# ADDIS ABABA UNIVERSITY
# OFFICE OF GRADUATE PROGRAM

## FACULTY OF SCIENCE
## DEPARTMENT OF STATISTICS

# REGIONAL DIFFERENCE AND DETERMINANTS OF CEREALS PRODUCTION IN ETHIOPIA

BY

LEGESSE ABATE

A THESIS SUBMITTED TO THE OFFICE OF GRADUATE PROGRAMMES OF
ADDIS ABABA UNIVERSITY, IN PARTIAL FULFILLMENT FOR THE
DEGREE OF MASTER OF SCIENCE IN STATISTICS

**August 2009**
**Addis Ababa**

# ADDIS ABABA UNIVERSITY
# OFFICE OF GRADUATE PROGRAM

## FACULTY OF SCIENCE
## DEPARTMENT OF STATISTICS

# REGIONAL DIFFERENCE AND DETERMINANTS OF CEREALS PRODUCTION IN ETHIOPIA

BY

LEGESSE ABATE

Approve by the Board of Examiners:

| | |
|---|---|
| **Sileshi Fanta (Asst. Prof.)** | |
| **Department Chairman** | **Signature** |
| **Eshetu Wencheko (Prof.)** | |
| **Examiner** | **Signature** |
| **Fentaw Abegaz (PhD)** | |
| **Examiner** | **Signature** |

## LIST OF ABRIVIATIONS

AgSS – Agricultural Sample Survey

AISCO – Agricultural Inputs Supply Corporation

ADLI – Agricultural Development Led Industrialization

ARDU – Arsi Regional Development Unite

CADU – Chilalo Agricultural Development Unit

CSA – Central Statistical Agency

CV – Coefficient of Variation

DAP – Di-Ammonium Phosphate

EAs – Enumeration Areas

ESE – Ethiopian Seed Enterprise

GDP - Gross Domestic Product

LSDV – Least Square Dummy Variables

MEDaC – Ministry of Economic Development and Cooperation

MOFED – Ministry of Finance and Economic Development

NMA – National Meteorological Agency

OLS – Ordinary Least Squares

PPS – Probability Proportional to Size

PSUs – Primary Sampling Units

SSUs – Secondary Sampling Units

UREA – Ammonium Nitrate

## Acknowledgements.

I wish to express my sincere thanks to my advisor, **Dr. Emmanuel Gebreyohannes**, for his constructive comments and suggestions throughout the study. Special thanks also to Ato Jemal Abdi, Ato Seid Jemal and Ato Habekristos Beyene, who have facilitated and supported in materials provision throughout the study.

Finally, I wish to express my heartfelt gratitude to all my friends and my family for their continuous support and encouragement.

# ABSTRACT

This study examines the cereals production technologies and the agro-ecological differences in selected regions of Ethiopia (1995-2008). The questions before the analysis of the data are whether the existing technologies of cereals production used by farmers, regional differences and time effects have significant impacts on cereal production or not. Panel data regression analysis is utilized and various panel data estimators tested. The paper indicates that, new technological inputs (fertilizer, improved seed varieties, pesticides), regional and time effects have statistically significant impacts on cereals production.

# TABLE OF CONTENTS

PAGE

**LIST OF ABRIVIATIONS** I

**AKNOWLEDGMENTS** II

**ABSTRACT** III

**CHAPTER 1. Introduction** 1

1.1 Background 1

1.2 Statement of the problem 6

1.3 Objectives 7

1.3.1 General Objectives 7

1.3.2 Specific Objectives 7

1.4 Significance of the study 7

1.5 Limitations of the study 8

1.6 Organization of the study 8

**CHAPTER 2. Literature Review** 9

**CHAPTER 3. Materials and Methodology** 19

3.1 Source of Data 19

3.2 Variables included in the study 19

3.3 Methodology 20

3.3.1 Introduction 20

3.3.2 Model specification 23

3.4 One-way Error Component Regression 23

3.4.1 Fixed Effects Regression Model 23

3.4.2 Random Effects Model 27

3.5 Testing the Significance of Fixed Effects 30

3.7 Testing the significance of Random Effects 31

3.8 The Hausman Specification Test 32

3.9 Tests for poolability of the data 33

3.10 Assumptions about this Analysis 34

**CHAPTER 4.Statistical Data Analysis** 37

4.1 Descriptive Statistics 37

4.2. Estimation of Fixed effects regression model 41

4.3. Estimation of Random effects regression model 44

4.4. Testing the Hypothesis 45

4.5. Examination of the assumptions 45

**CHAPTER 5. Summary and Conclusion** 50

REFERENCES 52

Annex 1: P-P and Q-Q plots for original data 53

Annex2: P-P and Q-Q plots for transformed data 56

Annex 3: Graphs of cereal production by year and region 56

Annex 4: Descriptive statistics for wheat production 58

Annex 5: Descriptive statistics for maize production 60

Annex 6: Plots of estimated values versus residuals 62

# CHAPTER ONE

## 1. Introduction

### 1.1 Background

The sound performance of agriculture warrants the availability of food crops. The prime role that agriculture plays in a country's political, economic and social stability makes measures of agricultural production extremely sensitive. This accomplishment in agriculture does not only signify the adequate acquisition of food crops to attain food security, but also heralds a positive aspect of the economy (*CSA, 2008*).

Agriculture remains to be the base of Ethiopian economy. In 1960's agriculture accounted for about 65 percent of the GDP. Recently it accounts nearly for 45.9 percent of the GDP, and for about 60 percent of exports. Agriculture provides raw materials for 70 percent of the country's large and medium sized agro-Industries. About 80 percent of the population depends on agriculture for its livelihood. It is growing by 9.4 percent *(National Bank of Ethiopia, 2006/2007).* It is estimated that 16.5 million hectares of land is under cultivation and grains are the most important field crop occupying 86 percent of area planted and being the chief element in the diet of most Ethiopians. The peasant agriculture has always remained the single important source of food in the country. However, in spite of its important contribution to the economy at large, the performance remains disappointing.

According to world development report of 1994, the annual growth rate of agriculture in the country was 1.9% between 1970-1980. The growth rate declined to 1.2% in the years 1980- 1992. In a similar fashion, the contribution of the sector to GDP was 56% in 1970 but only 48% in 1992.

Prior to the mid 1960's, Ethiopian agriculture was virtually unaffected by any modernization scheme. Improved farm inputs such as fertilizer, improved seed variety, insecticides, herbicides and improved farm tools were unknown in earlier dates. Some even argued that the country's conservative religious institutions, historical isolations and feudal structure had created an inward-looking peasantry not for agriculture innovation and development ( Cohen,1987, Pickett 1991). But since the mid 1960's the country has implemented programs in small holder agriculture with the objectives of providing services such as agronomic research, dissemination of research results, supplying improved inputs, provision of agricultural credit, marketing services, training of agricultural staff and developing infrastructure. To implement the programs different comprehensive integrated packages projects were initiated. These include the establishment of Chilalo Agricultural Development Unit (CADU) later called Arsi Regional Development Unit (ARDU).

The crop production sector accounts for about 60 percent of total agricultural out put (Seifu,1998). Generally speaking the farming system in Ethiopia falls into three major categories. These are

- Small holder mixed farming
- Pastoral agro-pastoral and
- Commercial farming

Small holder mixed farming contributes 96 percent of total annual crop production. The land area cultivated by small holders constitutes 98 percent of total cultivated land (Eshete, 1995).

Ethiopia has diversified agro-ecological and climatic conditions. The altitude ranges from as low as 180 meters below sea level to over 4600 meters above sea level. As a result of variability in altitude, Ethiopia experiences diverse climatic conditions and mean annual rainfall ranging from less than 200 mm in lowland areas to 2200mm in the south west highlands. Despite the rich resource base and suitable conditions for agriculture, Ethiopia has suffered reduced crop production and food insecurity since the 1970's (Benin and Echui, 2003). Recognizing the seriousness of the food problem in the country and the necessity of improving agricultural productivity and food security, the Federal Democratic Republic of Ethiopia initiated a broad based Agricultural Development Led Industrialization (ADLI) strategy in the early 1990's. The strategy concentrates on accelerating growth through focusing on the supply of fertilizer, improved seeds, and other inputs.

Through the implementation of ADLI, it is expected that domestic food supply would improve both quantitatively and qualitatively assuring the rural population adequate food and providing the urban population with sufficient food at affordable prices and raw materials. This could be achieved through improving the productivity of the existing land, labor and capital resources, through improved agronomic practices and through the introduction of improved new technology.

To attain the primary objective of food self sufficiency, agricultural research and extension has been geared to focus on the problems of small holder agriculture and to provide economically viable packages of technology to small holders operating in different farming systems. In order to disseminate the knowledge and the skills of using improved technology, field level extension workers are placed in each peasant association. The close proximity of extension workers to farming community may help to make inputs such as fertilizer, improved seeds and pesticides easily available to farmers. This is supposed to allow farmers to have unrestricted access to rural credit and extension services that are intended to increase the productivity of farmers.

From the gross production data of cereal crops, it is observed that the 2008 production of cereals is increased by more than twice compared to the 1998 cereals production. This may indicate the success of the effort made by the government policy (CSA, 2008).

Ethiopia grows large varieties of crops which include cereals (teff, corn, wheat, barley, sorghum, millet, oats, etc.); pulses (horse beans, chick-peas, haricot beans, field peas, lentils, soybean, etc); oilseeds (linseed, nigerseed, fenugreek, noug, rapeseed, sunflower, castor bean, groundnuts, etc.), stimulants (coffee, tea, chat, tobacco, etc.), fiber (cotton, sisal, flax, etc.), fruits (banana, orange, grape, papaya, lemon, menderin, apple, pineapple, mango, avocado, etc.), vegetables (onion, tomato, carrot, cabbage, etc.), root and tuber (potato, enset, sweet-potatoes, beets, yams, etc.) and sugarcane. This variety of crops is dominated by the production of cereals. Cereals account for over 90 percent of total major crop production. In terms of employment, over 80 percent of the rural population concentrates on cereal production. Cereals constitute the major proportion of the population diet. After all, Ethiopia is a country of cereals. The dominance of food grain production is also reflected on state farms. Over 80 percent of the cultivated area of state farms is used for the production of cereals mainly wheat, maize, and barley. Each region has its own contribution due to different capacity toward crop production.

Tigray, the northernmost region of Ethiopia bordering Sudan and Eritrea, has a cultivated area of about 820 000 ha farmed by some 775 000 households and 406 investors, the latter are located in the western lowlands. Usually classified as a food-deficit area due to its semi-arid climate and high population density, the region has embarked on major environmental rehabilitation programmes over the past ten years. Presently, it is in the process of linking food security issues to watershed management with the objective of improving employment and income generation opportunities in the central and eastern zones. The food deficit status of the region masks the fact that in most years there is surplus crop production from well organized, run-off based, peasant farming systems in the South Zone and from the fore-mentioned mechanized commercial enterprises in the western lowlands. It contributes about 7.34 percent of the total grain production of the country.

Amhara Region, located in the north, north-west of the country includes the nation's highest mountain ranges, lowland riverine valleys and plains as well as agriculturally productive plateaux with well-established mixed farming systems. Comprising 10 administrative zones, the region mostly produces around 33 percent of the national *meher* grain production. Following the national pattern of rainfall distribution and notwithstanding the within-zone vagaries of altitude, the western half of the region usually produces surplus grains from a substantial *meher* crop. The eastern half of the region has a less reliable *meher*.

Oromiya, comprising 14 administrative zones, is the largest region in the country extending in a "T" shaped landmass from near the Sudanese border in the west, across central Ethiopia near the eastern border with Somalia and southwards to the border with Kenya. It includes the most productive highland plateaux as well as drought-prone valley bottoms and lowland plains and mostly produces some 54 percent of the nation's cereals and pulses. In six of the southern zones a bimodal rainfall pattern is readily identifiable, usually providing a prolonged growing season and a wide range of cropping options. In the densely populated, high rainfall zones, the small size of peasant land holdings necessitates production of two or three crops annually from the same land if household needs are to be met. This places the farm families in a vulnerable position as the loss of a crop in a series cannot be compensated by increasing the area of the next crop in the sequence and increases the importance of the timeliness of operations at field level. It contributes about 48.40 percent of the total grain production of the country.

Southern Nations Nationalities and Peoples' Region (SNNPR), presently formed from 13 zones and 7 special *woredas*, is the most culturally diverse region in Ethiopia. The cultural diversity is matched by a wide range of agro-ecologies encompassing everything from rainforests to deserts. Bi-modal rainfall patterns exist throughout the region offering opportunities to crop two or three times per year on the same piece of land. Very small land holdings, however, create a structural vulnerability to dry spells at crucial times in the production cycles, as increased planting later in the year cannot easily compensate for lost opportunities. Such problems arise with the failure of *belg* rains. Fortunately, the majority of the rural population in SNNPR grows and eats enset. This perennial carbohydrate source, also known as false banana, is very resistant to rainfall fluctuations and provides a carbohydrate-based food safety-net for most farm families in the highland and middle altitude communities. The presence of eucalyptus and perennial cash crops including coffee and chat confirm the overall natural resources wealth of SNNPR in all but the lowland localities, where pastoralism is the main agricultural enterprise. SNNPR contributes about 7.96 percent of the total grain production of the country.

## 1.2 Statement of the problem

The agricultural sector in Ethiopia is characterized by farm fragmentation, stagnation in growth and instability due to climatic variation and other factors. The stagnation of agricultural production can be due to different factors. First farmers cultivate the land continuously due to scarcity of land. The natural fertility of the soil is declining due to this continuous cultivation. The fertility needs to be maintained by replacing the nutrients which are taken out of the soil. Different studies suggest the possibility of increasing agricultural production in Ethiopia by promoting the use of proper technology packages. But the question is, is there full technology package (which consists of high yield variety of improved seed (HYV), chemical fertilizer (urea and dap), insecticides, herbicides, irrigation practices) at regional level? Does regional environmental variation (which includes climate, soil fertility, etc) have a significant impact on crop production in Ethiopia? In light the above, this study attempts to investigate the regional effects and the significance of agricultural technologies used by farmers on crop production in Ethiopia.

## 1.3 OBJECTIVES

### 1.3.1 General Objective

Ethiopia has a complex diversified agro-ecological conditions and different mean annual rainfall. These resulted in different regions having their own potential for crop production. Thus, identifying such areas and investigating determinants of crop production in different regions is an important yardstick to suggest various interventions to enhance their output. The general objective of the study is to assess factors that significantly affect the output of cereals production of small holder farmers and to suggest appropriate policy measures to be considered by policy makers.

### 1.3.2 The specific objectives of this study are:

a) To compare and contrast the production of cereals of different regions.
b) To determine factors affecting cereals production.
c) To assess cereal production by region and year.
d) To asses appropriate panel model for cereal production.

## 1.4 Significance of the study

This study envisages making a useful empirical contribution on the following aspects:

1. To provide pertinent information that will be useful to enhance cereals production.

2. To contribute for further research in the area or regions of production.

## 1.5 Limitations of the study

This study focuses on the impact of regional variation and new technologies on crop production in Ethiopia. This thesis is limited to see the effect of the above factors on cereals production due to a number of reasons.

- Cereals are generally considered as the most important grain crops in Ethiopia.
- Currently cereals cover a large part of cultivated land (71% of the total land allocated to crops).
- Cereals account for the highest share of the total new technologies used by farmers. For example, cereals absorbed 90% of the total fertilizer and 96% of improved seed variety in 2007/2008 (CSA, 2008).

The study is limited to the four regions ( Tigrai, Amhara, Oromiya and SNNP) due to their high share (97%) of the total cereal production in Ethiopia.

## 1.6 Organization of the study

This study contains five chapters. Chapter one presents background of the study, objectives, statement of the problem, significance and limitation of the study. Chapter two deals with review of related literature on crop production and characteristics of crops in Ethiopia. Chapter three discusses the data and methodology of the study such as sources of data and variables to be included in the study. Methods of data analysis are also described in this chapter. Chapter four presents statistical data analysis and discussion. Finally, summary of the main findings and conclusions of the study are dealt in chapter five.

# CHAPTER TWO

## 2. Literature Review

The concern of agricultural development is a recent area of emphasis in development economics. The literature on agricultural development show that agriculture was not considered as an important contributor to economic growth before the 1950's. But this view has been changing from time to time. The literature puts the changing view of agriculture in economic development in historical perspective (Eicher & Staaz, 1990). Eicher & Staaz show that the development economy of the 1950's used to equate development with the structural transformation of the economy, that is, with the decline in agriculture's relative share in the national product and the labour force. Eicher and Staaz aim to facilitate the transformation by discovering ways to transfer resources (mainly labour) from traditional agriculture to industry. This was because industry was considered as an engine of growth and agriculture was often treated as a "black box from which people and food to feed them could be released". According to Lewis (2004) labour can be transformed from subsistence sector (agriculture) to the industrial sector without any loss of productivity because the marginal productivity of labour in agriculture is zero and the industrial sector can absorb all surplus labour from agriculture. Lewis' assumption was based on infinity elastic supply of labour. Therefore, the hope was that by doing so economic growth could be promoted.

By and large, agriculture in Ethiopia is subsistence. This is particularly true about the major food crops grown in the country. The major food crops are produced in almost all regions of the country in spite of the variation in volume of production across the regions. The variation may be attributed to the extent of area devoted to each crop type, weather change and a shift in preference for the crops grown (*CSA, 2008*).

It is estimated that 16.5 million hectares of land is under cultivation and grains are the most important field crop, occupying 86 percent of area planted and being the chief element in the diet of most Ethiopians. The principal grain crops are teff, wheat and barley, which are primarily cool-weather crops; and corn, sorghum, and millet which are warm weather grain crops. Teff is the most preferred crop grown in the cooler highlands, while sorghum is the principal lowland crop because it thrives well in semi-arid environments due to its hardy and drought resistant properties (*Reynolds, 2002).*

Cereals are the major food crops both in terms of the area they are planted on and the volume of production obtained. They are produced in larger volumes compared with other crops because they are the principal staple crops. Cereals are grown in all the regions with varying quantity. Out of the total grain crop area, 79.69% (8.7million hectares) was under cereals. Teff, maize, wheat and sorghum took up 23.42% (about 2.6 million hectares), 16.12% (about 1.8 million hectares), 13.01% (1.4 million hectares) and 14.01% (1.5 million hectares) of the grain crop area, respectively. Cereals contributed 85.11% (about 137.1 million quintals) of the grain production. Maize, wheat, Teff and sorghum made up 23.24% ( 37.5 million quintals), 14.36% ( 23.1 million quintals), 18.57% ( 29.9 million quintals) and 16.52% ( 26.6 million quintals) of the grain production in the same order (*CSA, 2008*).

*Reynolds (2002)* assessed cereal producing zones based on weather conditions favorable for each cereal type. Teff, wheat, and barley are cool weather crops grown predominantly in the Ethiopian highlands at optimum altitude ranges of 1800 to 2200 meters. Teff occupies the largest area (1.4 million hectares) and has the largest total cereal production. Teff, indigenous to Ethiopia, forms the staple diet of many Ethiopians and it furnishes the flour to make *injera*, an unleavened bread that is consumed in the highlands and in urban centers throughout the country. Teff is, however, a very delicate and fragile crop that requires a lot of work and care, and it has one of the lowest yields of the cereal crops. Barley is another major subsistence crop grown mostly between 2,000 and 3,500 meters, and also used in the production of *tella*, a locally produced beer.

Common warm weather cereal crops in Ethiopia are corn, sorghum, and millet, which are cultivated mostly at lower altitudes along the country's western, southwestern, and eastern peripheries. These three grains are the staple foods for a large part of the population and are major items in the diet for pastoralists. Sorghum and millet are drought resistant and grow well at low elevations where rainfall is less reliable. Sorghum is particularly important in northern Ethiopia, including in the highland areas of western Tigray. Corn is grown chiefly between elevations of 1500 and 2200 meters and requires large amounts of rainfall to ensure good harvests. Corn is particularly important in southwest Ethiopia, with the Oromiya Region producing the largest amount of corn.

Croppensted and Mulat (1997) conducted an empirical study on cereals production and technical efficiency of private farmers in Ethiopia. The study applies a panel model of mixed fixed-random regression approach to analyse data for cereal growing small-scale farmers in efficiency. The results show that land size is a major constraint and only small changes in cultivated area and land quality yield relatively high increments to output. Larger farms are relatively less productive, everything else being equal. Human capital in the form of literacy and experience are found to affect productivity positively. With regard to the inputs the findings show high degrees of input-specific technical inefficiency, especially so for labour and fertilizer. The age structure of the household, environmental factors and education are found to be weakly correlated with efficiency.

Alemu (2005) also conducted a study to asses factors with potential effects on variability in cereal production. The study applies descriptive and variance decomposition procedures. The result of the study indicated that, in Ethiopia, average cereal production between the period 1990 and 2000 did not change significantly compared to the period between 1974 and 1990. Between 1990 and 2000, cereal production was characterized by significant instability. It was found that production instability was caused more by increased yield instability than instability in an area. Yield instability could be the result of changes in technology, changes in policy and changes in weather conditions. The study concludes that instability regarding yield was predominantly the result of weather variability. This is because, in Ethiopia, rainfall fluctuation from the long-term average is becoming more common, the use of high-yield inputs is limited to a small number of farmers, production is at subsistence level, and farmers' responsiveness to policy changes is constrained by infrastructural, institutional and the existing land policy.

Kibara et al, (2008) conducted a study about trends in Kenyan smallholder agricultural productivity and factors with potential impact on productivity change of maize, tea, coffee, sugarcane, cabbages and Irish potatoes between the years 1997-2007. The study applied descriptive analysis to show trends in partial productivity measures such as crop output per unit of land and labor and fixed effects based on Cobb-Douglas production function to identify the major determinants of crop productivity growth on small holder farms after controlling for other factors, and to examine the significance of the various productivity determinants over eight different agro-regional zones. The results show that there was a consistent growth in maize production due to key factors that have contribution to productivity growth which include increased percentage of small holder households using fertilizers, more complete adoption of high-yielding seed varieties and an increased density of fertilizer retail outlets leading to a decline in the distances to sellers of agricultural inputs. In addition tea productivity has grown slightly, driven by increased fertilizer use and cabbage and Irish potato productivity fluctuated over the panel period.

Based on Alemu's findings, several factors could cause instabilities in cereal production in Ethiopia, which include fluctuations in areas sown, fluctuations in yield, and fluctuations in weather conditions and changes in pricing and marketing policies.

Change in areas sown: In Ethiopia, changes in areas sown constitute the major sources of production increase. This can be attributed to the domination of small-scale farmers, characterized by low input and low output rain-fed mixed farming with traditional technologies. Small-scale farmers constitute over 95% of total area sown and over 90% of agricultural output (MEDaC, 1999). In Ethiopia, expansion in areas sown as a potential source of production increase in recent years is being challenged by a decrease in holding sizes, which are presently estimated to be less than one hectare. Decrease in holding size is considered to be a direct consequence of the existing land policy, which disallows transfer of land and declares land the property of the state. Until 1989, new claims for land were entertained through land redistribution schemes. However, in 1989, land redistribution was officially banned because of decline of holding size. Consequently farmers who wish to plough, but who have no land, rely on family plots or enter into various forms of sharecropping arrangements. Cultivated land as a percentage of agricultural area has increased. It has risen from 21% between 1961 and 1975, to 32% between 1992 and 2000. The increase in the area cultivated is made possible by the expansion of cultivation to areas that were previously designated as permanent pasture or forests, or land previously categorized as unsuitable for farming. Forests covered 40% of the land area at the turn of the last century, but less than 4% today (Alemu, 2005).

Weather variability: As stated in the earlier sections, expansion of area sown, fluctuations in fertilizer availability and policy related factors have the potential to cause variations in cereal production. Given that only 25% and 2% of Ethiopian farmers utilize fertilizer and improved seeds respectively (MEDaC, 1999), that cereal production is predominantly rain-fed, and that fluctuations in rainfall from the long-term average are increasing (Webb et al, 1992), modern inputs are unlikely to dominate the effect of weather on cereal production. According to Jaeger (1991), "cereals are the most susceptible crop to moisture stress, and for most countries, variations in average yields of cereals result primarily from variations in weather."

A number of studies attribute the continued dependency of cereal production on weather change to inappropriate economic policies. According to these studies, events such as droughts do not happen suddenly. Drought results from an accumulation of a host of economic problems, which, over time, erode the capacity of farmers to cope (Webb *et al.*, 1992; Pickett, 1991). An attempt is made to prove this premise by comparing coefficient of variation of cereal production (CV) for the three economic systems that the country has experienced since the early 1960s. Results show that CVs were the highest between the period 1990 and 2000 when the country had non-conducive agricultural policies. Between 1975 and 1990 alone, the period when socialism was the economic system of the country, fluctuations in cereal production, solely attributable to weather variability, occurred in nine out of a total of 17 years *(Alemu et al, 2004).*

Education of the holders: The use of modern inputs requires awareness of their existence and knowledge of their advantages in order to compare the outputs under each technological input. The accuracy or quality of farmers' knowledge depends on their source of information and on their level of education. Education may have both cognitive and non cognitive effects upon labour productivity. Cognitive outputs of schooling include the transmission of specific information as well as information on general skills and proficiencies. Education also produces non-cognitive changes in attitudes, belief and habits, and may lead to greater willingness to accept risk, adopt innovations, save for investment and generally to embrace productive practices (Appleton, 1996).

Education also increases prior access to external sources of information and enhances the ability to acquire information through experiences with new technology. Wozniack (1987) proposes that education enhances the ability to adjust to technological changes. Education may indirectly increase output through interaction with other institutional variables. Education enables farmers to obtain wage employment by providing skills. This generates cash to finance agricultural investment. The empirical analysis made by Weir (1999) showed substantial internal (private) benefits of schooling for farmer's productivity in terms of efficiency gains in the case of Ethiopia.

When a new technology has uncertain returns and fixed costs of adoption, some individuals will make reallocation decisions faster than others. Differences on how quickly producers adopt to a changing set of production possibilities can be explained by differences in human capital and differences in knowledge of the new technology. Education helps to reduce subjective uncertainty and unnecessary anxiety as well as fatalistic acceptance of the status quo and thereby enhances the probability of adoption of new technology or practices by an individual (Netsanet, 1998). The dissemination of recommended agricultural technological innovations to farmers is chiefly carried out by extension agents. The Ministry of Agriculture and Rural Development has extension service department at all levels (National, Regional, Zonal, and woreda level) whose responsibility is to train and advice farmers on the use of improved technologies and better farming practices.

Fertilizers: The use of modern agricultural inputs such as chemical fertilizer, improved seed varieties, insecticides, and the corresponding cultural practices was not known in Ethiopia until the mid 1960's. Fertilizer refers to anything that is added to the soil and intended to increase the amount of plant nutrients available for crop growth. There are two types of fertilizers (natural and chemical). The natural fertilizer consists of the farm yard manure and wood ashes while the chemical type consists of DAP (Di- Ammonium phosphate) and UREA (Ammonium Nitrate). For farming system to remain productive and to be sustainable in the long term, it is necessary to replenish the reserves of nutrients, which are removed from the soil.

Regardless of the yield increasing effects of these modern agricultural inputs', the use of them have been minimal. Taking fertilizer as the major agricultural input, the per hectare consumption of arable land shows that Ethiopia uses the lowest amount in the world. Ethiopia, the third most populous country in Africa, is faced with dual problem of high population growth and declining food production (Yao, 1996). The perception of agricultural inputs supply company (AISCO) researchers and extensionists for the very low growth in agricultural productivity is that the effective demand for fertilizer in Ethiopia is not being met. This may be one reason why farmers are applying fertilizer at only half the recommended levels for cereals. While total fertilizer use has generally increased over the last 10 years, the rate of chemical fertilizer used in Ethiopia is less than 100 kg / hac. (*FAO, 2008*).

Many factors affect the use of fertilizers in the peasant agricultural sector in Ethiopia. Some of the factors are related to the general conditions of agriculture such as security, lack of roads, uncertain weather condition and the limited development of improved seed varieties which greatly augment the benefit of fertilizer use and still primitive farming techniques practiced by Ethiopian small holder farmers. (*FAO, 2008*).

Improved Seeds: By this we mean crop variety, which gives a significantly higher yield, and better quality compared to locally produced variety of seeds. The utilization by the peasant sector of improved seed varieties which are complimentary to fertilizer use has not been satisfactory. The Ministry of Agriculture and Rural Development in its annual report of 1998/99 has indicated that improved seed is widely regarded as a "leading input in agricultural development".

The national demand for seeds has been estimated at four million quintals. The Ethiopian seed Enterprise (ESE) is incapable to meet a higher demand. At first glance, it appears that the seed company is doing a poor job of transmitting technology to farmers if it is only distributing half of the available varieties and meeting 5% of seed demand. Second, the farmers are not familiar with them because subsistence farmers fear the risk of using new inputs such as new seed varieties and hence refuse to accept until they are well convinced about the results. The extension and research system so far has not been able to convince the farmers about the benefits of improved seed varieties.

Irrigation Practice: Lack of natural rainfall has been a major bottleneck to improving agricultural output in much of the world. Most countries rely heavily on rain fed agriculture in the process of agricultural production. This heavy reliance on unreliable natural rainfall has come to expose a number of developing countries to drought. For this reason many of them have diverted to adopt irrigation farming for the simple reason that it ensures food harvest.

Irrigation development at present has become one of the most important components for agricultural production in the world. Basically the objectives of irrigation development are based up on the benefits from irrigation. At present the benefits of irrigation are getting recognitions in many places. Specially getting higher yields from existing farms, bringing new lands under irrigation, increased crop intensity and in general achieving higher yields with a relatively greater certainty are some of the unique benefits of irrigation. All these benefits are reflected in the economy as a whole in the form of increased production and a decrease in prices of food items to meet the growing needs of the population. This also helps to save certain amount of foreign exchange that would otherwise be used for agricultural imports and to widen the scope for increased agricultural exports (FAO, 1998).

## CHAPTER THREE

# 3. MATERIALS AND METHODOLOGY

### 3.1 Source of Data

The data used in this study are secondary data obtained from annual post harvest agricultural sample survey conducted from 1995 to 2008 by Central Statistical Agency (CSA), and the National Meteorological Agency (NMA). The agricultural sample survey (AgSS) is the annual survey designed to provide quantitative information on estimates of the total crop area, volume of crop production and farm management practices which includes volume and type of inputs used which are essential for planning, policy making and monitoring & evaluation of food security in Ethiopia. The data were collected from private peasant holdings in rural sedentary areas of the country by interviewing the agricultural holders and physically measuring their fields. The sampling technique implemented for AgSS is a stratified two stage cluster sampling. The country was divided into different enumeration areas (EAs) and EAs were taken to be primary sampling unit (PSU). Enumeration areas from each stratum were selected systematically using probability proportional to size (PPS). Size being number of agricultural households. The agricultural households were taken to be secondary sampling units (SSUs) and were selected from the total agricultural households in each EAs using systematic sampling method.

### 3.2 Variables included in the study

Regions whose cereals production contributed more than five percent to total output were selected for the study. Data for production (in quintal), total area covered (in hectare), area under irrigation (in hectare), area under pesticide application (in hectare), number of literate holders, fertilizer used (in quintal) and improved seed use (in quintal) obtained from time series data on various statistical bulletins of CSA, and quarterly average rainfall (in millimeter) compiled from the National Meteorological Agency (NMA) over the years 1995-2008 in different regions are taken as study variables. Production (in quintal) is a response variable while the others are explanatory variables. Each variable was represented as follows.

$Y_{it}$ = Estimate of crop production (cereal) of $i^{th}$ region at time t.

$X^{(1)}{}_{it}$ = Estimate of total area covered by cereal in $i^{th}$ region at time t.

$X^{(2)}{}_{it}$ = Estimate of fertilizer used for cereals by $i^{th}$ region at time t.

$X^{(3)}{}_{it}$ = Estimate of improved seed used for cereals by $i^{th}$ region at time t.

$X^{(4)}{}_{it}$ = Estimate of area under pesticide of cereal in $i^{th}$ region at time t.

$X^{(5)}{}_{it}$ = Estimate of area under irrigation of cereal in $i^{th}$ region at time t.

$X^{(6)}{}_{it}$ = Number of literate holders of $i^{th}$ region at time t.

$X^{(7)}{}_{it}$ = First quarter average rainfall of $i^{th}$ region at time t.

$X^{(8)}{}_{it}$ = Second quarter average rainfall of $i^{th}$ region at time t.

$X^{(9)}{}_{it}$ = Third quarter average rainfall of $i^{th}$ region at time t.

$X^{(10)}{}_{it}$ = Fourth quarter average rainfall of $i^{th}$ region at time t.

## 3.3 Methodology

### 3.3.1 Introduction

Panel data analysis is used in this study. Panel data analysis is a method of studying a particular subject within multiple sites, periodically observed over a defined time frame. With repeated observations of enough cross-sections, panel analysis permits the researcher to study the dynamics of change with short time series. The combination of time series with cross-sections can enhance the quality and quantity of data in ways that would be impossible using only one of these two dimensions (Gujarati, 2004). Panel analysis can provide a rich and powerful study of a set of cross-section (Country, State Company, person, etc), if one is willing to consider both the space and time dimension of the data. Hsiao (2003) and Klevmarken (1989) list some of the benefits from using panel data. These include the following.

(1) Since panel data relate to individuals, firms, states, countries, etc., over time, there is bound to be heterogeneity in these units. The techniques of panel data estimation can take such heterogeneity explicitly into account by allowing for individual-specific effects. We use the term *individual* in a generic sense to include microunits such as individuals, firms, states, and countries.

(2) By combining time series of cross-section observations, panel data give "more informative data, more variability, less collinearity among variables, more degrees of freedom and more efficiency."

(3) Panel data are enable to study the *dynamics of adjustment*.

Cross-sectional distributions that look relatively stable hide a multitude of changes. Spells of unemployment, job turnover, residential and income mobility are better studied with panels. Panel data are also well suited to study the duration of economic states like unemployment and poverty, and if these panels are long enough, they can shed light on the speed of adjustments to economic policy changes. For example, in measuring unemployment, cross-sectional data can estimate what proportion of the population is unemployed at a point in time. Repeated cross-sections can show how this proportion changes over time. Only panel allow to estimate what proportion of those who are unemployed in one period can remain unemployed in another period. Important policy questions like determining whether families' experiences of poverty, unemployment and welfare dependence are transitory or chronic necessitate the use of panels. Deaton (1995) argues that, unlike cross-sections, panel surveys yield data on *changes* for individuals or households. It allows us to observe *how* the individual living standards change during the development process. It enables us to determine *who* is benefiting from development. It also allows us to observe whether poverty and deprivation are transitory or long-lived, the income-dynamics question. Panels are also necessary for the estimation of inter-temporal relations, lifecycle and intergenerational models. In fact, panels can relate the individual's experiences and behavior at one point in time to other experiences and behavior at another point in time. For example, in evaluating training programs, a group of participants and non-participants are observed before and after the implementation of the training program.

(4) Panel data can better detect and measure effects that simply cannot be observed in pure cross-section or pure time series data. For example, the effects of minimum wage laws on employment and earnings can be better studied if we include successive waves of minimum wage increases in the federal and/or state minimum wages.

(5) Panel data enables us to study more complicated behavioral models. For example, phenomena such as economies of scale and technological change can be better handled by panel data than by pure cross-section or pure time series data.

(6) By making data available for several thousand units, panel data can minimize the bias that might result if we aggregate individuals or firms into broad aggregates.

In short, panel data can enrich empirical analysis in ways that may not be possible if we use only cross-section or time series data. This is not to suggest that there are no problems with panel data modeling.

### Data Arrangment

A panel data set contains a series of observations per each of N entities (e.g., firms, states, regions). Each entity (or subject) includes T observations (1 through t time period). Thus, the total number of observations is NT.

### 3.3.2 Model specification

#### 1. One-way Error Component Regression Model

A panel data regression differs from a regular time-series or cross-section regression in that it has a double subscript on its variables, that is,

$$y_{it} = \alpha + X'_{it}\beta + v_{it} \qquad i = 1, 2, \ldots N \quad t = 1, 2, \ldots T \tag{3.1}$$

with $i$ denoting regions in the country (cross-section dimension) whereas t denoting time (time dimension). $\alpha$ is a scalar, $\beta$ *is* $k \times 1$ vector and $X_{it}$ is the $i^{th}$ observation at time $t$ on k explanatory variables. Most of the panel data applications utilize a one-way error component model for the disturbances, with $v_{it} = \tau_i + u_{it}$ where $\tau_i$ denotes the unobservable individual specific effect and $u_{it}$ denotes the remainder disturbance.

#### 1.1 Fixed Effects Regression Model

Fixed effects regression is the model to use when one wants to control for omitted variables that differ between cases but are constant over time. It lets you use the changes in the variables over time to estimate the effects of the independent variables on your dependent variable, and is the main technique used for analysis of panel data. One of the most important potential sources of bias in cross-sectional econometrics is the so-called heterogeneity bias arising from unobserved heterogeneity to both the dependent and independents variables. An example is the estimation of the effect of fertilizer on farm production in the presence of unobserved land quality (Baltagi, 2005). Consider fixed effects regression model where the slope coefficients are constant, but the intercept varies over Region.

$$y_{it} = \alpha_i + X'_{it}\beta + u_{it} \tag{3.2}$$

where $X'_{it} = \left[X^{(1)}{}_{it}, X^{(2)}{}_{it}, X^{(3)}{}_{it}, \ldots\ldots, X^{(k)}{}_{it}\right]$, $i = 1, 2, \ldots N,\ t = 1, 2, \ldots T$

$$\beta' = (\beta_1, \beta_2, \beta_3, \ldots\ldots, \beta_k)$$

Assumptions. $\operatorname{cov}(X_{it}, \alpha_i) \neq 0$ , $u_{it} \sim N(0, \delta^2{}_u)$

In this study regions (Tigray, Amahara, Oromiya, SNNPR) are taken as individuals and the time period is from 1995 to 2008. $\alpha_i$ denotes regional effect which is taken to be constant over time and specific to region *i*.

The marginal effects β of the set of k-vector of time-varying characteristics $X_{it}$ are taken to be common across *i* and *t* although this assumption can itself be tested.

### 1.1.1 Pooled ordinary Least Squares (OLS) Estimator

The simplest approach to the estimation of equation (3.2) assumes that the individual (regional) effects $\alpha_i$ are fixed and common across regions, *i.e* $\alpha_i = \alpha$ for all $i = 1, 2, 3, \ldots. k$. This leads straight forwardly to a classical linear regression formulation for which Ordinary least Squares will produce consistent and efficient estimates of α and β. Therefore equation (3.2) becomes

$$y_{it} = \alpha + X'_{it}\,\beta + u_{it}.$$

Stacking the individuals gives

$$y = X\beta * + u \tag{3.3}$$

where $X' = (l, X^{(1)}, X^{(2)}, \ldots\ldots X^{(k)})$, $\beta* = (\alpha, \beta_1, \beta_2, \ldots\ldots \beta_k)$ and $l$ is $NT \times 1$ column vector of ones.

This model (3.3) is known as the pooled least squares model, for which estimators of $\hat{\beta}*$ is formulated in the normal way, ie,

$$\underline{\hat{\beta}} * = (X'X)^{-1} X'Y \tag{3.4}$$

$$\operatorname{var}(\hat{\beta}*) = (X'X)^{-1}\sigma^2 u \tag{3.5}$$

Notice that this formulation does not distinguish in any way between two different individuals and the same individual at two points in time, a feature which undermines the accuracy of the approach when differences do exist between cross-sectional individuals. Nevertheless, the increase in sample size by pooling data across time generates an improvement in efficiency relative to a single cross-section.

### 1.1.2 The Least Square Dummy Variable (LSDV) estimator

This formulation of the model assumes that any differences across regions can be captured by shifts in the intercept term of a standard OLS regression. The $\alpha_i$ of model (3.2) is treated as an unknown parameter to be estimated. This leads to the least square dummy variable (LSDV) estimator of a fixed effects regression model.

$$y_{it} = \alpha_i + X'_{it}\beta + u_{it} \tag{3.6}$$

Let $y_i$ and $X_i$ be row vector of $T$ observations for $i^{th}$ unit, $l$ be a $T \times 1$ column vector of ones, and let $u_i$ be the associated $T \times 1$ vector of disturbances, that is,

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ . \\ . \\ y_{iT} \end{bmatrix}, \qquad X_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ . \\ . \\ X_{iT} \end{bmatrix}, \qquad u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ . \\ . \\ u_{iT} \end{bmatrix}$$

using these, equation (3.6) can be written as:

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_N \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ \beta_k \end{bmatrix} + \begin{bmatrix} l & 0 & \ldots & 0 \\ 0 & l & \ldots & 0 \\ . & . & \ldots & 0 \\ . & . & \ldots & 0 \\ 0 & 0 & 0 & l \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ . \\ . \\ \alpha_N \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_N \end{bmatrix}$$

$$y = X\beta + D\alpha + u \tag{3.7}$$

where $D = I_N \otimes l$, $I_N$ is the identity matrix of dimension $N$ and $\otimes$ represents the Kroneker product.

This model is usually referred to us the least squares dummy variable (LSDV) model.

The least squares estimator of $\gamma = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$ is given as:

$$\hat{\gamma}_{LSDV} = \left[(X \mid D)'(X \mid D)\right]^{-1}\left[X \mid D\right]' y$$

$$= (X' M_D X)^{-1} X' M_D y \tag{3.8}$$

where $M_D = I - D(D'D)^{-1}D'$.

### 1.1.3 The Within-Groups (WG) estimator

When the regional effects $\alpha_i$ are fixed but not common across $i = 1, 2, ......, N$, the pooled estimator is evidently biased through the misspecification of the mean equation $E(y_{it} / X_{it}) = \alpha_i + X'_{it}\,\beta$.

Due to this, another formulation of pooled regression is in terms of deviations from the group means which eliminates the group effect $\alpha_i$.

$$y_{it} - \bar{y}_{i.} = (X_{it} - \bar{X}_{i.})\beta + (u_{it} - \bar{u}_{i.}) \tag{3.9}$$

$$y^*_{it} = X^*_{it}\,\beta + u^*_{it}$$

This is called the within individual effect model. The least square estimator of $\beta$ is:

$$\hat{\beta}^{WG} = (X^{*\prime} X^*)^{-1} X^{*\prime} y^* \tag{3.10}$$

where $X^*_{it} = (X_{it} - \bar{X}_i.)$, $y^*_{it} = (y_{it} - \bar{y}_i.)$, $\bar{X}_i. = T^{-1}\sum_{t=1}^{T} X_{it}$, $\bar{y}_i. = T^{-1}\sum_{t=1}^{T} y_{it}$.

On the basis that $Var(u^*_{it}) = (\frac{T-1}{T})Var(u_{it})$, the covariance of $\hat{\beta}^{WG}$ is :

$$Var(\hat{\beta}^{WG}) = (X^{*\prime} X^*)^{-1}\sigma^2_{u^*} = \frac{(T-1)}{T}(X^{*\prime} X^*)^{-1}\sigma^2_{u} \tag{3.11}$$

Therefore the diagonal element of $Var(\hat{\beta}^{WG})$ is larger than the diagonal element of $Var(\hat{\beta}^*)$ given by equation (3.5) for large $T$ since the denominator of (3.11) which is $(X^{*\prime} X^*)$ is smaller than the denominator of equation 3.5 which is $(X'X)$. This demonstrates that the within groups estimator produces consistent estimates of β in the presence of regional fixed effects, but is inefficient relative to pooled OLS .

## 1.2 The Random Effects Model

The fixed effects model allows the unobserved individual effects to be correlated with the included variables. It is appropriate when differences between individual agents may reasonably be viewed simply as parametric shifts in the regression function itself. This model might be viewed as applying only to the cross-sectional units in the study, not to additional ones outside the sample. If the individual effects are strictly uncorrelated with the regressors, then it might be appropriate to model the individual specific constant terms as randomly distributed across cross-sectional units. This view would be appropriate if we believed that sampled cross-sectional unites were drawn from a large population (Green, 2000 and Maddala, 1992). Consider one way fixed effects model. Defining $\alpha_i = \alpha + \tau_i$. where $\tau_i$ represents an individual disturbance which is fixed over time and has a zero mean, the model becomes:

$$y_{it} = \alpha + X'_{it}\beta + \tau_i + u_{it} \qquad (3.12)$$

### 1.2.1 The Generalized Least Squares (GLS) estimator

To estimate the linear random effects model (sometimes called the variance components or random components model) requires a Generalized Least squares approach to deal with the more complex error structure inherent in (3.12) compared with the fixed effects model.

Assumptions:

$$
\begin{aligned}
&E(u_{it}/X) = E(\tau_i/X) = 0 \\
&E(u^2{}_{it}/X) = \sigma^2{}_u \\
&E(\tau^2{}_i/X) = \sigma^2{}_\tau \\
&E(u_{it}\tau_j) = 0 \ \ for\ all\ i,t,j \\
&E(u_{it}u_{js}/X) = 0 if\ t \neq s, i \neq j \\
&E(\tau_i\tau_j/X) = 0 \quad if\ i \neq j
\end{aligned}
$$

Let $\omega_{it} = u_{it} + \tau_i$ Then

$E(\omega_{it}) = 0$ for all $i,t$

$E(\omega^2{}_{it}) = \delta^2{}_u + \delta^2{}_\tau$ for all $i$, $t$

$E(\omega_{it}\omega_{is}) = \delta^2{}_\tau$ for all $t \neq s$

$E(\omega_{it}\omega_{js}) = 0$ for $i \neq j$or $t \neq s$

So, if we collect the $T$ disturbances for individual $i$ in a vector of the form

$\omega_i = (\omega_{i1}, \omega_{i2}, \ldots\ldots\ldots \omega_{iT})'$, we have that $E(\omega_i\omega'_i) = \Omega$

where
$$
\Omega = \begin{bmatrix}
\delta^2{}_u + \delta^2{}_\tau & \delta^2{}_\tau & \delta^2{}_\tau & \ldots & \delta^2{}_\tau \\
\delta^2{}_\tau & \delta^2{}_u + \delta^2{}_\tau & \delta^2{}_\tau & \ldots & \delta^2{}_\tau \\
\delta^2{}_\tau & \delta^2{}_\tau & \delta^2{}_u + \delta^2{}_\tau & \ldots & \delta^2{}_\tau \\
\ldots & \ldots & \ldots & \ldots & \delta^2{}_\tau \\
\delta^2{}_\tau & \delta^2{}_\tau & \delta^2{}_\tau & \ldots & \delta^2{}_u + \delta^2{}_\tau
\end{bmatrix}
$$

$$
= \delta^2{}_u I_T + \delta^2{}_\tau l_T l'_T \qquad (3.13)
$$

were $l_T$ is a $Tx1$ column vector of ones.

For the full panel of observations, the covariance matrix of $NT$ vector of disturbances $\omega = (\omega_1, \omega_2, \omega_3, \ldots\ldots \omega_N)'$ is given by:

$$V_{NT\times NT} = \begin{bmatrix} \Omega & 0 & 0 & \cdots & 0 \\ 0 & \Omega & 0 & \cdots & 0 \\ 0 & 0 & \Omega & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \Omega \end{bmatrix} = I_N \otimes \Omega \qquad (3.14)$$

Let y represent a stacked vector of $y_{it}$ formed in a similar fashion to ω (with a similar structure for X). The full system

$$y = X\beta + \omega \qquad (3.15)$$

may therefore be estimated by Generalized Least Squares, given the structure of the covariance matrix V. Generally, GLS estimation of a regression of the form (3.15) requires a transformation to remove the non-standard structure of the covariance matrix $E(\omega\omega') = V$. We define the weight matrix $P = V^{-\frac{1}{2}}$, and transform (3.15) by pre-multiplication, to give:

$$\begin{aligned} PY &= PX\beta + P\omega \\ Y^* &= X^*\beta + \omega^* \end{aligned}$$

Now we can get,

$$\begin{aligned} E(\omega^* \omega^{*\prime}) &= E(P\omega\omega' P) \\ &= PE(\omega\omega')P \\ &= P.V.P \\ &= I_{NT} \end{aligned}$$

which has common variances across *i* and *T*. So with knowledge of *P*, the GLS estimators of the regression function (3.15) may be derived as:

$$\hat{\beta}_{GLS} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}Y\,. \qquad (3.16)$$

Note that that $\hat{\beta}_{GLS}$ includes $\hat{\alpha}$ .

For the random effects model, one can generate a specific form for the weight matrix $P = V^{-\frac{1}{2}}$. From (3.13) and (3.14) we have:

$$V^{-\frac{1}{2}} = I_N \otimes \Omega^{-\frac{1}{2}}$$

$$= I_N \otimes \frac{1}{\sigma_u}\left[ I_T - \frac{\theta}{T} l_T l'_T \right] \qquad (3.17)$$

$$where \ \theta = 1 - \frac{\sigma_u}{\left(T\sigma^2_\tau + \sigma^2_u\right)^{1/2}}$$

So, the appropriate transformation for the random effects model is to pre-multiply each $y_i = (y_{i1}, y_{i2}, \ldots\ldots, y_{iT})'$ $by\ \Omega^{-\frac{1}{2}}$ to get:

$$y_i{}^* = \Omega^{-\frac{1}{2}}.y_i = \frac{1}{\sigma_u}\begin{bmatrix} y_{i1} - \theta\bar{y}_i \\ y_{i2} - \theta\bar{y}_i \\ . \\ . \\ y_{iN} - \theta\bar{y}_i \end{bmatrix} \qquad (3.18)$$

with similar transformations to generate each $X_i{}^* = \Omega^{-\frac{1}{2}} X_i$ .

### 1.3 Testing the Significance of the Regional Effects.

The *F-test* can be used for a test of the hypothesis that $\alpha_i$ are all equals.

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = ..... = \alpha_N$$
$$H_A : H_0 \ \ is \ not \ true \ \ for \ at \ least \ two \ regions$$
$$at \ \alpha - level \ of \ significance.$$

To test this hypothesis, one can either estimate a pooled OLS on an intercept and $K$ regressors $X_{it}$ (giving a coefficient of determination $R^2_p$) or LSDV model (with coefficient of determination $R^2_{DV}$). The F-ratio used for this test is

$$F = \frac{\dfrac{(R^2{}_{DV} - R^2{}_P)}{(N-1)}}{\dfrac{(1-R^2{}_{DV})}{(NT-N-K)}} = \frac{(R^2{}_{DV} - R^2{}_p)(NT-N-K)}{(1-R^2{}_{DV})(N-1)} \qquad (3.19)$$

which is distributed as an $F_{N-1, NT-N-T}$ under the null hypothesis.

### 1.4 Testing for Random Effects

Breusch and Pagan (1980) have devised a Lagrange multiplier test for the random effects model based on the OLS residuals. The specific hypothesis under investigation is the following.

$$H_o : \sigma^2{}_\tau = 0$$
$$H_A : \sigma^2{}_\tau > 0 \ at \ \alpha - level \, of \ significance.$$

From inspection of (3.14), one can see that $V = \sigma^2{}_u I_{NT}$ under the null hypothesis, so that the Random Effects Model (REM) reduces to a pooled OLS regression.

The test of this hypothesis, based on OLS residuals $\hat{u}_{it}$ from pooled regression model, requires the LM statistic:

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}(\sum_{t=1}^{T}\hat{u}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^{\,2}} - 1\right]^2 \qquad (3.20)$$

Under the null hypothesis, $LM$ is distributed as chi-squared with one degree of freedom.

## 1.5 The Hausman Specification Test

We may be interested in comparing directly the random effects estimator $\hat{\beta}_{GLS}$ with the fixed effects estimator $\hat{\beta}_{LSDV}$.

$$H_0 : COV[\alpha_i, X_{it}] = 0$$
$$H_A : COV[\alpha_i, X_{it}] \neq 0 \ \ at\ \alpha - level\ of\ significance.$$

Under the null hypothesis, i.e. in the absence of correlation between the regressors $X_{it}$ and regional effects $\alpha_i$, the GLS estimator ($\hat{\beta}_{GLS}$) is consistent and efficient while the LSDV estimator ($\hat{\gamma}_{LSDV}$) is consistent but inefficient. In the presence of correlation between the regressors $X_{it}$ and regional effects $\alpha_i$, the GLS estimator is inconsistent, while the LSDV estimators are consistent.

<u>Test Statistics</u>: Under the null hypothesis, the two estimates should not differ systematically, and a test can be based on the difference between $\hat{\beta}_{GLS}$ and $\hat{\gamma}_{LSDV}$.

Here the variance of the difference $\hat{\beta}_{GLS}$ - $\hat{\gamma}_{LSDV}$ may be derived as:

$$Var(\hat{\beta}_{GLS} - \hat{\gamma}_{LSDV}) = Var(\hat{\beta}_{GLS}) + Var(\hat{\gamma}_{LSDV}) - Cov(\hat{\beta}_{GLS}, \hat{\gamma}_{LSDV}) - Cov(\hat{\beta}_{GLS}, \hat{\gamma}_{LSDV}) \qquad (3.21)$$

Hausman's essential result is that the *covariance of an efficient estimator with its difference from an inefficient estimator is zero.* This implies that

$$Cov(\hat{\beta}_{GLS} - \hat{\gamma}_{LSDV}, \hat{\gamma}_{LSDV}) = Cov(\hat{\beta}_{GLS}, \hat{\gamma}_{LSDV}) - Var(\hat{\gamma}_{LSDV}) = 0$$

Or that $Cov(\hat{\beta}_{GLS}, \hat{\gamma}_{LSDV}) = Var(\hat{\gamma}_{LSDV})$ (3.22)

where $0$ is a $k+1 \times k+1$ matrix of zeros.

Inserting (3.22) in to (3.21) produces the required covariance matrix for the test.

Therefore $Var(\hat{\beta}_{GLS} - \hat{\gamma}_{LSDV}) = Var(\hat{\beta}_{GLS}) - Var(\hat{\gamma}_{LSDV}) = \Sigma$ (3.23)

The Hausman test of the null hypothesis of no correlation can therefore be conducted using the Wald statistic:

$$W = \left(\hat{\beta}_{GLS} - \hat{\gamma}_{LSDV}\right)\hat{\Sigma}^{-1}\left(\hat{\beta}_{GLS} - \hat{\gamma}_{LSDV}\right) \tag{3.24}$$

For $\Sigma$, we use the estimated covariance matrices of the slope estimator in the LSDV model and the estimated covariance matrix in the random effects model, excluding the constant term. Under the null hypothesis, $W$ has a limiting chi-squared distribution with K-1 degrees of freedom, K being the number of regressors in $X_{it}$.

### 1.6 Tests for poolability of the data.

The question of whether to pool the data or not naturally arises with panel data. This question asks if slopes are the same across groups or over time. The poolability test is undertaken under the assumption $u \sim N(0, \sigma^2 I_{NT})$. The hypothesis to be tested is

$$H_o : \beta_{ik} = \beta_k$$
$$H_A : H_o \textit{ is not true}$$

we remember that slopes remain constant in fixed and random effect models. Only intercepts and error variances matter. In order to conduct this test, we need to run group by group OLS regressions. This test uses the F statistic, which is exactly the chow test presented by chow (1960).

$$F_{obs} = \frac{(e'e - \sum_{i=1}^{N} e'_i e_i)/(N-1)k}{\sum_{i=1}^{N} e'_i e_i / N(T-k)} \qquad (3.25)$$

where $e'e$ is the error sum of square of the pooled OLS and $e'_i e_i$ is the error sum of square of the OLS regression for group i.

Under the null hypothesis, $F_{obs}$ is distributed as an $F_{((N-1)k,\ N(T-k))}$ . Hence the critical region for this test is defined as

$$\{F_{obs} > F((N-1)k, N(T-k)); \alpha_o\}$$

where $\alpha_0$ denotes the level of significance of the test. If the null hypothesis is rejected, the panel data are not poolable. Under this circumstance, we may go for the random coefficient model or hierarchical regression model

### 1.7 Assumptions for this Analysis

Statistical inferences are based only in part up on the observations. An equally important base is formed by prior assumptions about the underlying situation. After choosing the econometric technique for the estimation of the model, the econometrician should state explicitly the assumptions of this technique and examine their implications for the estimates of the parameters. Even in the simplest cases there are implicit or explicit assumptions about randomness and independence about distributional models, perhaps prior distributions for some unknown parameters, and so on. As it was stated in the model specification part, the model applied for this study follows a classical regression.

## Assumptions of classical regression

The linear regression model is based on certain assumptions, some of which refer to the distribution of the random variable ***u***, some to the relationship between ***u*** and the explanatory variables, and finally some refer to the relationship between the explanatory variables themselves.

### 1. No perfect multicollinearity

The term *multicollinearity* is due to Ragnar Frisch. Originally it meant the existence of a “perfect” or exact linear relationship among some or all explanatory variables of a regression model. Therefore, multicollinearity is the presence of linear relationships among some or all explanatory variables of a regression model. When multicollinearity exists between the explanatory variables, we may observe very high standard errors and low values of the t-statistic, unexpected changes in coefficient magnitudes or signs, or non-significant coefficients despite a high R-square.

### 2. Normality

The residuals are normally distributed with mean zero and constant variance, that is,

$u_{it} \sim N(0,\sigma^2{}_u) \ \ for\ all\ i,t$

## 3. Linearity

The model specifies a linear relation between dependent and independent variables, that is, $Y_{it} = X_{it1}\beta_1 + X_{it2}\beta_2 + \ldots + X_{itk}\beta_k + U_{it}$

If the relationship between independent variables (IV) and the dependent variable (DV) is not linear, the results of the regression analysis will *under-estimate* the true relationship. This under-estimation carries two risks: increased chance of a *Type II* error for the independent variable which is not linear, and in the case of multiple regressions, an increased risk of *Type I* errors (over-estimation) for other independent variables that share variance with that non-linear independent variable (Pedhazur, 1997).

# CHAPTER FOUR

# STATISTICAL DATA ANALYSIS

## 4.1. Panel regression models

This section considers the fixed effects model which includes ordinary least squares (OLS), least square dummy variable model (LSDV), random effects model which includes feasible generalized least square (FGLS), tests for hypothesis which includes tests for poolability, tests for regional and time effects and comparisons of each model. The data for these panel analyses is obtained from two organizations namely, Central Statistical Agency (CSA) and the Ethiopian Meteorological Agency (EMA). Four regions whose contribution to total cereal production in Ethiopia is greater than 5 percent were selected for analysis. The data for each region is the annual time series extending from 1995 to 2008 except rainfall. Rainfall is the average quarterly time series data extending from 1995 to 2008. The analysis provides empirical results to various models formulated in chapter three. Combinations of econometric software packages used for empirical analysis of this study are STATA (version 10), SAS (version 8.2) and SPSS (version 17). After confirming the assumptions of each model used for analysis each model was applied. The major factors that may influence production of cereals at the study regions are discussed.

## 4.2 Descriptive Statistics

Cereals are the major food crops both in terms of the area they are planted and volume of production obtained. From Table 4.1, out of the total cereal production within the last 14 years, 52 percent is from Oromiya, 32 percent is from Amhara, 9 percent is from SNNP and 7 percent is from Tigray. Cereal production is increasing from year to year in a similar fashion like all inputs used except for some years.

**Table 4.1 Total cereal production (in 000's qnt)**

| Year | Tigray | Amhara | Oromiya | SNNP | Total | Average |
|---|---|---|---|---|---|---|
| 1995 | 2897 | 19912 | 31512 | 6106 | 60427 | 15107 |
| 1996 | 4927 | 24774 | 42515 | 8460 | 80677 | 20169 |
| 1997 | 5628 | 27266 | 43053 | 8223 | 84169 | 21042 |
| 1998 | 4284 | 18249 | 33637 | 6420 | 62591 | 15648 |
| 1999 | 6082 | 24481 | 37309 | 6515 | 74386 | 18596 |
| 2000 | 5657 | 23286 | 39289 | 6620 | 74852 | 18713 |
| 2001 | 5494 | 27514 | 45785 | 11533 | 90326 | 22582 |
| 2002 | 5816 | 27247 | 44852 | 7366 | 85281 | 21320 |
| 2003 | 5837 | 27369 | 45283 | 7956 | 86444 | 21611 |
| 2004 | 5857 | 27492 | 45714 | 8546 | 87608 | 21902 |
| 2005 | 5349 | 30055 | 53974 | 8775 | 98153 | 24538 |
| 2006 | 7568 | 36499 | 60286 | 8977 | 113330 | 28332 |
| 2007 | 9613 | 40188 | 65351 | 10154 | 125307 | 31327 |
| 2008 | 10026 | 43611 | 68621 | 11172 | 133430 | 33358 |
| Total | 85035 | 397943 | 657180 | 116823 | 1256981 | 314245 |

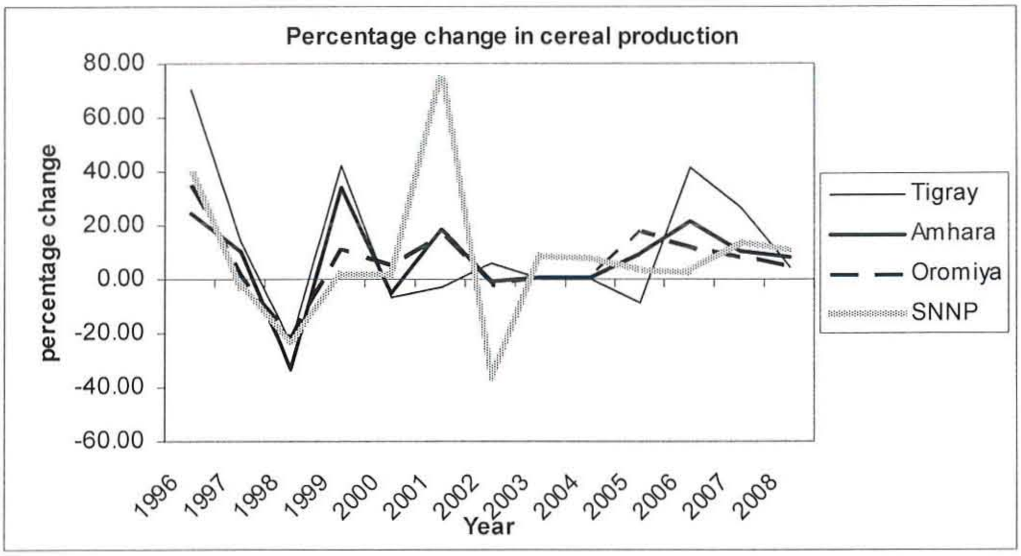


**Fig. 4.1 Percentage change in cereal production from preceding year**

Cereal production decreased in 1998 and 2002 by about 25 and 5 percent, respectively, from preceding years. As the Disaster Prevention and Preparedness Commission (DPPC) and the World Food Program (WFP) reported, this decrease in cereal production is due to the impact of the consecutive occurrence of drought during those years. But there was a continuous increase in cereal production for the remaining years.

The panel models applied for this study were categorized into three groups, namely pooled OLS, fixed effects model and random effects model. After fitting temporary models, the study attempts to test for the best model which fits the data on hand.

## 4.3 Tests for Fixed effects versus Pooled OLS

Pooled ordinary least squares assumes that there is no a regional difference as well as time effect across regions. That is, disregard the space and time dimensions of the pooled data and just estimate the usual OLS regression by stacking the observations for each region one on top of the others.

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$$

$$H_A : H_0 \text{ is not true for at least two regions.}$$

$$at\ \alpha = 0.05 \text{ level of significance.}$$

This hypothesis is tested by the F test, which is based on loss of goodness-of-fit.

$$F = \frac{\dfrac{(R^2{}_{DV} - R^2{}_{P})}{(4-1)}}{\dfrac{(1 - R^2{}_{DV})}{(56-4-9)}} = \frac{(0.9703 - 0.9563)(43)}{(1 - 0.9703)(3)} = 6.756457$$

The tabulated $F_{(3,43)}$ at 0.05 level of significance is **2.83.**

From the above results, we have evidence to reject the null hypothesis. This implies that, the fixed effects model is better than the pooled OLS model.

### 4.4 Tests for Fixed effects versus Random Regional effects

The challenge facing a researcher is which model is better, FEM or REM. The answer to this question hinges around the assumption one makes about the likely correlation between the individuals or cross-section specific error component $\alpha_i$, and the regressors ($X$). If it is assumed that $\alpha_i$ and $X's$ are uncorrelated, REM may be appropriate, otherwise FEM may be appropriate.

$$H_0 : COV[\alpha_i , X_{it}] = 0$$
$$H_A : COV[\alpha_i , X_{it}] \neq 0$$

The Hausman specification test gives the following results:

Regional effects model

| *Wald* | Prob> chi2 |
|---|---|
| 24.21 | 0.0071 |

Time effects model

| *Wald* | Prob> chi2 |
|---|---|
| 2.51 | 0.9907 |

From the Hausman specification test, the estimated test statistic is 24.21 with a P-value of 0.0071. Since the P-value is far smaller at any test of significance level, the hypothesis that regional differences are uncorrelated with the other regressors in the model can be rejected. Therefore, the Hausman test suggests that the regional differences are correlated with the other variables in the model which implies that of the two alternatives we have considered, fixed regional effects model is the better choice.

With a similar procedure for time effects model, since the P-value is far larger at any test of significance level, the hypothesis that time effects are uncorrelated with the other regressors in the model can be accepted. Therefore, the Hausman test suggests that the time effects are uncorrelated with the other variables in the model which implies that of the two alternatives we have considered, the random time effects model is the better choice.

## 4.5 Fixed effects model for regional effects

One way to take into account the "individuality" of each region or each cross-sectional unit is to let the intercept vary for each region but still assume that the slope coefficients are constant. We call it the fixed effects or least-squares dummy variable (LSDV) regression model. The results of the LSDV regression for cereals (using equation 3.8) is given in Table 4.2

Note: $\alpha_1$, $\alpha_2$, $\alpha_3$, and $\alpha_4$ represent regional difference of Tigray, Amhara, Oromiya and SNNP, respectively. Since we have four regions, we have used only three dummies to avoid falling in to the dummy-variable trap (that is, the situation of perfect collinearity). Therefore, we dropped one dummy variable arbitrary.

**Table 4.2 Estimates of the coefficients for the fixed effects model for cereals**

| Variables | Coefficients | Std.error | t-value | P-value |
|---|---|---|---|---|
| Intercept | -2.7945680 | 1.2398800 | -2.25 | 0.030 |
| $\alpha_1$ | -0.0407675 | 0.1080439 | -0.38 | 0.708 |
| $\alpha_2$ | -0.6907747 | 0.1487118 | -4.65 | 0.000 |
| $\alpha_3$ | -0.6100883 | 0.1509919 | -4.04 | 0.000 |
| Area | 1.2691080 | 0.1081349 | 11.74 | 0.000 |
| Fertilizer | 0.3212123 | 0.0434658 | 7.39 | 0.002 |
| Improved seed | 0.0488153 | 0.0184210 | 2.65 | 0.017 |
| Pesticides | 0.0646200 | 0.0182637 | 3.54 | 0.001 |
| Irrigation | -0.0301411 | 0.0302832 | -1.00 | 0.325 |
| Rf1 | 0.0271853 | 0.0163087 | 1.67 | 0.103 |
| Rf2 | 0.0395991 | 0.0302125 | 1.31 | 0.197 |
| Rf3 | 0.2072042 | 0.0681329 | 3.04 | 0.004 |
| Rf4 | -0.0132888 | 0.0179720 | -0.74 | 0.464 |

$R^2 = 0.9703$ F= 536.42 P-value 0.0000 $\sigma^2{}_\alpha = 0.0890503$ $\sigma^2 u = 0.0353605$

Corr.($\alpha_i, X\hat{\beta}$)= 0.7893 F-test for all $\alpha_i$=0 is, F=3.78 p-value=0.0171

In the fixed effects model estimation, the variable "educational status of holders" was discarded since it was found to have a large variance inflation factor (VIF). The large F-statistic (P-value $< 0.0001$) indicates that the model is a good fit.

The null hypothesis of no regional effects ($H_0 : \alpha_i = 0 \ for \ all \ i$) is rejected at the 5% level of significance (P-value=0.0171). This indicates that there is a significant regional variation in cereals production. This may be due to unique features of regions such as soil fertility, temperature, altitude, etc.

The least square dummy variable model assumes that the correlation between regional difference and the explanatory variables is different from zero. The estimated value of correlation coefficient is 0.7893. The quantity of chemical fertilizer, total area covered by cereals, improved seeds, pesticides and the third quarter rainfall have positive and statistically significant impact on cereals production. The regional differences $\alpha_2$ and $\alpha_3$ on cereal production are also statistically significant. However, the regional effect $\alpha_1$ and area under irrigation have no significant impact at the 5% level.

Separate fixed effects regression models are fitted for wheat and maize production. The results are given in Tables 4.3 and 4.4. In both models, the hypothesis of no regional difference is overwhelmely rejected. Unlike the model for aggregate cereals production, area under irrigation has a significant impact on both wheat and maize production. This may be an indication that irrigation practice has primarily focused on these cereals than on others.

**Table 4.3 Estimates of the coefficients for the fixed effects model for wheat**

| Variables | Coefficients | Std.error | t-value | P-value |
|---|---|---|---|---|
| Intercept | -1.21898 | 0.94932 | -1.28 | 0.2060 |
| $\alpha_1$ | -0.22695 | 0.13535 | -1.68 | 0.1009 |
| $\alpha_2$ | -0.67720 | 0.16520 | -4.10 | 0.0002 |
| $\alpha_3$ | -0.51900 | 0.16691 | -3.11 | 0.0033 |
| Area | 1.22200 | 0.08097 | 15.09 | <0.0001 |
| Fertilizer | 0.13896 | 0.04393 | 3.16 | 0.0029 |
| Improved seed | 0.12800 | 0.02994 | 4.27 | 0.0001 |
| Pesticides | 0.07687 | 0.02061 | 3.73 | 0.0013 |
| Irrigation | 0.06307 | 0.01989 | 3.17 | 0.0028 |
| Rf1 | 0.00612 | 0.02426 | 0.25 | 0.8020 |
| Rf2 | 0.09821 | 0.03668 | 2.68 | 0.0110 |
| Rf3 | 0.14407 | 0.08797 | 1.64 | 0.1090 |
| Rf4 | -0.03367 | 0.01991 | -1.69 | 0.0980 |

$R^2 = 0.9892$ F= 328.51 P-value<0.0001 $\sigma^2_\alpha = 0.151104$ $\sigma^2 u$=0.013934

Corr.($\alpha_i, X\hat{\beta}$)= 0.8100 F-test for all $\alpha_i$=0 is, F=9.76 p-value=0.0000

**Table 4.4 Estimates of the coefficients for the fixed effects model for maize**

| Variables | Coefficients | Std.error | t-value | P-value |
|---|---|---|---|---|
| Intercept | 9.03183 | 1.15690 | 7.81 | <0.0001 |
| $\alpha_1$ | -1.04673 | 0.21520 | -4.86 | <0.0001 |
| $\alpha_2$ | -0.19104 | 0.23150 | -0.83 | 0.4138 |
| $\alpha_3$ | 0.67815 | 0.19183 | 3.54 | 0.0010 |
| Area | 0.24803 | 0.07742 | 3.20 | 0.0026 |
| Fertilizer | 0.14762 | 0.03142 | 4.70 | <0.0001 |
| Improved seed | 0.08839 | 0.02623 | 3.37 | 0.0018 |
| Pesticides | 0.08239 | 0.03018 | 2.73 | 0.0219 |
| Irrigation | 0.12859 | 0.03187 | 4.04 | 0.0009 |
| Rf1 | -0.01832 | 0.03966 | -0.46 | 0.6465 |
| Rf2 | 0.04897 | 0.02338 | 2.09 | 0.0420 |
| Rf3 | 0.17972 | 0.08856 | 2.03 | 0.0490 |
| Rf4 | 0.02572 | 0.03095 | 0.83 | 0.4105 |

$R^2 = 0.9773$ F= 154.28 P-value<0.0001 $\sigma^2_\alpha = 0.7493388$ $\sigma^2 u$=0.0311062

Corr.($\alpha_i, X\hat{\beta}$)= 0.5515 F-test for all $\alpha_i$=0 is, F=32.61 p-value=0.0000

## 4.6 Random effects model for cereals

As was confirmed by a test, the time effects are strictly uncorrelated with the regressors. That is. the time specific constant terms are randomly distributed across cross-sectional units. The model is given as

$$y_{it} = \alpha + X'_{it}\beta + \tau_i + u_{it} \qquad i=1, \ldots, 4 \quad t=1, 2, \ldots, 14$$

where $\tau_i$ is the random time effect. The estimates for the parameters can be computed as (equation 3.16):

$$\hat{\beta}_{GLS} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}Y$$

The estimation of the results are given in Table 4.5.

**Table 4.5 Estimates of the coefficients using random time effects for cereals**

| Variables | Coefficients | Std. error | z-value | P-value |
|---|---|---|---|---|
| Constant | 3.952515 | 0.4053411 | 9.75 | 0.000 |
| Area | 0.2826053 | 0.0687213 | 4.11 | 0.000 |
| Fertilizer | 0.4362751 | 0.058746 | 7.43 | 0.000 |
| Improved seed | 0.1012741 | 0.0416766 | 2.43 | 0.002 |
| Pesticides | 0.0646939 | 0.0308313 | 2.10 | 0.036 |
| Irrigation | 0.0378445 | 0.0638441 | 0.59 | 0.553 |
| Rf1 | 0.0234697 | 0.0469202 | 0.50 | 0.617 |
| Rf2 | 0.0869315 | 0.065986 | 1.32 | 0.188 |
| Rf3 | 0.2579592 | 0.0904554 | 2.85 | 0.004 |
| Rf4 | 0.0105918 | 0.0319764 | 0.33 | 0.740 |

$R^2$=0.9532 theta ($\hat{\theta}$)= 0.514419 Wald = 1550.51 P-value =0.0000

From the Wald test it can easily be seen that the model is a good fit. From Table 4.5, the interpretation of the coefficients is similar to the fixed regional effects model. Total area covered by cereals each year, quantity of fertilizer used, improved seed, pesticides and third quarter rainfall (July, August, September) have a positive and significant effect on cereal production. The significance of the random time effect needs another test called Lagrange Multiplier test.

### 4.7 Tests for Random Time Effects

The null hypothesis is that time variance components are zero. Breusch and Pagan (1980) developed the Lagrange multiplier (LM) test statistic given by:

$$LM = \frac{NT}{2(N-1)}\left[\frac{\sum_{t=1}^{T}(\sum_{i=1}^{N}\hat{u}_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^{2}}-1\right]^2 = \frac{56}{2(4-1)}\left[\frac{\sum_{t=1}^{14}(\sum_{i=1}^{4}\hat{u}_{it})^2}{\sum_{i=1}^{4}\sum_{t=1}^{14}\hat{u}_{it}^{2}}-1\right]^2$$

The LM test statistic is distributed as chi-squared with one degree of freedom. The results of this test are displayed in Table 4.6.

**Table 4.6 Breusch and Pagan Lagrangian multiplier test for random time effects**

| variable | Variance | sd = sqrt (Var) |
|---|---|---|
| Production | 0 .7910521 | 0 .889941 |
| Overall error | 0.0269797 | 0.164255 |
| Random time error | 0.0218609 | 0.147854 |

Test: $Var(\tau) = 0$ $LM$ = 6.41 P-value = 0.0114

The small P-value (0.0114) implies that we have evidence to reject the null hypothesis in favor of random time effects. Therefore, random time effects for cereal production are statistically significant.

## 4.8 Examination of the assumptions.

Most statistical tests rely upon certain assumptions about the variables used in the analysis. When these assumptions are not met the results may not be trustworthy, resulting in a considerable Type I or Type II error, or over- or under-estimation of significance or effect size(s). As Pedhazur (1997) notes, "Knowledge and understanding of the situations when violations of assumptions lead to serious biases, and when they are of little consequence, are essential to meaningful data analysis". However, as Osborne et al, (2001) observe, few articles report having tested assumptions of the statistical tests they rely on for drawing their conclusions.

### 1. Tests of Normality

Skewness is a measure of asymmetry, or more precisely, lack of symmetry. The coefficient of skewness for the normal distribution is zero and symmetric data should have a skewness near zero. Kurtosis is a measure for detecting normality. It is a measure of whether the data are peaked or flat relative to a normal distribution. The result of the skewness-kurtosis test for the residual is given in table 4.7.

**Table 4.7 Skewness/Kurtosis tests for Normality**

| Variable | Pr(Skewness) | Pr(Kurtosis) | adj chi2(2) | P-value |
|---|---|---|---|---|
| Residual | 0.005 | 0.079 | 9.27 | 0.0097 |

As we can notice from skewness-kurtosis test, the p-value is very small which implies that we reject the null hypothesis that the distribution is normal. The probability plots (Fig. A1 in the annex) also indicate that the distribution is not normal as the points show a clear departure from the 45 degree line. Therefore, we need to transform our data to achieve normality. After looking different methods of transformation, the log transformation was found to be appropriate to correct the departure from normality.

The skewness-kurtosis test of the residuals after the log transformation of our data is presented in Table 4.8. The results indicate that the null hypothesis of normality is not rejected. Fig. A2 in the annex also shows that almost all points are scattered around the 45 degree line, supporting the results of the skewness-kurtosis test.

**Table 4.8 Skewness/Kurtosis tests for Normality**

| Variable | Pr(Skewness) | Pr(Kurtosis) | adj chi2(2) | P-value |
|---|---|---|---|---|
| Residual | 0.657 | 0.283 | 1.40 | 0.4958 |

### 2. Multicollinearity

Variance Inflation factor (VIF)

Variance Inflation factor (VIF) shows how the variance of an estimator is inflated by the presence of multicollinearity. If any one of the VIF's exceeds 10, then this is an indication that the associated regression coefficients are poorly estimated due to MC. The VIF and tolerance of each predictor is presented in Table 4.9.

**Table 4.9 Variance Inflation Factors of predictors (VIF).**

| Variable | VIF | 1/VIF |
|---|---|---|
| Area | 7.15 | 0.139889 |
| Fertilizer | 5.67 | 0.176291 |
| Rf4 | 4.02 | 0.24856 |
| Rf3 | 3.85 | 0.259869 |
| Rf2 | 3.42 | 0.292427 |
| Irrigation | 3.27 | 0.306064 |
| Pesticides | 3.24 | 0.308234 |
| Improved seed | 2.78 | 0.360056 |
| Rf1 | 1.63 | 0.614293 |
| Mean VIF | 3.89 | |

The VIF column shows by how much other coefficients' variances (and standard errors) are increased due to the inclusion of that predictor. We can see from the table that the VIF values are less than 10. Thus, we can conclude that MC is not a serious problem.

### 3. Linearity

Pedhazur (1997) suggests ways to detect linearity. A preferable method of detection is examination of residual plots (plots of the standardized residuals as a function of standardized predicted values). Figure 4.2 shows scatter plots of standardized residuals with standardized predicted values of cereals production. From the figure, we can detect no systematic patter for the cloud of points and the lowess line is almost overlapping with the fitted line. This situation shows there is a linear relationship between the dependent and independent variables.

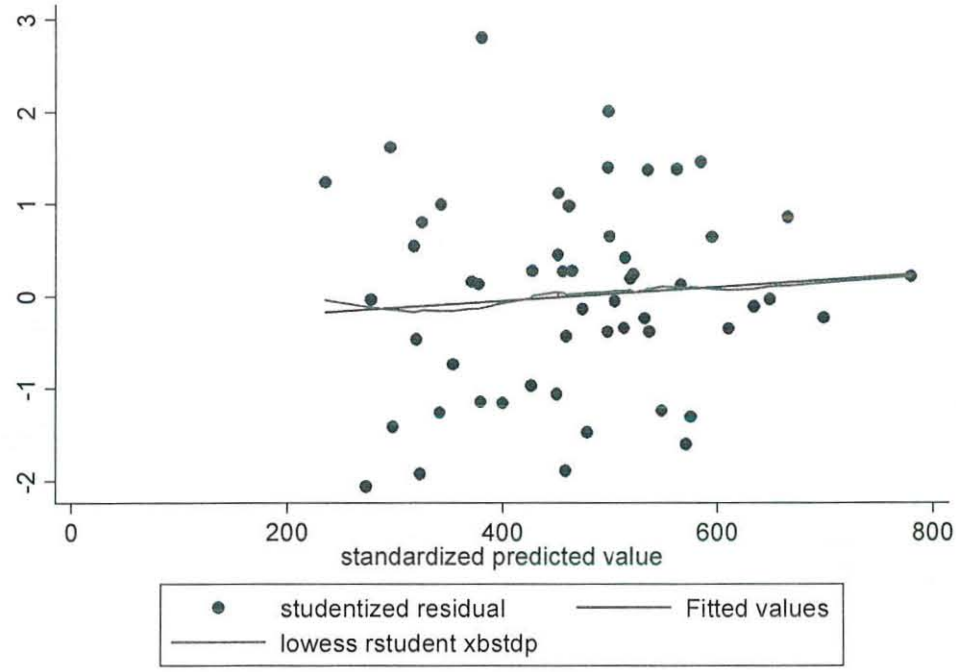


**Figure 4.2.** Scatter plots of standardized residuals by standardized predicted values.

## 4. Detecting outliers

Cook's distance is a useful measure of an aggregate influence, showing the effect of the $i^{th}$ observation on all $n$ fitted values. It is suggested that observations with large $D_i$, say $D_i > 1$, may have a significant influence on all $n$ fitted values (Cook and Weisberg, 1982). In our case, none of the $D_i$'s exceeds one. Thus we can say that there are no outliers in our data.

# CHAPTER FIVE

## SUMMARY AND CONCLUSION

The lessons from the "green revolution'' in east Asia reveal that appropriate farming practices and intensive utilization of improved technologies such as high yielding varieties (HYV), chemical fertilizers (Urea and Dap), pesticides and irrigation technologies greatly influence crop production.

The government of Ethiopia has made an effort to introduce and utilize agricultural technologies to increase food grain production. This study attempted to asses effects of new technology used and whether the variation of regional effects and time effects have an impact on cereal production. The study considered cereals since they are staple food for most people in the country. The study regions namely, Tigray, Amhara, Oromiya and SNNP were selected because of their large cereal plots and higher percentage of cereal production.

The estimated fixed effects (Least square dummy variable) regression model, and random effects regression model for both individual and time dimensions were compared for their compatibility for cereal production data.

For regional effects model, the fixed regional effects regression model is a better choice than the ordinary least squares and random regional effects model. However, for time effects model, the random time effects model is a better choice for cereal production data. Using the selected model, different factors that influence cereal production were assessed. New technological inputs like, fertilizer, improved seed variety, pesticides and regional effects have a significant impact on cereal production although the magnitudes of their influences vary.

In general, the results of the regression analysis showed that cereal production is highly influenced by the total area covered by cereals, quantity of fertilizer used, improved seed variety, pesticides and regional differences. There is a need for these factors to promote agricultural production. The other two factors namely, irrigation, first and forth quarter rainfall have no statistically significant effect on the total cereal production of the country. But they may have significant impact on different crops separately like area under irrigation has significant effect on wheat and maize production.

The time effect and quarterly average rainfall on cereal production were also analyzed using random time effects model. Considering different years ( 1995-2008), the time effect and the third quarter (July, August, September) average rainfall have a significant effect on cereal production.

## Recommendations

Several factors affect cereal production in Ethiopia. From the results of this study, the following points are recommended for sustainable increase in cereals production.

- At present the benefits of irrigation are getting recognitions in many places. Irrigation increases crop intensity and gives higher production from existing farms. These benefits are reflected in the economy as a whole in the form of increasing production. Therefore, expanding the irrigation practice which is significant for wheat and maize production has potential role for increasing crop production in Ethiopia.
- Increasing investment in developing and transferring appropriate new technologies like utilization of improved seed varieties, chemical fertilizers pesticides for all crops under cultivation is also recommended for increasing cereal productions.
- Expansion of cultivated land is also taken as a potential source of cereals production increase. Therefore, it is recommended that, increasing area sown has a significant role for sustainable increasing cereal production.

# REFERENCES

Alemu, Z.G. (2005): Causes of instability in Cereal Production in Ethiopia. Working paper, Department of Agricultural Economics, University of Free State.

Alemu, Z.G., Oosthuizen, L.K. and Van Schalkwyk, HD. (2004): Agriculture versus non agriculture in the Ethiopian economy: a study of comparative performance. University of Free State, Department of Agricultural Economics. www.uovs.ac.zg.

Appleton, B.A. (1996): Education and Agricultural Productivity. Working papers WPS/96-5.

Baltagi, B.H (2005) : Econometric Analysis of Panel Data. Third Edition, John Wiley and Sons Ltd, England.

Benin, S. and Echui, S (2003): Determinants of Cereal Diversity in Communities and on Household Farms of The Northern Ethiopian High lands, 03-14.

Breusch, T.S. and pagan, A.R (1980): Testing for random effects and special lag dependence in panel data models. Journal of Statistics and Probability letters, V 78, pp 3304-3306, University of Connecticut.

Chow, G.C. (1960): Tests of equality between sets of Coefficients in two Linear Regressions. **www.princeton.edu/~gchow/.**

Cook, R.D and Weisberg, S (1982): Residuals and Influence in Regression. Chapman and Hall University of Minnesota.

Cohen, J.M. (1987): Integrated Rural Development. The Ethiopian Experience and the Debate. The Scandinavian Institute of African Studies, Uppsala.

Croppensted, A. and Mulat Demeke (1997): Applied Economics. Volume 29, issue 9, pages 1217-1226. Addis Ababa University, Ethiopia.

CSA (2008): Annual report of Agricultural Sample Survey. Addis Ababa, Ethiopia

Deaton, A. (1995): Working paper on Panel data and Econometric Tools for Development Analysis. Amsterdam and New York, Princeton University, Elsevier Science Bv.1995, V 3A, pp 1785-1882.

Eicher, C.K and Staaz, J.M (1990): International Journal of Educational Development. Volume 17, Issue 3, pages 303-311.

Eshete Bogale (1995): Determinants of productivity of small holder farmers in Ethiopia. The case of wheat production in two districts of Sidama zone. Unpublished MSc. thesis, Addis Ababa University.

FAO (1998): Crop and Food Supply Assessment Mission to Ethiopia. Annual report on food production and socio-economic situations.

FAO (2008): Crop and Food Supply Assessment Mission to Ethiopia. Annual report on food production in 2008 and socio-economic situations.

Greene, W.H. (2000): Applied Econometrics Analysis. Fifth Edition. Stern School of Business, New York University. Pearson Education LTD, Inc., Upper Saddle River, New Jersey 07458.

Gujarati, D.N. (2004): Basic Econometrics. Fourth Edition, the McGraw-Hill Companies, New York.

Hsiao, C. (2003): Analysis of panel data, Second edition, University of Southern California.

Jaeger, W. (1991): The impact of policy in African agriculture. An empirical investigation. WB working Paper No. 147.

Kibra,B., Ariga,J., Olwande,J. and Jayne,T.S. (2008): Trends in Kenyan agricultural Productivity (1997-2007). Egerton University, Nairobi.

Klevmarken, N.A. (1989): Journal of European Economic Review. V 33, Issue (2-3), pp 523-529.

Lewis, W.A. (2004): "Economic Development with unlimited supplies of Labour". Working paper, University of Manchester.

Maddala, G.S. (1992): Introduction to Econometrics. Second edition, University of Florida and Ohio State University, Macmillan Publishing Company, New York.

MEDaC (Ministry of Economic Development and Cooperation) (1999): Survey of the Ethiopian economy. Review of post reform developments (1992/93-1997/98).

National Bank of Ethiopia (2006/2007): Annual report of Ethiopian economy.

Netsanet Asfaw (1998): Factors that affect the use of Agricultural Credits. Unpublished MSc. thesis, Addis Ababa University.

Osborne, J. W., Christensen, W. R., & Gunter, J. (April, 2001): Educational Psychology from a Statistician's Perspective: A Review of the Power and Goodness of Educational Psychology Research. Paper presented at the national meeting of the American Education Research Association (AERA), Seattle, WA.

Pedhazur, E. J., (1997): Multiple Regression in Behavioral Research. (3rd ed.), Orlando, FL: Harcourt Brace.

Pickett, J. (1991): Economic Development in Ethiopia: Agriculture, the market and the state. Development Center Studies of the organization for development cooperation and development, Paris.

Reynolds, C (2002): "Crop Production Zones in Ethiopia" Production Estimates and Crop assessment Division (to Ethiopia).

Seifu Wubeshet (1998): Technical Efficiency of cereals production farmers. Unpublished MSc. thesis, Addis Ababa University.

Webb, P., Von Brown, J. and Yohannes, Y. (1992): Famine in Ethiopia: Policy

implications of coping failure at national and household levels. IFPRI, Research

Report 92.

Weir, S. (1999): The effect of education on farmer productivity in Rural Ethiopia. Working paper series/99-7, University of Oxford.

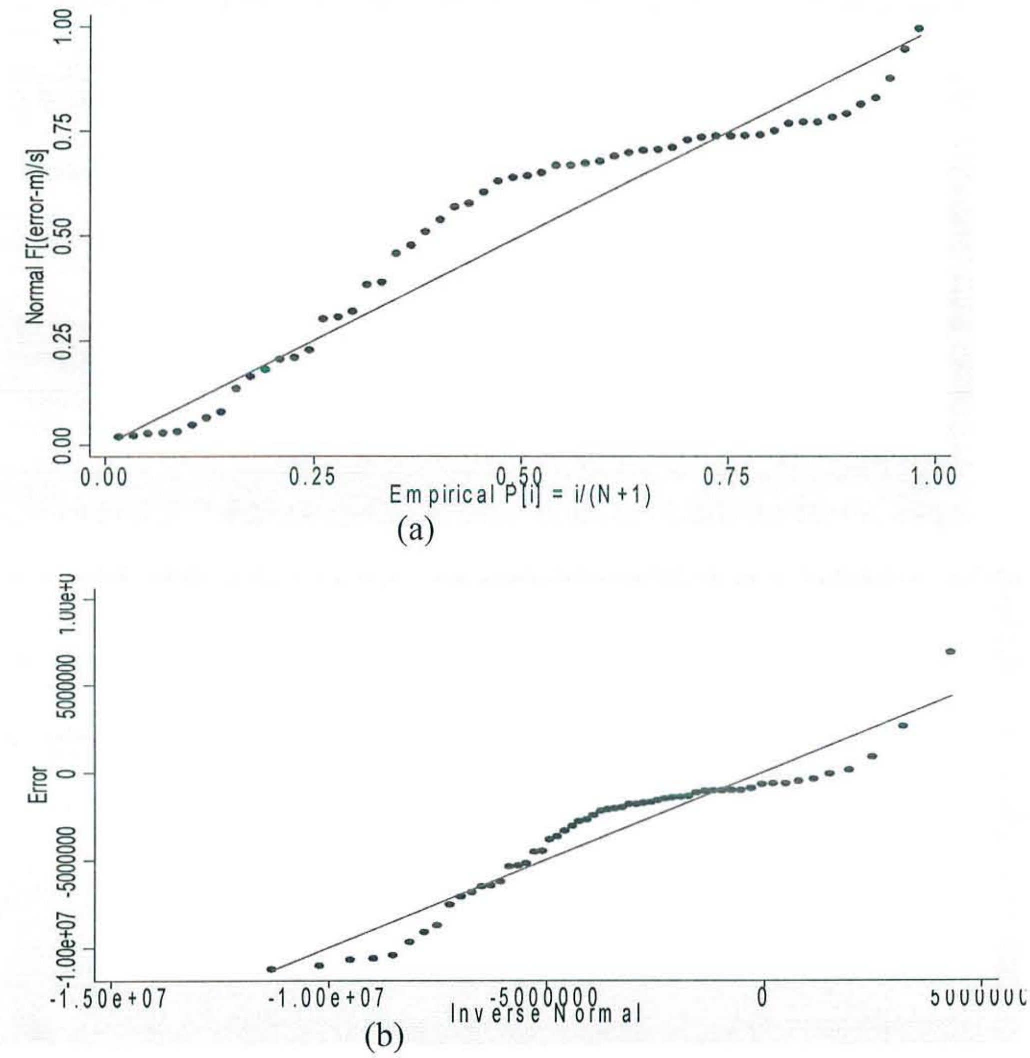
Wozniak, G.D. (1987): Human Capital, Information and Early Adoption of New Technology. The Journal of Human Resources, XXII.I, University of Tulsa.

Yao, S. (1996): The Determinants of Cereals Crop Productivity of Peasant farm sector in Ethiopia. Journal of international Development, V 8, No. 1,pp 69-82.

ANNEXES

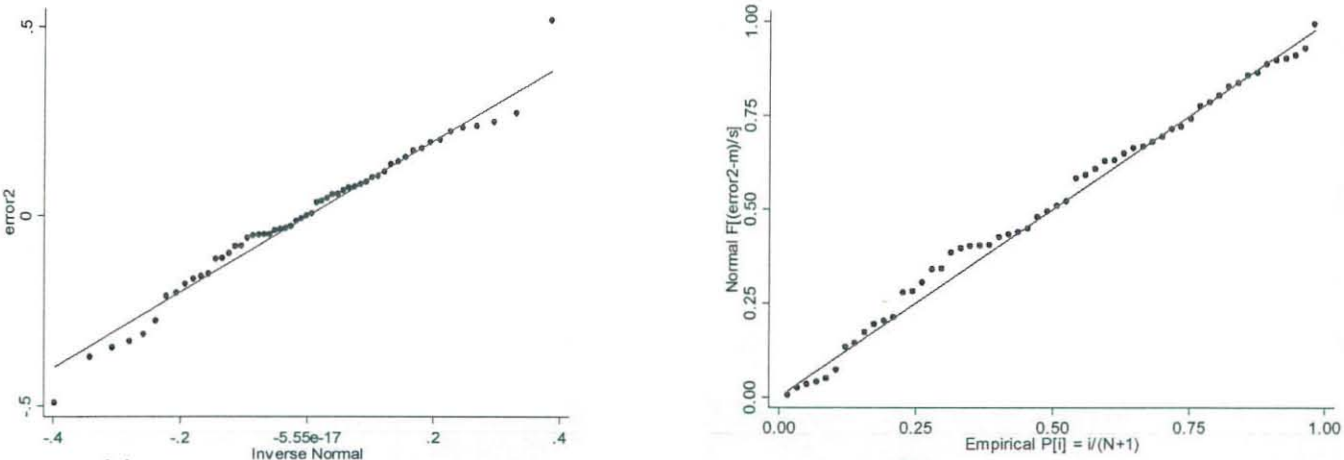
Annex 1, P-P and Q-Q plots for untransformed data

Fig.A1 P-P and Q-Q plots for error terms



Annex2, P-P and Q-Q plots for transformed data

Fig. A2 Normal probability plot. (Q-Q and P-P plots) after transformation



(a) (b)

Annex 3. Graphs of cereal production by year and region

Fig. A 3 a. Column chart for total cereal production by year

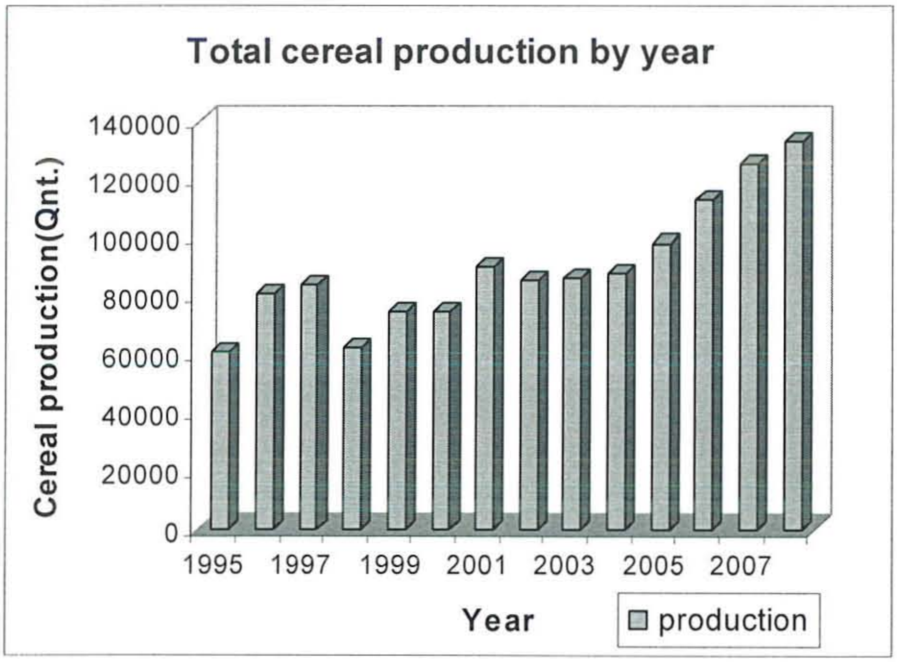


Fig.A3 b. Column chart for total cereal production by Region

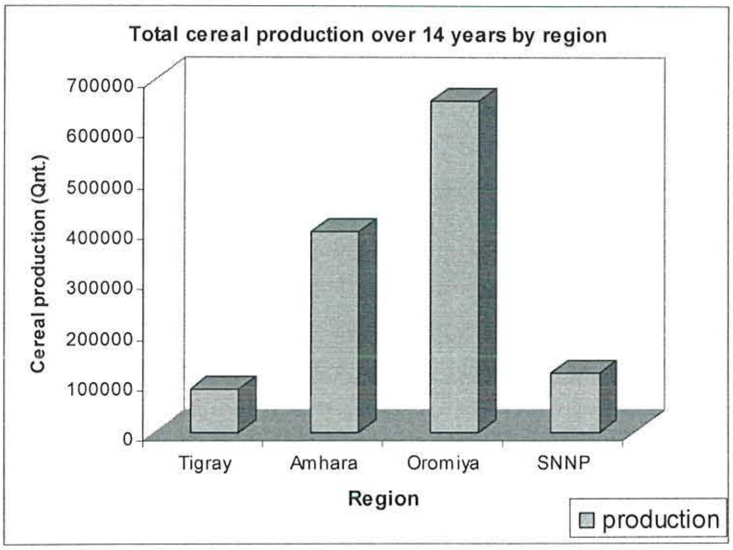
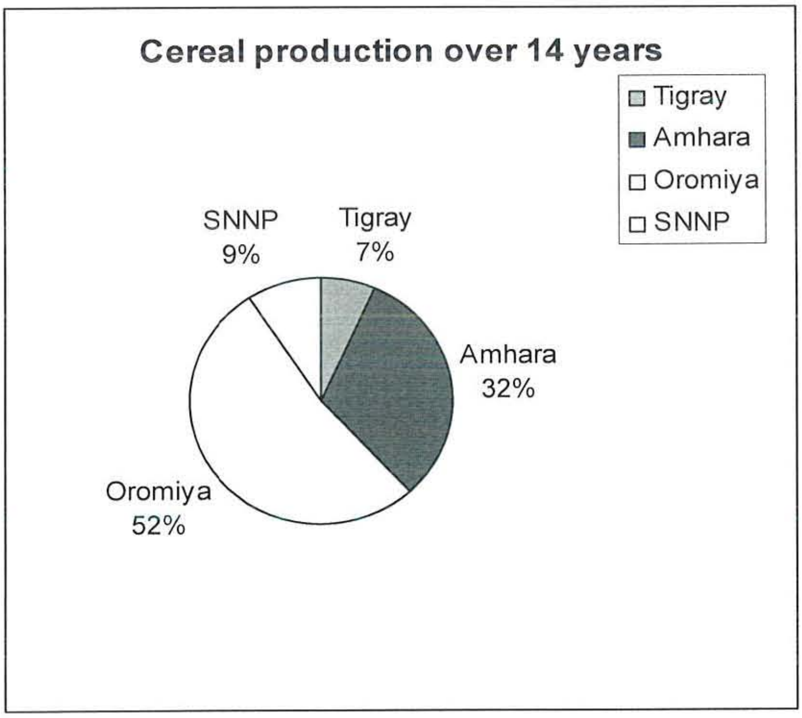


Fig.A3 c. Pie chart for total cereal production by region



Annex 4, Descriptive statistics for wheat production

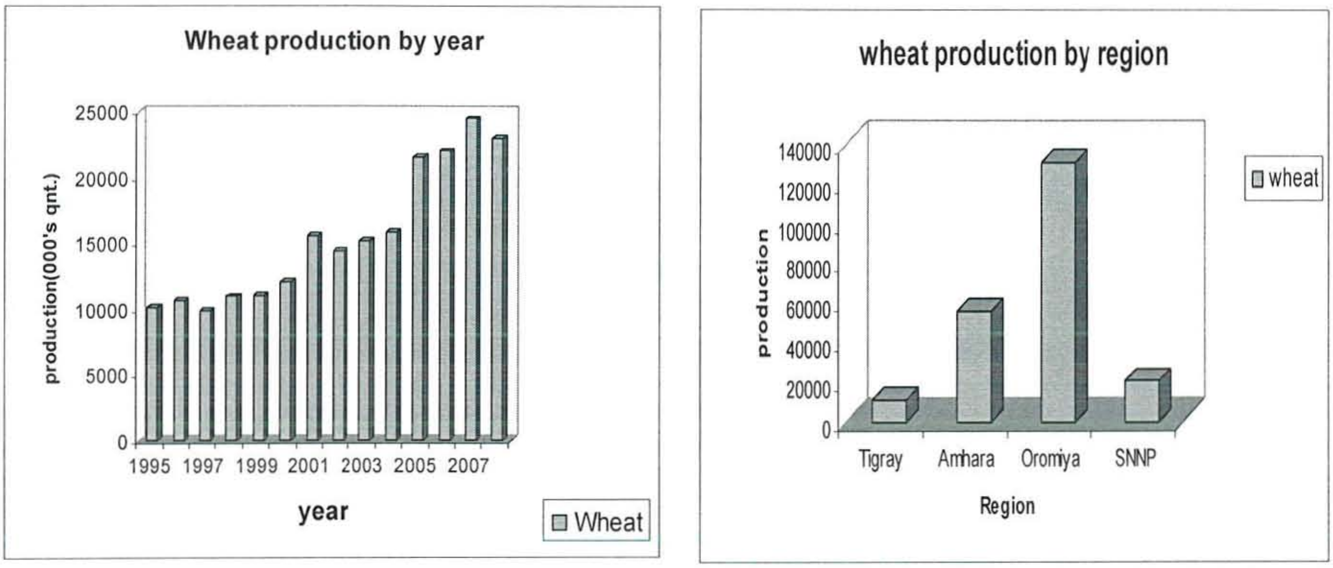
Table A4 a, Total Wheat production(in 000's qnt.)

| Year | Tigray | Amhara | Oromiya | SNNP | Total | Average |
|---|---|---|---|---|---|---|
| 1995 | 183 | 3219 | 5678 | 1122 | 10202 | 2551 |
| 1996 | 847 | 2314 | 6385 | 1106 | 10652 | 2663 |
| 1997 | 559 | 2072 | 6043 | 1255 | 9930 | 2482 |
| 1998 | 353 | 2085 | 7608 | 961 | 11007 | 2752 |
| 1999 | 578 | 2466 | 7113 | 934 | 11090 | 2773 |
| 2000 | 721 | 3272 | 7331 | 750 | 12074 | 3019 |
| 2001 | 682 | 3225 | 9905 | 1800 | 15612 | 3903 |
| 2002 | 745 | 3551 | 8773 | 1377 | 14446 | 3611 |
| 2003 | 806 | 3934 | 8951 | 1555 | 15246 | 3812 |
| 2004 | 867 | 4146 | 9129 | 1734 | 15875 | 3969 |
| 2005 | 859 | 5694 | 13028 | 1963 | 21544 | 5386 |
| 2006 | 1007 | 6075 | 13177 | 1799 | 22058 | 5515 |
| 2007 | 1481 | 6816 | 14273 | 1945 | 24514 | 6129 |
| 2008 | 1477 | 6212 | 13402 | 1975 | 23066 | 5766 |
| Total | 11163 | 55081 | 130797 | 20276 | 217317 | 54329 |

Table A4b,Percentage change in wheat production

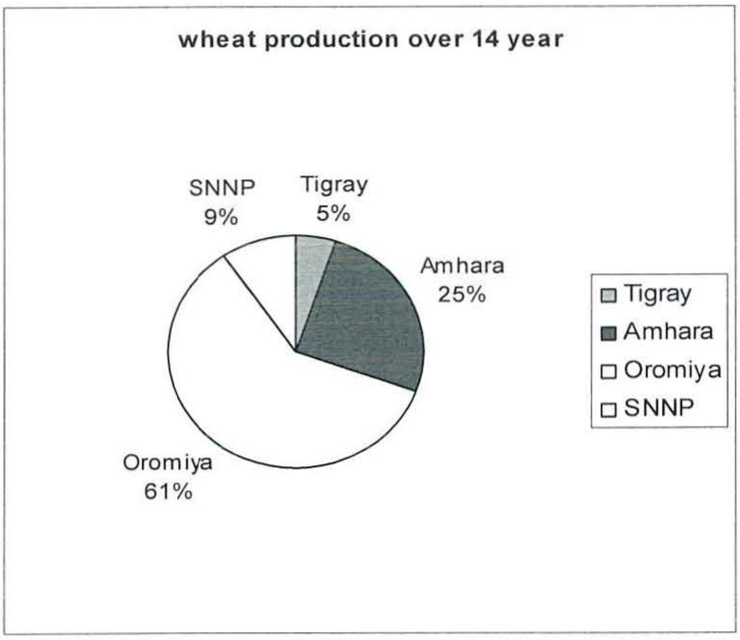
| Year | Tigray | Amhara | Oromiya | SNNP | Total change |
|---|---|---|---|---|---|
| 1995 | * | * | * | * | * |
| 1996 | 363.12 | -28.12 | 12.46 | -1.48 | 4.41 |
| 1997 | -33.92 | -10.46 | -5.36 | 13.50 | -6.78 |
| 1998 | -36.85 | 0.64 | 25.89 | -23.45 | 10.85 |
| 1999 | 63.49 | 18.26 | -6.51 | -2.76 | 0.76 |
| 2000 | 24.80 | 32.70 | 3.07 | -19.69 | 8.87 |
| 2001 | -5.40 | -1.44 | 35.10 | 139.97 | 29.30 |
| 2002 | 9.24 | 10.10 | -11.43 | -23.52 | -7.47 |
| 2003 | 8.18 | 10.80 | 2.03 | 12.95 | 5.54 |
| 2004 | 7.56 | 5.37 | 1.98 | 11.47 | 4.12 |
| 2005 | -0.89 | 37.35 | 42.72 | 13.22 | 35.71 |
| 2006 | 17.23 | 6.69 | 1.14 | -8.33 | 2.39 |
| 2007 | 47.04 | 12.20 | 8.32 | 8.08 | 11.13 |
| 2008 | -0.21 | -8.87 | -6.10 | 1.53 | -5.91 |

Fig.A4. Column chart for total wheat production by year and region



(a) (b)

Pie-chart for total wheat production by region



(c)

## Annex 5, Descriptive statistics for maize production

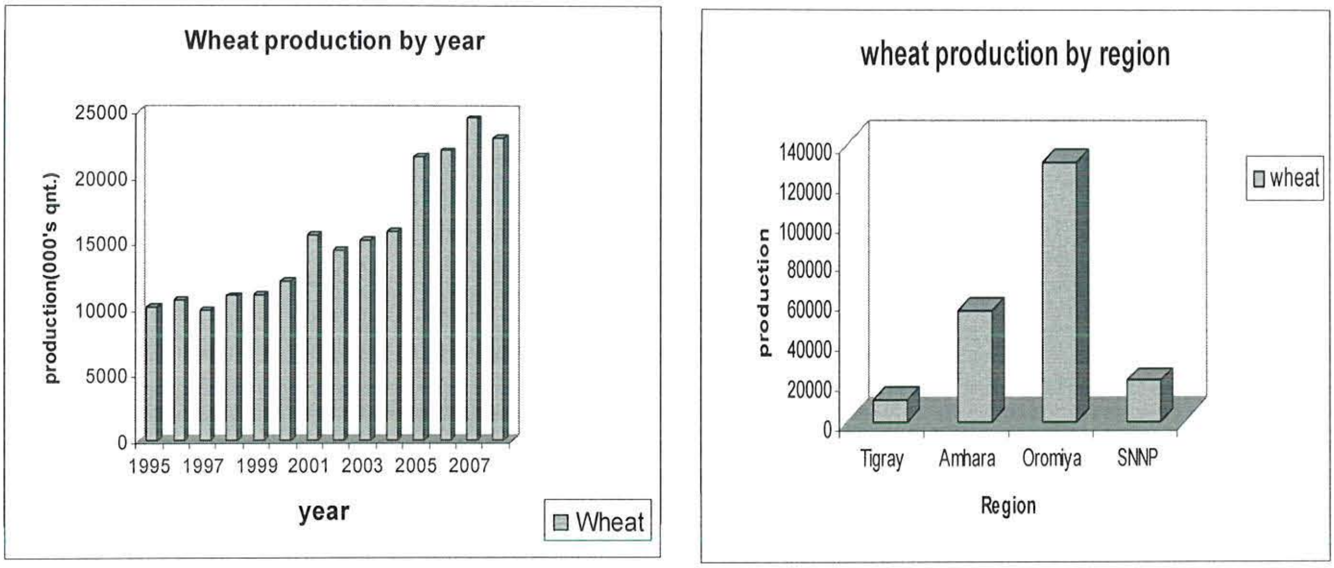
Table A5 a, Total maize production(in 000's qnt.)

| Year | Tigray | Amhara | Oromiya | SNNP | Total |
|---|---|---|---|---|---|
| 1995 | 302 | 3149 | 10599 | 2394 | 16444 |
| 1996 | 680 | 5667 | 14660 | 3597 | 24604 |
| 1997 | 828 | 5301 | 15417 | 3113 | 24659 |
| 1998 | 954 | 3773 | 10314 | 3246 | 18287 |
| 1999 | 1318 | 5285 | 13778 | 2883 | 23264 |
| 2000 | 1345 | 5249 | 14438 | 3339 | 24371 |
| 2001 | 861 | 6489 | 17247 | 5793 | 30389 |
| 2002 | 895 | 6399 | 16848 | 3323 | 27465 |
| 2003 | 856 | 5522 | 16182 | 3586 | 26145 |
| 2004 | 638 | 4644 | 15516 | 3848 | 24646 |
| 2005 | 539 | 4972 | 14526 | 3156 | 23193 |
| 2006 | 803 | 7257 | 20317 | 3740 | 32118 |
| 2007 | 1027 | 8688 | 22158 | 4360 | 36232 |
| 2008 | 942 | 9212 | 21470 | 4398 | 36022 |
| Total | 11987 | 81607 | 223470 | 50777 | 367841 |

Table A5b,Percentage change in wheat production

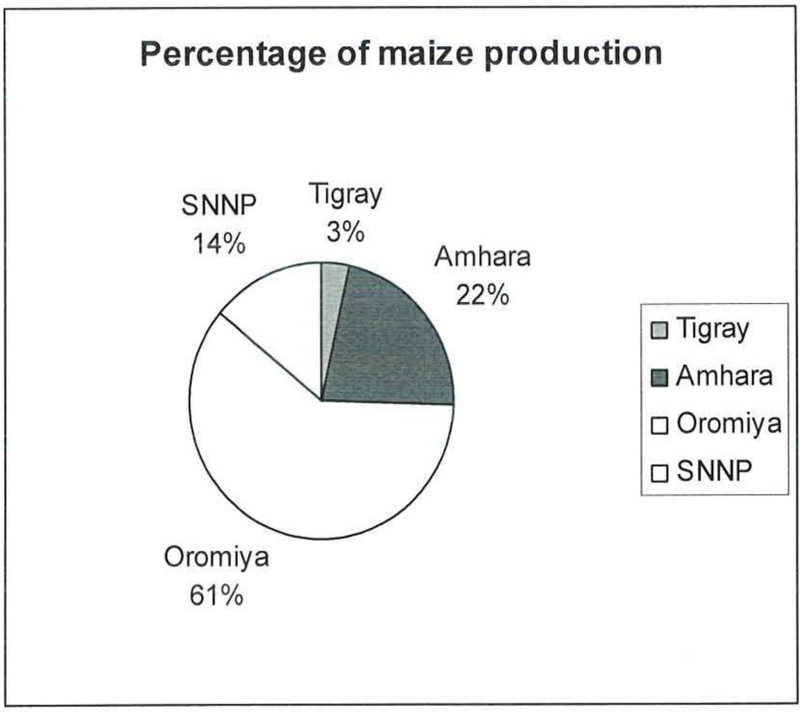
| Year | Tigray | Amhara | Oromiya | SNNP | Total change |
|---|---|---|---|---|---|
| 1995 | * | * | * | * | * |
| 1996 | 125.07 | 79.98 | 38.32 | 50.21 | 49.62 |
| 1997 | 21.76 | -6.46 | 5.16 | -13.43 | 0.22 |
| 1998 | 15.24 | -28.83 | -33.10 | 4.27 | -25.84 |
| 1999 | 38.20 | 40.07 | 33.58 | -11.18 | 27.21 |
| 2000 | 2.04 | -0.67 | 4.79 | 15.81 | 4.76 |
| 2001 | -36.01 | 23.62 | 19.46 | 73.47 | 24.69 |
| 2002 | 4.00 | -1.38 | -2.31 | -42.64 | -9.62 |
| 2003 | -4.33 | -13.71 | -3.95 | 7.91 | -4.81 |
| 2004 | -25.50 | -15.89 | -4.12 | 7.33 | -5.73 |
| 2005 | -15.47 | 7.06 | -6.37 | -18.00 | -5.89 |
| 2006 | 48.99 | 45.96 | 39.87 | 18.52 | 38.48 |
| 2007 | 27.80 | 19.72 | 9.06 | 16.57 | 12.81 |
| 2008 | -8.21 | 6.04 | -3.11 | 0.88 | -0.58 |

Fig.A5 Column chart for total maize production by year and region



(a) (b)

Pie-chart for total maize production by region



(c)

Annex 6, Plots of estimated values versus residuals

Fig A6 a. Estimated values versus residuals plots for fixed regional effects model

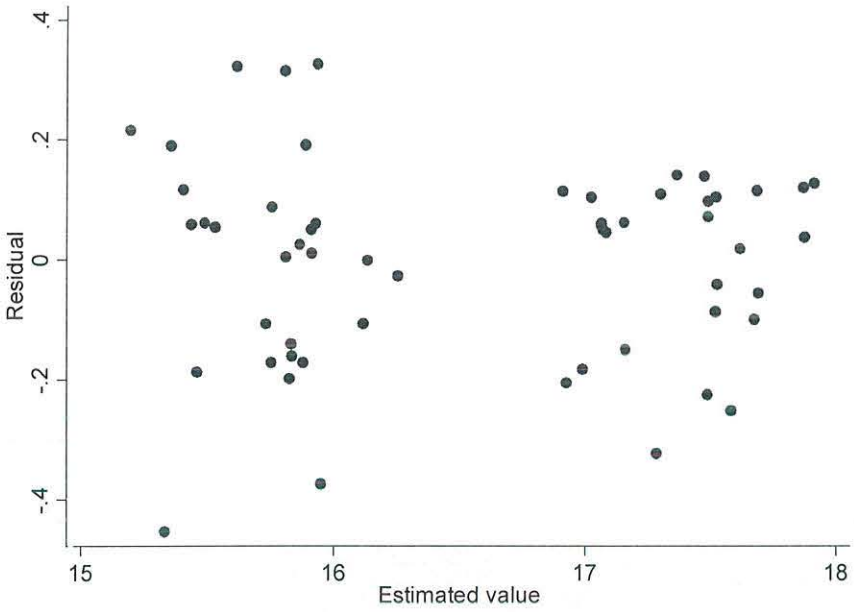
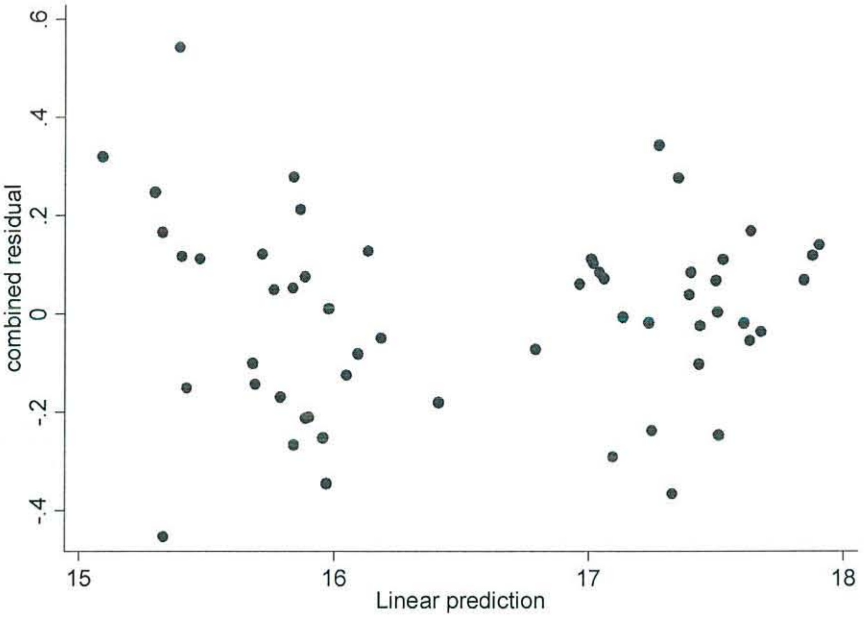


Fig A6b. Estimated values versus residuals plot for random time effects model



# DECLARATION

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

Name: Legesse Abate

Signature: ..............................

Place: Faculty of Science, Addis Ababa University

Date: August 2009

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

University Advisor.

.............................. 28/08/09 ..............................

Dr. EMMANUEL GEBREYOHANNES