors associated with cereal crop yield. Moreover, data from 2017/2018, 2018/2019, and 2019/2020 calendar years of agricultural sample surveys are used to assess changes over the three years. The study covered a total of 29,229 agricultural households from different regions of Ethiopia except Afar and Somali regions. Of these, the highest proportion, 29% of agricultural households were from Amhara and another 29% from Oromia followed by 19% from Tigray region. The pie chart in Figure 4.1 corroborates this fact.

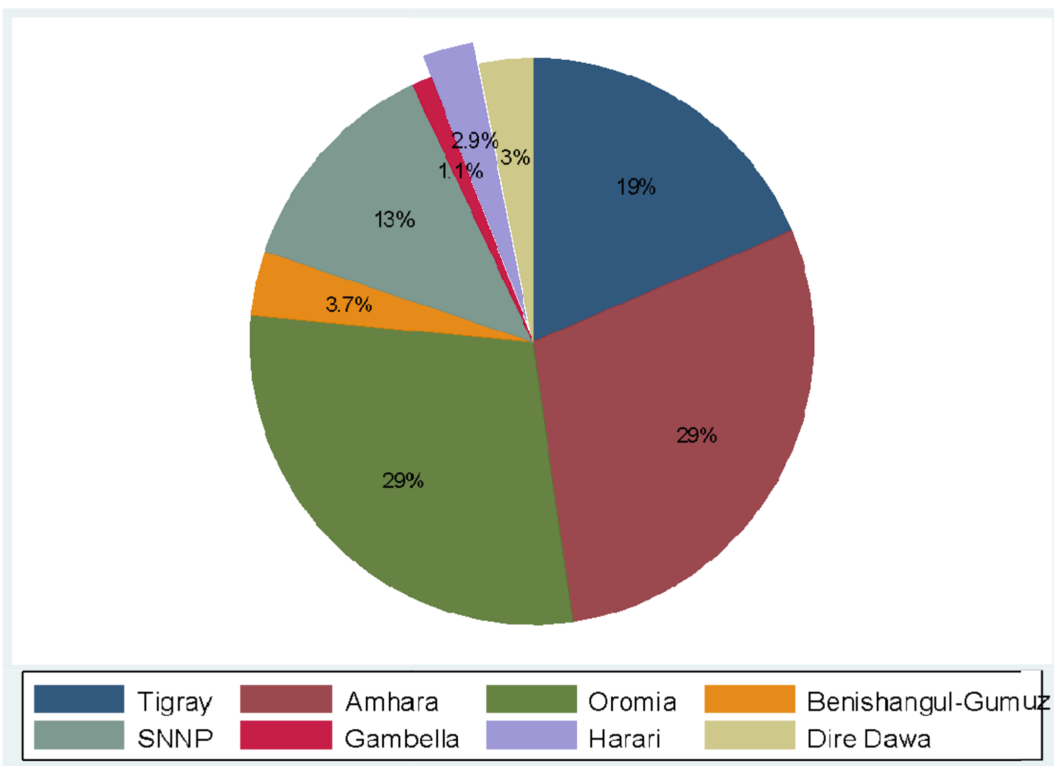

Figure 4.1 *Regions sample contribution for this study*

Table 4.1 provides a summary of average cereal crop yield in kilograms by region and cereal type. As can be seen from Table 4.1 below, Benishangul-Gumuz and Gambella are the two regions with the highest and second-highest average cereal crops yield of 637kg and 564kg respectively followed by Oromia with an average yield of 532kg. Conversely, Diredawa and Harari are the two regions with the least average cereal crop yield of 217 and 203kg respectively.

Table 4.1. Regional average cereal crops yield

Region	Cereal Crops Type					
	BARLEY	MAIZE	SORGHUM	TEFF	WHEAT	Grand Total
Tigray	147	276	839	357	212	374
Amhara	241	336	597	607	337	375
Oromia	433	445	580	559	848	532
Benishangul Gumuz	201	614	707	607	550	637
SNNP	135	401	196	135	135	307
Gambella	375	458	698	0	0	564
Harari	38	97	242	0	44	203
Dire dawa	0	32	238	0	0	217

Cereal crop “teff” is not included in this data analysis from Gambella and Harari regions, and Diredawa city administration, while cereal crop “wheat” is not included from Gambella region and Dire dawa due to the unavailability of records in the data source.

4.1.2 Cereal Crops Yield Trend

The trend of cereal crops yield over three consecutive years (2017/2018 to 2019/2020) is presented in Figure 4.2.

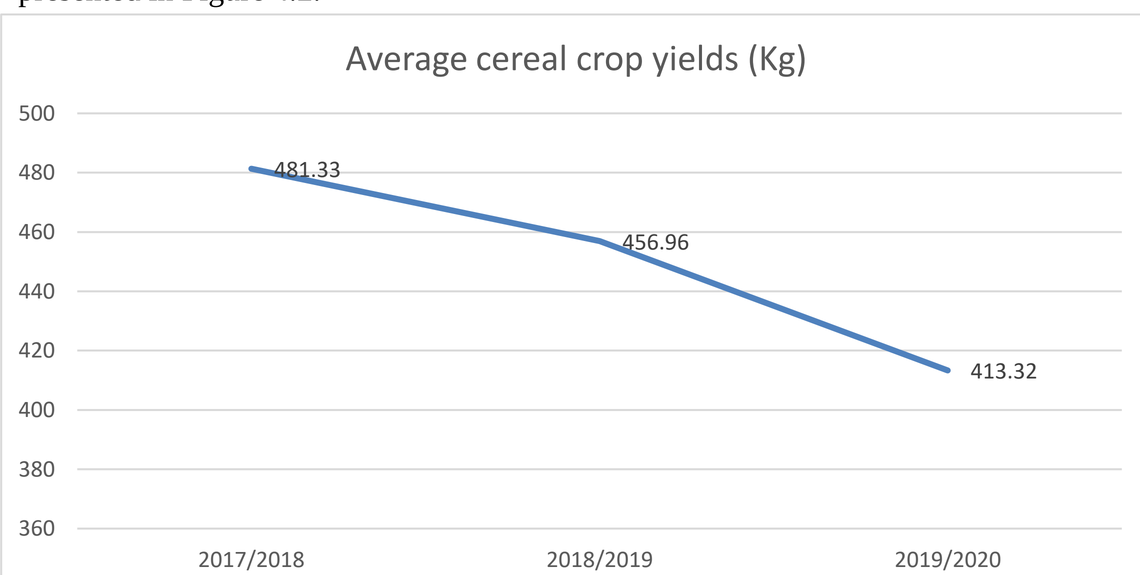

Fig 4.2 Average cereal crops yield for three consecutive years.

Fig 4.2 depicts that there is a decreasing yearly average yield of cereal crops in Ethiopia over the last three years.

4.2 Assumptions

4.2.1 Linear relationship between the dependent variable and covariates

For this study, since the scatter plot of each covariate versus the dependent variable failed to meet the assumption of linearity, the logarithms of both the dependent variable and covariates have been used. Scatter plots of log-transformed cereal crop yield in kg versus log-transformed covariates are presented in Figure 4.3.

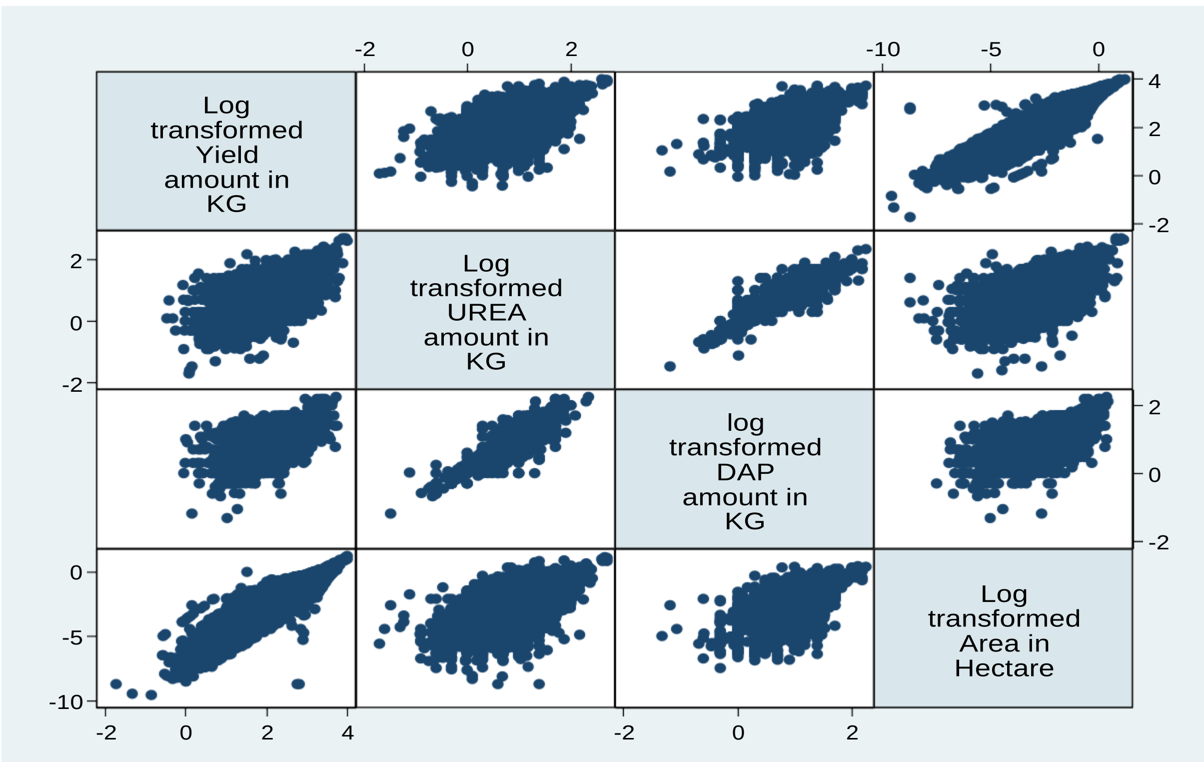

Figure 4.3 *Scatter plots of log-transformed cereal crop yield in kg versus log-transformed covariates.*

The scatter plots of the log-transformed response/dependent variable i.e. cereal crops yield measured in kilogram versus the log-transformed covariates (amount of urea measured in kilogram, amount of DAP measured in kilogram, quantity of mixed fertilizer, and area size measured in hectare) suggest linear relationships as depicted in Figure 4.3.

4.2.2 Multicollinearity

The existence of multicollinearity has been tested using the Variance Inflation Factor (VIF). First, the correlations between the independent variables were assessed and no pair of independent variables possess a strong correlation i.e. the correlation coefficients between any pair of independent variables have not exceeded 0.8. The VIFs computed for each independent variable are presented in Table 4.2.

Table 4.2.Variance Inflation Factor corresponding to the continuous independent variables

Variable	VIF	1/VIF (Tolerance)
DAP	6.4	0.156322
UREA	5.96	0.16786
Area size	1.65	0.60687
Mean VIF	4.67	

As a rule of thumb, VIF greater than 10 indicates the existence of multicollinearity or a strong correlation between independent variables. In our case, as the VIF values are all well less than 10, there is no serious problem of multicollinearity.

4.3 Modelling

4.3.1 Analysis of Covariance (ANCOVA)

The results of our ANCOVA are presented in Table 4.4. A significant factor variable suggests the existence of mean difference in the cereal crop yields for the different categories of the factor variable controlling for the effects of the covariates. Conversely, a significant covariate implies that the covariate is significantly contributing to the mean change of cereal crop yields keeping the effect of all other independent variables constant.

Table 4.3: ANCOVA results with all main effects included

Coefficients	Estimate	Std. Error	|t| value	Pr(>|t|)	
(Intercept)	3.24	0.03225	100.34	< 2e-16**	P-value < 2.2e-16 R-Squared= 0.8806 Adj. R-squared = 0.8799 Residual error= .2187, df=1927
Sloppy field orientation	-0.0199	0.017891	-1.121	0.26253	
Somewhat sloppy field orientation	0.001184	0.01064	1.108	0.26803	
Irrigation	-0.00477	0.024238	-0.197	0.84401	
Land degradation	-0.0137	0.019873	-0.693	0.48856	
Seed type	0.102648	0.012773	8.036	1.6e-15**	
Natural Fertilizer	0.0089	0.098452	0.093	0.92568	
Both natural and chemical fertilizer	-0.003747	0.010678	-0.342	0.73257	
Field prevention from erosion	-0.03357	0.01668	-2.013	0.04424*	
Area	0.42687	0.005032	84.657	< 2e-16**	
DAP	0.086861	0.027369	3.17	0.00155**	
UREA	-0.00371	0.026452	-0.133	0.89439	

The above analysis of covariance results indicate that the quantity of DAP and area size are the two covariates that have significant effect on the change in the mean amount of cereal crops yield at 95% level of significance.

Likewise, keeping the effects of all other covariates and factor variables constant, there is statistically significant difference in the average cereal crops yield for the categories of seedling type and field prevention from erosion.
The variables that are not significant have been removed from the model using the stepwise variable selection method and the results are presented in Table 4.4.

Table 4.4: ANCOVA model fit results with only significant main effects

Coefficients	Estimate	Std. Error	|t| value	Pr(>|t|)	
(Intercept)	3.198294	0.023318	137.16	< 2e-16 ***	P-value < 2.2e-16 Multiple R-Squared= 0.8762 Adj R-squared = 0.876 Residual error= .2171, df=2632
Seed type	0.091352	0.011189	8.164	4.956e-16 ***	
Field prevention from erosion	-0.027119	0.013129	-2.066	.039**	
Area	0.419287	0.004235	98.998	< 2e-16 ***	
DAP	0.084226	0.01201	7.013	2.96e-12***	

Table 4.4 provides estimates of parameters of the model for variables having significant effects on the change in the average cereal crops yield. This results in Table 4.4 reveal that there is a statistically significant difference in the mean cereal crops yield between those categories of seed type and the categories of field prevention from erosion at 95% level of significance. Table 4.4 reveals that both harvested area size and the quantity of DAP utilized for cereal crops are the two covariates that have significant effect on the mean amount of cereal crops yield. The variation in cereal crop yield explained by the significant main effect variables is about 87.6 percent.
When the interaction of the two significant covariates included in the model was tested for significance, field prevention from erosion has been found to be not significant and the outcome is presented in Table 4.5.

Table 4.5: ANCOVA model fit results involving interaction effect

Coefficients	**Estimate**	**Std. Error**	**|t| value**	**Pr(>|t|)**	
(Intercept)	3.101203	0.026448	117.258	<2e-16***	P-value < 2.2e-16 Multiple R-Squared= 0.8769 Adj R-squared = 0.8767 Residual error= .2171, df=2632
Seed type	0.090304	0.011159	8.092	8.84E-16	
Area	0.396788	0.006745	58.825	<2e-16***	
DAP	0.1558	0.020212	7.708	1.80e-14***	
Area:DAP	0.0256555	0.005879	4.364	1.33e-05***	

Table 4.5 shows the significant variables and their interaction influencing the average cereal crops yield. In this model, the type of seed used, area size, the quantity of DAP, and the interaction between area size and quantity of DAP have significant effect on the average cereal

crops yield at 5% level of significance.

The multiple coefficients of determination imply that about 87.67 percent of the variation in the amount of cereal crops yield is explained by this model. This value shows a slight increase in the amount of variation in the cereal crops yield explained by this model when compared to the main effect model given in Table 4.4.

As the variation in the mean amount of cereal crops yield is better explained by Model 2 (model in Table 4.5 with R-square .8767 and AIC = -580.8165) compared to model 1(model in Table 4.4 with R-square = .876 AIC = -566.0758), the second model (Model 2) is the preferred and final model and interpretation of results will be made using it.

4.3.2 Dummy Variable Multiple Linear Regression Analysis

4.3.2.1 Intercept-only model

The intercept-only model was fitted with the results shown in Table 4.6.

Table 4.6. Intercept-only model fit results

<table>
<tr><th>Source</th><th>Sum of square</th><th>Degree of freedom</th><th>Mean Square</th><td colspan="3" rowspan="4">Number of Observations=29,223
F(0, 29222)
Prob> F = 0.000
R-Squared= 0.00
Adj R-squared = 0.000
Root MSE = 0.64474</td></tr>
<tr><td>Model</td><td>0</td><td>0</td><td></td></tr>
<tr><td>Residual</td><td>12147.2309</td><td>29,222</td><td>0.41569</td></tr>
<tr><td>Total</td><td>12147.2309</td><td>29,222</td><td>0.41569</td></tr>
<tr><th>Yield</th><th>Coef.</th><th>Std. error</th><th>t</th><th>P > |t|</th><th colspan="2">[95% conf. Interval</th></tr>
<tr><td>Constant</td><td>2.242</td><td>0.0037716</td><td>594.41</td><td>0</td><td>2.2344</td><td>2.2492</td></tr>
</table>

This fitted model with only intercept indicates that the average amount of cereal crops yield measured in kilograms in Ethiopia for the year 2019/2020 is $Y_i = 10^{2.242}$ which is 174.58 kilograms when the effects of all independent variables are removed from the model.

4.3.2.2 Full/final model

A multiple linear regression model has been fitted after removing the non-significant variables using stepwise variable selection method.

In this stepwise variable selection procedure, a combination of both forward elimination and backward variable selection procedures have been used. The results of fitted model is presented in Table 4.7.

Table 4.7: Multiple linear regression model fit results

Source	Sum of square	Degree of freedom	Mean Square			
Model	878.511	4	219.629	Prob> F = 0.000 R-Squared= 0.8769 Adj R-squared = 0.8767 Root MSE = 0.21645		
Residual	123.311	2,632	0.0485			
Total	1001.83	2,636	0.3801			
Vaiable/factor	**Coefficient**	**Std. error**	**t**	**P > |t|**	**[95% conf. Interval**	
Seed type	0.0903	0.0111	8.09	<0.0001*	.0684	.1122
Dap	0.1558	0.0202	7.71	<0.0001*	.1162	.195
Area size	0.3968	0.007	58.82	<0.0001*	.384	.41
Dap:Area size	0.0257	0.006	4.36	<0.0001*	.0141	0.0372
Constant	3.1012	0.0264	117.26	<0.0001*	2.2344	2.2492

The results in Table 4.7 depict that only one categorical independent variable and two covariate variables significantly affect average cereal crops yields. Moreover, the interaction of DAP and area size also have a significant effect on the mean cereal crops yields.

Since the data for this study is log transformed to meet the assumption of the multiple linear regression model, the dummy variable multiple linear regression equation of the full model is written as:

$$log(Y_i) = 3.101 + .0903\theta_{1i} + .156log(x_{2i}) + .397log(x_{3i}) +.0257log(x_{2i}x_{3i}) \quad where,$$

The value of adjusted R^2= .8767 indicates that about 87.67% of the variation in the average cereal crop yield is explained by the explanatory variables included in the model.

The intercept term 3.101 can be interpreted as the amount of cereal crops yield by removing the effect of all independent variables from the model i.e. $\log_{10}(Y_i) = 3.101$ the expected average cereal crops yield in Ethiopia is $Y_i = 10^{3.101} = 1261.83$ when the effect of all variable removed from the model.

The coefficient of the dummy coded factor variable “seed type” 0.0903 is interpreted by taking the antilogarithm of the coefficient value i.e. $Y_i = 10^{0.903} = 1.23$ by keeping the effect of all other explanatory variables constant. The value 1.23 indicates that the average cereal crops yield per hectare increases by 1.23kg when using “improved seed” compared to using non- improved seed.

The coefficient, 0.156 for the “quantity of DAP” shows that an increase in DAP by 1kg will result in an increase of log average yield by 0.156 or 15.6%.
Therefore, a one percent change in the quantity of DAP measured in kilogram used is associated with 15.6% change in the amount of cereal crop yield in Ethiopia.

Likewise, the coefficient (0.397) for “area size” shows that an increase in area size by 1 hectare will result in an increase of log average yield by 0.0.397 or 39.7%.

Therefore, a one percent change in area size is associated with 39.7% change in the amount of cereal crop yield in Ethiopia.

4.3.2.3 Cereal crops yield regression modeling for each Region

In order to see whether the same variables contribute to the change in average cereal crop yield, a separate regression model was fitted for each region using the stepwise variable selection method as in the case for the overall full fitted model presented in Table 4.7.

Tigray Region:
The fitted model is

$$log(Y_i) = 3.08 + 0.032\tau_{2i} + .019log(x_{2i}) + .398log(x_{5i}) + .034log(x_{2i}x_{5i})$$

The variables contributing to the mean amount of change in cereal crop yields in Tigray region are, Fertilizer type used, the quantity of DAP, area size, and the interaction of both covariates.
If the effects of all variables are kept constant in the model, by taking antilogarithm both sides the estimated average cereal crop yield (per hectare) in Tigray region is about 1202 kilograms for the year 2019/2020. And if the effect of all other variables kept constant using both natural and chemical fertilizers together results in 1.076 kilogram increment of cereal crops yield when compared to using only chemical fertilizers.
While using fertilizer DAP results in an increase of 1.9 percent cereal crops yield in Tigray keeping the effect of all other variables constant.

Amhara Region:
The fitted model is

$$log(Y_i) = 3.03 + 0.032\tau_{2i} + .0142log(x_{2i}) + .406\ log\ (x_{5i})$$

The variables contributing to the mean change in the amount of cereal crop yields in Amhara region are, Fertilizer type used, the quantity of DAP, and area size.
If the effects of all variables are removed from the model, the expected average cereal crop yield (per hectare) in Amhara region is about 1071 kilograms for the year 2019/2020. And if the effect of all other variables kept constant using both natural and chemical fertilizers together results in 1.076 kilogram increment of cereal crops yield when compared to using only chemical fertilizers.
While using fertilizer DAP results in an increase of 14.2 percent cereal crops yield in Amhara region keeping the effect of all other variables constant.

Oromia Region:

The fitted model is:

$$log(Y_i) = 3.233 - .69\tau_{2i} + .0699log(x_{2i}) + .396log(x_{5i})$$

The variables contributing to the mean amount of change in cereal crop yields in Oromia region are, Fertilizer type used, the quantity of DAP, and area size.

If the effects of all variables are removed from the model, the expected average cereal crop yield in Oromia region is about 1710 kilograms for the year 2019/2020. If the effect of all other variables kept constant using both natural and chemical fertilizers together results in 4.9 kilogram increment of cereal crops yield in Oromia region when compared to using only chemical fertilizers.

And using fertilizer DAP results in an increase of 6.99 percent cereal crops yield in Oromia region keeping the effect of all other variables constant.

Benishangul Gumuz Region:

The fitted model is:

$$log(Y_i) = 3.41 + 0.898\theta_{1i} + .011\rho_{1i} + .42log(x_{5i})$$

The variables contributing to the mean amount change in cereal crop yields in Benishangul Gumuz region are seed type, fertilizer use, and area size.

If the effect of all variables is removed from the model, the average cereal crop yield in Benishangul Gumuz is estimated to be about 2570 kilograms for the year 2019/2020.

If the effect of all other variables kept constant using improved seed results in 7.9 kilograms increment when compared to using non-improved seed type for cereal crops and an increase of 1.03 kilograms of cereal crops yield if fertilizers are used for harvesting keeping the effect of all other variables constant.

SNNP Region:

The fitted model is:

$$log(Y_i) = 3.065 - .0274\gamma_{2i} + 0.22\theta_{1i} + .128log(x_{1i}) + .397log(x_{5i}))$$

The variables contributing to the mean amount of change in cereal crop yields in SNNPR region are field orientation, seed type, the quantity of UREA used, and area size.

If the effects of all variables are removed from the model, the expected average cereal crop yield in the SNNP region is about 1161 kilogram for the year 2019/2020.

In SNNP regions there is about 1.07 kg reduction in cereal crops yield keeping all other variables constant in the model when somewhat sloppy field orientation harvested compared to flat field orientation. And 12.8 percent increase in the yield amount keeping the effect of all other variables constant and when using the fertilizer UREA.

There is also a 1.67 kilograms of cereal crops yield increase when using improved seed instead of non-improved seed type, and keeping the effect of all other variables constant in SNNP region.

Gambella Region:

The fitted model is:

$$log(Y_i) = 3.36 + .445log(x_{5i})$$

The variables contributing to the mean amount change in cereal crop yields in Gambella region is area size only.

If the effect of all variables is removed from the model, the average cereal crop yield in Gambella region is about 2290 kilogram for the year 2019/2020.

And in Gambella region only the size of harvested land would has a significant effect on the average amount of cereal crops yield change.

Harari Region:

$$log(Y_i) = 3.27 + .448log(x_{5i})$$

The variable contributing to the mean amount change in cereal crop yields in Harari region is area size only. If the effect of all variables is removed from the model, the average cereal crop yield in the Harari region is about 327 kilograms for the year 2019/2020.

As in the case of Gambella region, also in the Harari region only the size of harvested land would has a significant effect on the average amount of cereal crops yield change.

Dire Dawa City Administration

$$log(Y_i) = 3.153 + 0.33\lambda_{1i} + .446log(x_{5i})$$

The variables contributing to the mean amount change in cereal crop yields in Dire Dawa are irrigation, area size, and the interaction of both variables.

If the effect of all variables is removed from the model the average cereal crop yield in Dire Dawa is about 1422 kilograms for the year 2019/2020.

The factor variable, irrigation, significantly affects the average cereal crops yield amount in the Dire Dawa city administration. There is about 2.14 kilograms increase in the amount of cereal crops yield when irrigation is used compared to cereal crops yield in the non-irrigated area, keeping the effect of all other variables/factors constant.

The fitted separate models for the regions depict that the average cereal crop yield across all studied regions is affected by the area size of the harvest, and in most regions, the quantity of DAP used for harvest also had a significant effect on the average cereal crop yields.

4.4 Model Diagnostics

Model diagnostics have been undertaken only for the final fitted model for all regions considered together.

4.4.1 Homogeneity of covariate regression coefficients

The assumption of homogeneity of regression slopes has been tested in the context of ANCOVA. The test was performed for each covariate at each factor level separately.

If the interaction effect of "area size" and "seed type" is found statistically significant, we say that the model fails to meet the assumption of homogeneity of regression slopes. The constant slope test results are presented in Table 4.8. The results for the tests of between-subjects effects are presented in Table 4.8.

Table 4.8: Area size regression slope test

Tests of Between-Subjects Effects

Dependent Variable: Log transformed Yield amount in KG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	10378.582[a]	3	3459.527	5.657E4	.000
Intercept	41127.573	1	41127.573	6.725E5	.000
Seed type	32.163	1	32.163	525.944	.000
Area log hectares	4950.651	1	4950.651	8.096E4	.000
Seed type * Area log hectares	.137	1	.137	2.237	.135
Error	1786.994	29222	.061		
Total	159027.487	29226			
Corrected Total	12165.576	29225			

a. R Squared = .853 (Adjusted R Squared = .853)

The non-significant F test for the interaction term implies that the regression slopes of both improved and non-improved seed with respect to the association between the covariate which is area size and the response variable which is "cereal crops" yield are equal within sampling fluctuation.

The results of the tests of between-subjects effects are presented in Table 4.9. The test results have been used for checking the plausibility of the assumption of regression slopes between the seed type used with respect to the association between "quantities of DAP" used and "cereal crop yields" measured in kilograms.

Table 4.9: Quantity of DAP regression slope test

Tests of Between-Subjects Effects

Dependent Variable: Log transformed Yield amount in KG

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	426.745[a]	3	142.248	634.953	.000

Intercept	617.207	1	617.207	2.755E3	.000
Seed type	.725	1	.725	3.238	.072
Dap	291.252	1	291.252	1.300E3	.000
Seed type * Dap	.780	1	.780	3.481	.062
Error	590.318	6635	.224		
Total	13962.530	6639			
Corrected Total	1017.063	6638			

a. R Squared = .420 (Adjusted R Squared = .419)

The non-significant F test results imply that the slopes of both improved and non-improved seed with respect to the association between quantities of DAP used and cereal crops yield are equal within sampling fluctuation.

4.4.2 Constant error variance

The assumption of constant error variance or homoscedasticity has been tested after fitting the linear regression model. This test was used to determine whether the seed groups have equal error variance in the cereal crops yield controlling for the effect of both the "quantity of DAP" measured in kilograms and "area size" measured in hectares. The result of Leven's test of equality of error variance are presented in Table 4.10, followed by graphical method.

Table 4.10: Levene's test of homoscedasticity

Dependent Variable: Log transformed Yield amount in KG

F	df1	df2	Sig.
2.339	1	2637	.126

The null hypothesis of this test is that the error variance of the dependent variable is equal across groups. Leven's test results are that the value of the statistic is F=2.339 and the corresponding p-value is 0.126. Since this p-value is greater than 0.05, we do not to reject the null hypothesis that the error variance of the dependent variable is equal across groups. This implies that the error variance of the dependent variable is equal across the groups of the categorical independent variable i.e. "seed type".

The test of homoscedasticity or constant error variance has also been performed by plotting the predicted values versus individual error as shown in Figure 4.4.

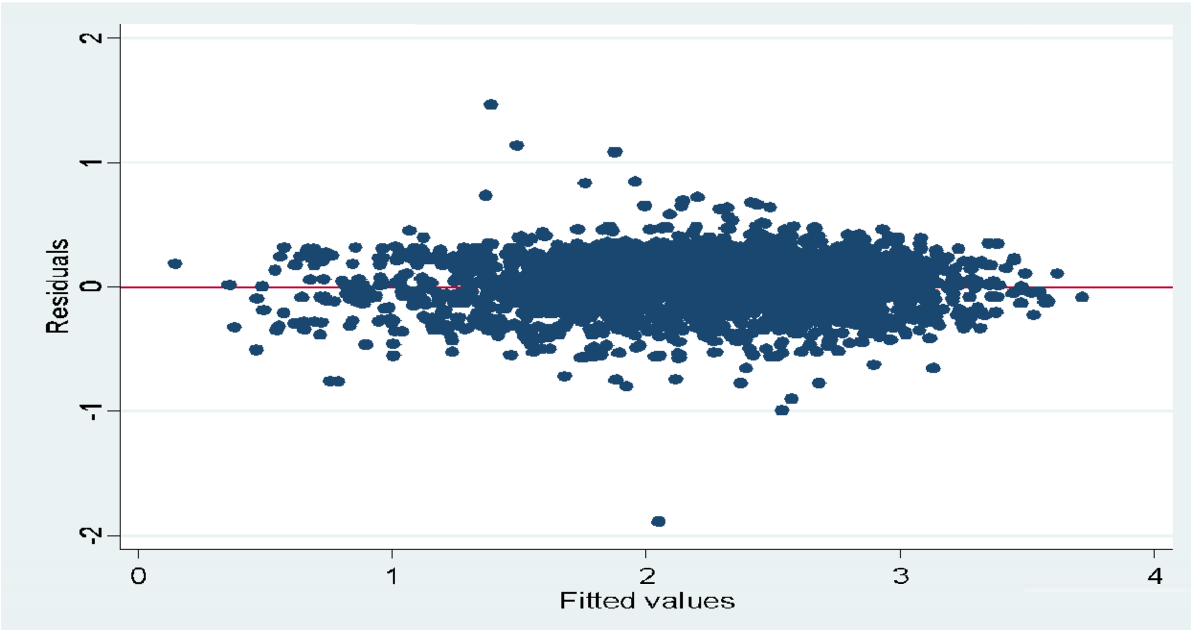

Figure 4.4: Plot of residuals versus fitted values

Figure 4.4 depicts constant variance in the error terms.

4.4.3 Normality

To test the normality assumption of error terms, both graphical and statistical tests methods have been performed.

Z scores from both kurtosis and skewness of cereal crops yield measured in kg were computed separately for the categories of seedling type.

Table 4.11 shows the computed Z score values obtained by dividing kurtosis by its standard error and skewness by its standard error value while Figures 4.5 and 4.6 show two graphical methods for tests of normality: the normal p-p plot and the histogram respectively.

Table 4.11: Z-value computed from skewness and kurtosis

	Seedling type		Statistic	Std.error	Computed z-value
Log transformed Yield amount in KG	Non-Improved	Skewness	-0.0538	0.06	-0.897
		Kurtosis	0.0361	0.031	1.165
	Improved	Skewness	-0.055	0.035	-1.571
		Kurtosis	0.1033	0.69	0.150

As shown in Table 4.11, the computed Z –values lie between −1.96 and 1.96 indicating that the data exhibit the pattern of standard normal distribution.

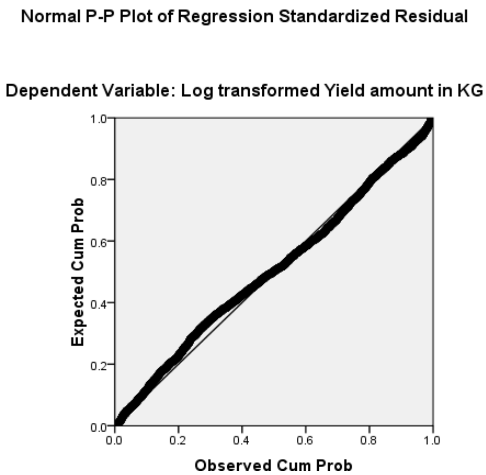

Figure 4.5

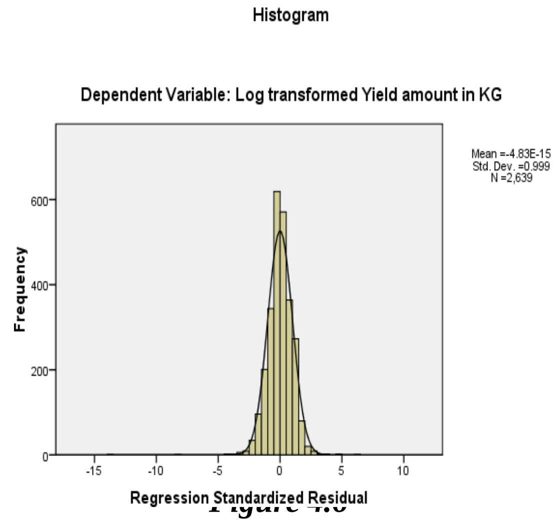

Figure 4.6

The normal probability plot of the residuals in Figure 4.5 is approximately linear implying that the error terms are normally distributed and the histogram of residuals in Figure 4.6 indicate that the residuals (and hence the error terms) are normally distributed.

5. CHAPTER 5

5.1 Discussions

The objective of this research study was to identify the non-climatic factors associated with Ethiopian cereal crops yield and test the existence of statistically significant difference between the factors levels by controlling the effect of covariates.

In the current study ANCOVA and dummy variable regression have been employed to identify non-climatic factors significantly contributing to the average change in the amount of cereal crops yield in Ethiopia and determine if there are statistically significant differences between the factor levels of the categorical predictor variable on the mean response variable (average yield of cereal crops measured in kg). The results of the analyses indicated that some covariates and factors are significantly contributing to the change in average cereal crops yields in Ethiopia.

Among the explanatory variables considered in this study, “quantity of DAP” measured in kg and “area size” measured in hectares are the two covariates that affect the change in average cereal crops yields while “seedling type” is the significant factor which contributed to the change in the average cereal crops yield in Ethiopia using the 2019/2020 agricultural survey data. Thus, all conclusions and recommendations drawn from this study are based on the model which includes the three aforementioned variables and the interaction of the covariates. It was demonstrated that the average cereal crops yield between the categories of seedling type is significantly different.

Consistent with our findings that variables like area size, fertilizers, specifically DAP, and seed type are significantly contributing to the change in the average yield amount in Ethiopia, Albanian J. (2013) argued that there is a strong positive correlation between crop yield and fertilizer indicating that an increase in fertilizer access to farmers will lead to a corresponding increase in yield and also farm size has a positive correlation with yield.

The findings of this study also aligns with the article published by Kassa (2015) that fertilizer and improved seed utilization are among factors that would significantly contribute to crop yield in the Tigray region of Ethiopia, showing that fertilizer type used, area size and quantity of DAP used are the variables that are significantly contributing to the change in the yield amount in the Tigray Region of Ethiopia.

Even though, there is a climatic difference between Ethiopia and Uganda, in agreement to our finding, a study in Uganda by Epule et al. (2018), stated that when farmers have adequate access to fertilizers, their crop yields are likely to increase. Consistent to our finding, Yengoh (2012) found out that improved seeds do significantly improve yields. This finding is also consistent with the finding by Merga and Haji (2019).

The result of this study indicated that using fertilizers specifically DAP for agricultural input would results in an increase in the amount of cereal crops yield by 15.6% percent per one hectare of the harvested land in Ethiopia. This result is consistent with the study by Zare et al. (2013) an increase in fertilizers access to farmers will lead to a corresponding increase in yield.

Despite the fact that there exist a regional variation in the type of factor variables that contributes to the change in the mean amount of cereal crops yield in Ethiopia, the effect of using natural and chemical fertilizers together results in an increased in the amount of cereal crops yields in three regions namely, Tigray, Amhara and Oromia. But, the amount of change in cereal crops yields due to utilization of both natural and chemical fertilizer together leads to the same amount of change in the average cereal crops yield by 1.076 kilograms in Tigray and Amhara regions, while the effect of using this variable leads an increase in average cereal crops yield by 4.9 kilograms in Oromia region.

The factor variables "seed type" and "field orientation" contributes to the change in the average amount of cereal crops yield in SNNP region indicating that harvesting "somewhat sloppy field orientation" results in declining average amount of cereal crops yield in the region when compared to the "flat field orientation". Moreover, using "improved seed" in the region instead of non-improved seed type results in an average increase of 1.66 kilograms of cereal crops yield per hectare.

The factor variable irrigation contributed to the change in the average cereal crops yields in Dire Dawa city administration, there is an increase in average cereal crops yield by 2.14 kilograms per hectare when the field of harvest is irrigated compared to the non-irrigated field.

Across all regions, an increase in "area size" results in the average cereal crops yield increment. For every one hectare increase in the area size, the average cereal crops yield is expected to increase by 39.7 percent.

The factor variables "seed type" and "fertilizer use" contribute to the change in the average amount of cereal crops yield in Benishangul Gumuz region. Using improved seed results in an average increase of 7.91 kilograms of cereal crops yield compared to non-improved seed and using fertilizer results in an average increase of 1.29 kilograms of cereal crops yield per hectare in the region.

The variables that are contributing to the change in the average amount of cereal crops yield in three regions i.e. Tigray, Amhara and Oromia are the same, which are the utilization of fertilizer type, quantity of DAP and area size. However, cereal crops yield amount in Tigray region is also affected by the combined effect of "area size" and "quantity of DAP" used.

The average cereal crops yield amounts in two regions namely, Gambella and Harari are affected by "area size" only.

CHAPTER 6

6. Conclusions and Recommendations

6.1 Conclusions

Beyond its contribution to Ethiopia's largest GDP, agriculture is an irreplaceable source of feeding and industrial input in Ethiopia. In this study, data obtained from the 2019/20 agricultural sample survey conducted by the Central Statistical Agency of Ethiopia were used to assess non-climatic factors associated with cereal crop yield. The study covered a total of 29,231 agricultural households from different regions of Ethiopia. Of these, 29% of agricultural households were from Amhara and another 29% from Oromia. The study results showed that seed type used, DAP, area size, and the interaction of both DAP and area size were statistically contributing to the change in the amount of cereal crop yields in Ethiopia,

This study indicated regional variation in the factors contributing to the changes in the amount of cereal crop yields. The use of DAP fertilizer and seed type are the two major factors influencing the changes in the amount of cereal crop yields in Ethiopia.

The results of the study also revealed that the area size harvested affects the mean change in the amount of cereal crop yields significantly across all regions except Afar and Somali regions of Ethiopia.

6.2 Recommendations

The Ministry of Agriculture and all engaged stakeholders have to create awareness among farmers across all regions regarding the importance of improved seed and fertilizer since it is clear that fertilizers and improved seeds do have impact on crop yield.

Hence, standardized utilization methods for fertilizers specifically for DAP and improved seeds should be established so that cereal crop yield is increased.

References

Alemayehu Seyoum Taffesse, Paul Dorosh and Sinafikeh Asrat, 2011. Crop Production in Ethiopia: Regional Patterns and Trends. Ethiopia Strategy Support Program II (ESSP II) ESSP II Working Paper No. 0016 March 2011

Agricultural Growth in Ethiopia (2004-2014): Evidence and Drivers. IFPRI Working Paper 81. October 2016, p. 5

Bodirsky BL, Rolinski S, Biewald A, Weindl I, Popp A, Lotze-Campen H. Global food demand scenarios for the 21st century. PLoS One 2015; 10(11): e0139201. [http://dx.doi.org/10.1371/journal.pone.0139201]

Bulti Merga & Jema Haji (2019). Factors Impeding Effective Crop Production in Ethiopia. Journal of Agricultural Science; Vol. 11, No. 10; 2019 CIMMYT, Poor soils a huge limitation for Africa's food security," 2015 DOI:10.5539/jas.v11n10p98

H. Charles J. Godfray, John R. Beddington, Ian R. Crute, Lawrence Haddad, David Lawrence, James F. Muir, Jules Pretty, Sherman Robinson, Sandy M. Thomas, Camilla Toulmin , 2010. Food Security: The Challenge of Feeding 9 Billion People

Chong-Ho Wang (2014) Farming Methods Effects on the Soil Fertility and Crop Production Under a Rice-Vegetables Cropping Sequences, Journal of Plant Nutrition, 37:9, 1498-1513, DOI: 10.1080/01904167.2014.881876

CSA (Central Statistics Agency). 2016. Agricultural Sample Survey 2015/2016 (2008 E.C.). Volume I. Report on Area and Production of Major Crops (Private Peasant Holdings, Meher Season). Statistical Bulletin 584, Addis Ababa

Denning G, Kabambe P, Sanchez P, Malik A, Flor R, Harawa R, Nkhoma P, Zambia C, Banda Magombo C. Inputs subsidies to improve smallholder maize productivity in Malawi: towards an African green revolution. PLoS Biol. 2009;7:e1000023.

Epule, T.E., Ford, J.D., Lwasa, S. *et al.* The determinants of crop yields in Uganda: what is the role of climatic and non-climatic factors?. *Agric& Food Secur* **7,** 10 (2018). https://doi.org/10.1186/s40066-018-0159-3/

FAO. 2017. The future of food and agriculture – Trends and challenges. Rome.

FAO. 2010. Investment Centre Database of Projects. Rome, Food and Agriculture Organization of the United Nations.

Fayera Deresa and Tsetadergachew Legesse, 2015. Cause of Land degradation and its impacts on livelihood of the population in Toke kutaye woreda, International Journal of Scientific and Research Publications.

Getachew Ahmed Abegaz: Cereal productivity in Ethiopia: Ethiopian Journal of Economics, Volume XX No. 2, October 2011

Godfray HCJ, Beddington JR, Crute IR, Haddad L, Lawrence D, Muir JF, et al (January 28, 2010). Food Security: The challenge of feeding 9 billion people. Science. 327(5967) DOI: 10.1126/science.1185383

Kassa (2015). Factors affecting Agricultural production in Tigray Region, Northern Ethiopia

Kerr, J and J. Pender. 2005. Farmers' perceptions of soil erosion and its consequences in India's semiarid tropics. Land Degradation and Development (in press)

Knox, O. G., Leake, A. R., Walker, R. L., Edwards, A. C., & Watson, C. A. (2011). Revisiting the multiple benefits of historical crop rotations within contemporary UK agricultural systems. Journal of Sustainable Agriculture, 35(2), 163-179. https://doi.org/10.1080/10440046.2011.539128

Linear Models in Statistics (2nd Edition) by A.C Rencher and G.B Schaalje (John Wiley, Inc.), New York

Mesfin, A. H., & Zemedu, L. (2015). Improved Rice Seed Production and Marketing: Challenges and Opportunities; the Case of Fogera District of Ethiopia. Journal of Agriculture and Environmental Sciences, 1(2), 1-21

Negin J, Remans R, Karuti S, Fanzo JC. Integrating a boarder notion of food security and gender empowerment into the African green revolution. Food Security.

Oldfield EE, Bradford MA, Wood SA. Global meta-analysis of the relationship between soil, organic matter and crop yields. The Soil. 2019

Paul C Hay, Extension Educator University of Nebraska-Lincoln Extension in Gage County, 2016. Top Seven Factors in Crop Production.

Reta Hailu Belda, 2014. Population dynamism and agrarian transformation in Ethiopia. Addis Ababa University, College of Development Studies, Addis Ababa, Ethiopia

Tandzi Ngoune Liliane and Mutengwa Shelton Charles (July 15th 2020). Factors Affecting Yield of Crops, Agronomy - Climate Change & Food Security, Amanullah, IntechOpen, DOI: 10.5772/intechopen.90672. Available from: https://www.intechopen.com/chapters/70658

Thirtle C, Irz X, Lin L, Mckenzie-Hill V, Wiggins S. Relationship between changes in agricultural productivity and the incidence of poverty in developing countries. In: DFID Report No. 7946. 2001

United Nations, The Least Developed Countries Report, New York, 2015

Wang C-H. Farming methods effects on the soil fertility and crop productionn under a rice—Vegetables cropping sequences. Journal of Plant Nutrition. 2014;37:1498-1513. DOI: 10.1080/01904167.2014.881876

World Bank Group, Development Research Group Poverty and Inequality Team, November 2014.

Yadaw, R. L., Dwivedi, B. S., & Pandey, P. S. (2000). Rice-wheat cropping system: Assessment of sustainability under green manuring and chemical fertilizer inputs. Field Crops Research, 65(1), 15-30. https://doi.org/ 10.1016/S0378-4290(99)00066-0

Yengoh, G.T. Determinants of yield differences in small-scale food crop farming systems in Cameroon. *Agric& Food Secur* **1,** 19 (2012). https://doi.org/10.1186/2048-7010-1-19

ZARE M., SIASINEJAD M , GHASEMI S AND AZARPANAH (2013). Various irrigation cycles effect on grain yield, proline and adaptive metabolits in some wheat genotypes. Albanian j. agric. sci. 2013;12 (4): 653-657

Zerihun Gudeta Alemu ,Lache Oosthuizen and Herman Van Schalkwyk (2002). Agricultural development policies of Ethiopia since 1957**.**South African Journal of Economic History 17(1). DOI: 10.1080/10113430209511142

Appendices

.

YIELD_KG	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
SLOPPY_FLD_~T	23.53244	15.78607	1.49	0.136	-7.408949	54.47383
Land_degrad~n	-28.26941	14.37595	-1.97	0.049	-56.44689	-.0919272
Seed_type	289.9275	13.98109	20.74	0.000	262.524	317.3311
Corp_damage	0	(omitted)				
IRRIGATION_	-147.5289	38.07418	-3.87	0.000	-222.1559	-72.9018
FERTILIZER	-4.605891	14.84918	-0.31	0.756	-33.71094	24.49916
Natural_fer~r	-235.9728	13.47709	-17.51	0.000	-262.3885	-209.5571
Both_fertil~e	-236.6067	16.70253	-14.17	0.000	-269.3444	-203.869
FALLOW_LAND	-312.5665	83.30508	-3.75	0.000	-475.8481	-149.2849
CROPFILED_L~D	-413.603	76.43397	-5.41	0.000	-563.4169	-263.789
VIRGIN_LAND	-191.6193	169.9569	-1.13	0.260	-524.7423	141.5037
_cons	938.8685	76.6877	12.24	0.000	788.5572	1089.18

```
. regress Yield_trans SLOPPY_FLD_ORIENT Land_degradation Seed_type Corp_damage I
> RRIGATION_ FERTILIZER Natural_fertilizer Both_fertilizer_type FALLOW_LAND CROP
> FILED_LAND VIRGIN_LAND
note: Corp_damage omitted because of collinearity
```

Source	SS	df	MS			
				Number of obs	=	29,758
				F(10, 29747)	=	317.06
Model	6605.57996	10	660.557996	Prob > F	=	0.0000
Residual	61974.5936	29,747	2.08338971	R-squared	=	0.0963
				Adj R-squared	=	0.0960
Total	68580.1736	29,757	2.30467364	Root MSE	=	1.4434

Yield_trans	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
SLOPPY_FLD_~T	.073012	.0265368	2.75	0.006	.0209987	.1250252
Land_degrad~n	-.0722227	.0241662	-2.99	0.003	-.1195896	-.0248558
Seed_type	.4761834	.0234993	20.26	0.000	.4301238	.5222431
Corp_damage	0	(omitted)				
IRRIGATION_	-.5152349	.0639942	-8.05	0.000	-.6406663	-.3898035
FERTILIZER	.2351413	.0249613	9.42	0.000	.1862162	.2840665
Natural_fer~r	-.9674002	.0226524	-42.71	0.000	-1.0118	-.9230005
Both_fertil~e	-.6397822	.0280735	-22.79	0.000	-.6948074	-.5847569
FALLOW_LAND	-.5229235	.1400171	-3.73	0.000	-.797363	-.2484839
CROPFILED_L~D	-.6425672	.1284684	-5.00	0.000	-.894371	-.3907635
VIRGIN_LAND	-.2501344	.2856592	-0.88	0.381	-.810039	.3097702
_cons	5.945748	.1288953	46.13	0.000	5.693108	6.198388

.