EFFECT OF SOYBEAN COMMERCIALIZATION ON INCOME OF SMALLHOLDER FARMERS IN BUTERE SUB-COUNTY, KENYA

MARTHA WANJIRU GACHUHI

A Thesis Submitted to the Graduate School in Partial Fulfillment of the Requirements for the Masters of Science Degree in Agribusiness Management of Egerton University

EGERTON UNIVERSITY

NOVEMBER, 2021

DECLARATION AND RECOMMENDATION

Declaration

This thesis is my original work and has not been presented in this university or any other for the award of a degree.

Signature:  Date: 26/10/2021

Martha Wanjiru Gachuhi

KM19/14457/15

Recommendation

This thesis has been submitted with our approval as University supervisors.

Signature:  Date: 28/10/2021

Prof. George Owuor, PhD

Department of Agricultural Economics and Agribusiness management

Egerton University

Signature:  Date: 26/10/2021

Dr. Edith Gathungu, PhD

Department of Agricultural Economics and Agri-Business Management

Egerton University.

COPYRIGHT

DEDICATION

I dedicate this thesis to my parents Mr. Samuel G.Maina and Mrs. Mary Wambui G.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the Almighty Father above for the gift of life, courage, strength, confidence, patience, determination and peace throughout my entire study period. Further, I would like to express my heartfelt gratitude to Egerton University for giving me this study opportunity. Special thanks also go to the entire staff of the Department of Agricultural Economics and Agribusiness Management under the good leadership of Professor Patience Mshenga for their sincere, steadfast, and honest support, and for granting me a favourable learning environment since I enrolled for my studies. I would also like to convey my sincere gratitude and utmost appreciation to my supervisors Professor George Owuor and Dr. Edith Gathungu for their tireless and invaluable effort in guiding and supporting me during the entire research period. Indeed, their constant support and comments were significant in ensuring the successful completion of this study. Overwhelmingly, I want also to extend my sincere appreciation to my family for their love, prayers, financial and moral support throughout the entire study period. Heartfelt gratitude also goes to my colleagues from Egerton University for their positive criticism and useful ideas during the entire period of study and research. I also wish to thank the enumerators who assisted me during data collection. Special thanks also go to Mr. Felix and Mr. Ogutu for their support and guidance during reconnaissance and data collection. Lastly, to all parties who had input in this thesis from its inception to the final production, who are not mentioned above, thank you so much for your support. Indeed, your efforts and contributions in my research cannot be exhausted, may God bless you all.

ABSTRACT

In developing countries, Kenya included, agricultural commercialization has remained one of the ways to improve farmers' welfare. A result, Kenyan government through many efforts has tried to commercialize soybean for its nutritional and industrial value. Despite the efforts, the level of soybean commercialization in Butere Sub County has remained low for unknown reasons. Furthermore, information about effect of soybean commercialization on household income remain scanty. Therefore, study was geared towards filling these gaps, by determining factors influencing soybean commercialization, factors influencing the choice of soybean main market outlet, as well as the effects of soybean commercialization on household income of smallholder farmers in Butere Sub-County, Kenya. Data were gathered from 201 smallholder soybean farmers using a semi structured questionnaire in a face to face interview. Double Hurdle, multinomial logistic and Endogenous Switching Regression models were used for analysing objective 1, 2 and 3, respectively. Descriptive results reveal that soybean commercialization level was 56.72 % among soybean producing households in the study area. Double hurdle model results disclose that, schooling years, total land size under soybean production, and the number of extension visits positively influence decision and intensity of soybean commercialization at $p<0.01$. However, age of household head, primary occupation of household head, household size and distance to the nearest market center have a negative effect on soybean commercialization decision and intensity at $p<0.05$. Multinomial logistic regression results reveal that household size has a negative influence while education level of household head and access to credit show a positive influence on the choice of Middlemen at $p<0.1$. Additionally, the age of household head, marital status of household head, and soybean price per unit have a negative, as number of extension visits and access to credit show a positive effect on the choice of choice of cooperatives over direct marketing at $p<0.100$. Endogenous Switching Regression model reveal that off-farm income and group membership have positive effects on annual household income per capita for commercialized soybean households at $p<0.010$. In conclusion; among other factors, years of schooling, land under soybean production and extension visits positively influence soybean commercialization. Years of schooling and credit access positively influence the choice of brokers and cooperatives in contrast to consumers. Finally, soy bean commercialization has a positive effect on household income. Enhancing extension services, credit access and group membership are recommended.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

CGIAR	-	Consultative Group for International Agricultural research
DHM	-	Double Hurdle Model
ESR	-	Endogenous Switching Regression
FAO	-	Food and Agriculture Organization
FAOSTAT	-	Food and Agriculture Organization Corporate Statistical Database
FIML	-	Full Information Maximum Likelihood Method
GDP	-	Gross Domestic Product
HA	-	Hectare
HCI	-	Household Commercialization Index
ICRISAT	-	International Crops Research Institute for the Semi-Arid Tropics
KES	-	Kenyan shillings
MLR	-	Multinomial linear regression
MNL	-	Multinomial Logistic Regression
NACOSTI	-	National Commission for Science, Technology and Innovation
OECD	-	Organization for Economic Co-operation and Development
OLS	-	Ordinary Least Squares
SSA	-	Sub-Saharan African

CHAPTER ONE
INTRODUCTION

1.1 Background

The agricultural sector plays an important role in contributing to food security and economic development in Sub-Saharan Africa (SSA). According to Organization for Economic Co-operation and Development- Food and Agriculture Organization (OECD-FAO, 2016), the agricultural sector, on average, contributes 15% of the region's GDP (Otsuka *et al*., 2013). Furthermore, the sector provides employment to over half of SSA's population. Notably, agriculture is the main source of livelihood in the rural areas. It is crucial to note that smallholder farmers, producing mainly for subsistence purposes, dominate the sector. Despite the prominence of agriculture in the region, the response of the sector to the increasing demand for food as a result of demographic changes is slow (OECD-FAO, 2016 & Van Ittersum *et al.*, 2016). This is attributed to low productivity of agricultural resources, policy inconsistencies, rapid population growth, limited diversification, climate change, and poor infrastructure.

Crop production is the most dominant agricultural sub-sector in SSA. The crop sub-sector has over the last two decades accounted for an average of 85% of the total value of agricultural production in the region (OECD-FAO, 2016). However, Otsuka *et al*. (2013) and Ravallion (2009) observed that despite many regions in the world recording increased crop productivity, the average yields in SSA have stagnated. This is largely attributed to poor soils caused by extensive nutrient mining and limited access to input and output markets by smallholder farmers. In addition, there is much emphasis on cereal crops compared to other crop enterprises. Therefore, given the relative importance of the crop sub-sector in SSA, regional governments can enhance the contribution of the agricultural sector to the improvement of rural livelihoods by encouraging diversity in crop mix. Importantly, agricultural stakeholders should emphasis legume production. This is because legumes are important sources of protein and oil for human diet and contribute to the improvement of soil fertility.

Kenya is one of the SSA countries with the most developed agricultural sector. However, the sector is also dominated by smallholder farmers who have limited capability in dealing with crop production challenges. For instance, in dealing with input use constraints and the effects of climate

change, smallholder farmers in Kenya are increasingly practicing intercropping. According to Wambugu and Muthamia (2009), about 90 percent of Kenyan farmers grow maize, which is the staple food crop in the country. Smallholder farmers commonly intercrop maize with common bean as an integrated soil fertility management practice and as a strategy for increasing productivity per unit area. This has made common bean to be the most popular legume in Kenya, pushing other legumes to the periphery.

However, it is crucial to acknowledge that given the human and industrial demand for other legumes such as soybean, *Glycine max (*L.) *Merrill*, research organizations are increasingly developing and disseminating soybean technology with the view of encouraging soybean production. Overtime, leguminous crops such as soybean represent an essential economic development opportunity in many developing countries. These crops enable smallholder farmers to gain higher yields and increase their profits and family income which further help them promote improved living standards (Osmani *et al*., 2014). Therefore, in recognition of the role of soybean in improving the nutritional status of poor rural households, governments and private organizations are increasingly promoting soybean production in developing a region.

According to Mahasi *et al*. (2011), the Kenyan soybean production remains low and insufficient to meet human consumption and industrial demand. Soybean production and importation are insufficient in meeting human consumption demand of between 10,000 and 14,000 metric tonnes and industrial demand of 120,000 metric tonnes per annum (International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), 2013; Odendo *et al.* 2013). The domestic production of soybean was 2866 and 2804 metric tonnes in 2012 and 2013, respectively (FAO, 2017). Furthermore, Kenya imported about 1926 and 6346 metric tonnes of soybean in 2012 and 2013, respectively. This makes Kenya the lowest producer and the highest importer of soybean in East Africa. This occurs against the backdrop of Kenya having a high potential of producing soybean. Western Kenyan counties such as Kakamega, Bungoma, Busia, and Vihiga are the highest producers of soybean in Kenya. However, these counties are yet to meet their production potential. Other soybean producing areas include Uasin Gishu, Kisii, Embu, Nyeri, Meru, and Machakos counties. Mahasi *et al*. (2011) estimates that about 2,500 hectares of land are under soybean

production. Further, Kenyan's production per hectare is averagely 0.8 tonnes, which falls short of an average potential of 1.5 to 3.0 tonnes per hectare.

The multiple uses of soybean make it an important industrial crop. Supposedly, this should be an incentive for farmers to commercialize soybean production as they take advantage of the increasing prices resulting from increased demand. For instance, soybean is increasingly used internationally in biofuels and in the manufacturing of nutritious and healthy food product (Kim & Moschini, 2018). Domestically, soybean is used as raw material for manufactured food products (AGRA, 2017). It is one of the major source of protein thus nutritionally good for human diets. These are opportunities for encouraging soybean commercialization especially among smallholder farmers in Kenya. Commercialization would be a major milestone for improving soybean productivity, improving smallholder farm incomes, and meeting the domestic consumption and industrial demand for soybean.

Demographic shifts have led to an increase in demand for safer and nutritious food like soybean (Consultative Group for International Agricultural research (CGIAR), 2016). Consequently, this has led to increasing industrial demand for soybean as a raw material for production of human food and livestock feed. Therefore, increasing farmer participation in soybean value chain in western Kenyan counties is paramount to the utilization of the regions' soybean production potential. This would meet the increasing demand for soybean processed products such as beverages, soy milk, soy meat, and livestock feed. In addition, increasing production and marketing of soybean in Kakamega County is key to improving the livelihood, income and food security of smallholder farmers. To this effect, a number of efforts have been focused on enabling smallholder farmers to participate in soybean market or commercialized it especially in Butere sub-county. This is because in Butere Sub-county, the majority of the farmers are practicing soybean production as part of their livelihood and, with low commercialization efforts, reasons that motivated the choice of the sub-county as a study area. The government strategies being implemented in the sub-county aim at encouraging production and commercialization of soybean for not only meeting the increasing demand for soybean and soybean products, but also for improving smallholder farmers' livelihoods.

1.2 Statement of the Problem

Increasing soybean production and marketing remain key to improving the livelihood, income and food security status of smallholder farmers. In this regard, the government of Kenya among other stakeholders through a number of programs have come up with strategies to boost soybean commercialization in Kenya, especially in Butere Sub-County in Kakamega County. Despite these efforts as well as an increased marketing opportunity, commercialization of soybean in Kenya, has remained low for unknown reasons thus resulting to yield gap. Importantly, commercial oriented soybean farmers, tend to engage in at least a market outlet as a selling platform. Soybean farmers, therefore, sell through a market outlet whenever they decide to commercialize, however, literature on factors influencing the choices of different market outlets remains scarce. This implied that there was a limited research on factors that impede smallholder farmers' decisions to commercialize soybean production as well as those that influence the choice of marketing outlets. Therefore, understanding factors influencing soybean commercialization and choice of market outlets among smallholder farmers through this study remains valuable in meeting the current and future demand for soybean and soybean products, while contributing to improved rural livelihoods. To this effect, this study sought to determine the factors influencing soybean commercialization and the choice of soybean market outlets, as well as its effects on income of smallholder farmers.

1.3 Objectives of the Study

1.3.1 General Objective

To contribute towards improved smallholder farmers' livelihood through enhanced production, commercialization and consumption of soybean in Kenya.

1.3.2 Specific Objectives

i. To determine factors influencing intensity of soybean commercialization among smallholder farmers in Butere Sub-County, Kenya.

ii. To determine factors influencing the choice of soybean market outlets among smallholder farmers in Butere Sub-County, Kenya.

iii. To determine the effect of soybean commercialization on smallholder farmers household income in Butere Sub-County in, Kenya.

1.4 Research Questions

i. What are the factors influencing intensity of soybean commercialization among smallholder farmers in Butere Sub-County, Kenya?

ii. What are the factors influencing the choice of market outlets for soybean in Butere Sub-County, Kenya?

iii. How does soybean commercialization influence household income of smallholder soybean farmers in Butere Sub-County, Kenya?

1.5 Justification

Smallholder agriculture in Kenya is characterized by disproportionate commercialization. Whereas commercialization of cereals and beans is relatively high, it is low for important legumes such as soybean. Hence, addressing the soybean commercialization challenges requires an understanding of the basic concept of production and marketing. Therefore, the starting point for understanding commercialization should be grounded on knowledge of production and marketing. Much needs to be discerned about low farmer participation in soybean production in Kakamega County despite its production potential. Furthermore, gaining knowledge and understanding of marketing systems is important for designing interventions aimed at enabling smallholder farmers to increase production and enhance market participation.

Commercialization is important for the expansion of soybean production in response to the current and anticipated demand for soybean and soybean products. First, commercialization of soybean is appropriate strategy for improving smallholder farmer livelihoods by contributing to increased farm incomes as a result of improvements in production. Second, commercialization of soybean is also instrumental to strengthening of smallholder farmer links to the other sectors which further enhance employment. Lastly, commercialization of soybean encourages successful financing of smallholder agriculture which would, in turn, stimulate investment in diverse farm enterprises. These help in improving smallholder farmers' incomes and wealth. This in turn contributes in meeting broader economic goals such as the big four agenda and sustainable development goals (SDGs).

1.6 Scope of the Study

This study was restricted only to smallholder farmers producing soybean. The study was conducted in Butere Sub-County in Kakamega County in Western Kenya due to fact that the region is characterized as a high potential area for soybean production in the country.

1.7 Limitations of the Study

The study encountered some difficulties in getting farmers to participate in the survey, however, with a help of well-known village guides and thorough consent seeking process, the expected sample size was achieved. Lastly, language barrier and poor record keeping among some farmers was also assumed to affect the accuracy of data collected, however, through probing was done by the enumerators to avoid the problem. However, the assumption that the respondents were to provide correct, truthful and honest information was met. Lastly, the study assumes that whenever a soybean farmers engage or sell his produce via a market outlet it is, therefore, regarded as being commercial oriented.

1.8 Definition of Terms

Commercialization: Farmers decision to engage in any process of managing or running an activity principally for financial gain. For this study, it measured farmers' decision to participate in soybean marketing as well as extent of such participation measured in term of Household Commercialization Index. This also involve soybean farmers' participation in an market outlet as a selling avenue.

Gross margin: Gross margin refers to farm total sales revenue minus its cost of goods sold used as a proxy of farm income.

Household: Is an independent male or female and his /her dependants who must have lived together for at least six months. The members are answerable to one person as the head and share the same eating arrangement

Household Commercialization Index: Refers to proportion of soybean sold over total amount of soybean produced.

Income: Amount of money received from farm investment measured in terms of gross margin.

Market outlet: Individual or institution that the farmers sell soybean output i.e direct consumers, cooperatives, middlemen and processors. Participation in a market outlet means that a farmer is commercial oriented.

Smallholder farmer: A farmer who does his/her production activities on an area of land less than 0.4 ha of land either owned or leased.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

Many studies have been done on soybean, most of which focused on its dissemination, production, marketing as well as its contribution to households' welfare across different agricultural zones. However, this review concentrated much on theoretical and past empirical literatures on factors determining leguminous crop production, participation in output market, commercialization level, as well as its effectiveness in improving smallholder welfare. Without taking into consideration a given set timeframe, the bulk of this literature was obtained from online databases and some printed texts, where more than twenty articles from both international and local studies were reviewed. From the literature review, several key topics on leguminous crop production and commercialization specifically for soybean were identified and discussed, however, only relevant topics were discussed in detail because of word constraints.

This chapter is organized as follows: The first section provides determinants of leguminous crop production. The second section provides insights on smallholder commercialization decision and levels. The third section discusses determinants of smallholders' choice of output market outlet. The fourth section discusses the effect of leguminous crop production and commercialization on improving smallholder welfare. The fifth section provides a theoretical framework upon which the study was developed. Finally, the chapter ends with a conceptual framework outlining the determinants of soybean commercialization, choice of soybean market outlet, and the resulting welfare effect.

2.2 Determinants of Leguminous Crop Production

Overtime, leguminous crops such as soybean represent an essential economic development opportunity in many developing countries. These crops enable smallholder farmers to gain higher yields and increase their profits and family income which further help them promote improved living standards (Osmani *et al*., 2014). Therefore, in recognition of the role of soybean in improving the nutritional status of poor rural households, governments and private organizations are increasingly promoting soybean production in developing a region. Despite this effort, soybean production in developing nations is significantly low as compared to major producers such as the

United States. It is against this background that Houngnandan *et al.* (2015) used ordered probit regression to study determinants of soybean production system in Benin. The study revealed that smallholder farmers do not actually practice correct soybean cropping practices. Again, they reported that availability of improved seeds, fertilizers, plant density, fallow, and farmer characteristics are major factors significantly influencing soybean production in Benin. However, this study failed to apply strong empirical and methodological approach to study farmers' decision to participate in soybean markets.

Mbanya (2011) also assessed the constraints to soybean production in Northern Ghana. The study revealed that technical and socioeconomic factors were the major constraints to soybean production. Poorly organized marketing channels as well as lack of processing infrastructure constrained farmer participation in soybean production. Furthermore, the study established that low output prices negatively affected soybean production. Specifically, low market prices were a major disincentive to increased soybean production. Also, low adoption of improved soybean as well as pest and disease constrained soybean production. Despite the contribution of Mbanya (2011) to the understanding of constraints to soybean production, the study employed a weak empirical and methodological approach which limited consistent and efficient estimation of causality. This study overcame this weakness by using an empirical specification that would establish the extent of the influence of different factors on soybean commercialization.

Research into economic and institutional factors that influence farmer allocation of land to particular crops is increasing. To this end, Garrett *et al*. (2013) assessed the effect of biophysical, economic and institutional conditions on the variation of the size of land allocated to soybean production in Brazil. The study used the von Thünen's spatial market model to estimate the relative effect of each determinant of regional variations in the area of land under soybean. The study then used two-stage least squares model to analyze the association between soybean yield and planted area. Contrary to the expectation of the study, land tenure did not have an observable effect on the area of land under soybean production. On the other hand, access to credit and marketing of soybean through cooperatives had a positive and statistically significant effect on the planted area. Further results established by the study was a positive effect of soybean yields on land allocation to soybean production. The study results indicate that increasing soybean production can greatly

be enhanced in developing countries by facilitating smallholder farmers' access to credit and enhancing their participation in local institutions. These would enable them to overcome constraints to soybean production and commercialization that they are currently facing.

Similarly, Kadam and Suryawanshi (2011), Mbanya (2011), and Singh *et al*., (2012) investigated constraints to soybean production in India. Besides the two studies investigating individual-specific and socioeconomic characteristics of soybean growers, they also focused on institutional factors that facilitate or hinder soybean production. Kadam and Suryawanshi (2011) and Singh *et al*. (2012) found that lack of technical knowledge and information on soybean and soybean production was a major constraint to farmer participation in soybean production. Furthermore, the studies found consistent results with respect to the farmer access to credit. The studies indicated that lack of access to credit in time and high-interest rates negatively impacted on soybean production. Further emphasized was marketing, institutional and economic factors such as high input costs, cooperative membership, and prices which were important towards influencing farmer participation in soybean production and productivity. However, these studies were limited to the extent that they relied on descriptive statistics in identifying the constraints without establishing the extent of the association between the different factors and soybean marketing. Therefore, this study established the extent of the influence of various factors on farmers' participation in soybean marketing using appropriate econometric models.

Legumes are critical to sustainable agricultural intensification in regions that are characterized by poor yield and low use of improved technologies. Consequently, Mhango *et al.* (2013) used a mix of analytical approaches to determine opportunities and constraints to legume production in Malawi. Results indicated that legume production faces a myriad of constraints and opportunities. The study found that land devoted to legume production relative to maize area was low. Soil fertility and properties were among key edaphic constraints to legume production. Furthermore, the study findings showed that farm-level decision-making, especially with regard to input use, was a key constraint to legume production. On the other hand, Yen *et al*. (2013) found that low soil fertility and land degradation were major constraints to agricultural productivity in Northern Vietnam. In contrast, Yen *et al*. (2013) findings revealed that low level of technology use coupled with land fragmentation negatively impacted on agricultural production.

Although the Kenyan agricultural sector is relatively developed compared to other developing countries in Sub-Saharan Africa, it is not devoid of challenges and has not reached its potential. Besides maize and beans, Kenya has a huge potential of enhancing the contribution of other cereals and legume crops to addressing the persistent food shortages. Informed by this observation, Muui *et al*. (2013) examined the socioeconomic factors that influence production and use of sorghum in three lower eastern counties in Kenya. Results indicated that a majority of farmers in the region were resource-poor hence lacked income to purchase fertilizer and chemicals for sorghum production. Furthermore, inadequate access to quality seed as well as high susceptibility of existing varieties to pest and diseases were negatively associated with sorghum production. However, the study relied on measures of central tendency such as Analysis of Variance and Least Significant Differences. These methods are not adequate in circumstances when the interest is estimating cause-effect relationship. This attempted to address this limitation using cause-effect econometric technique that does not merely rely on mean and mode but is capable of producing robust cause-effect association between variables.

Despite the important role of legume crops in global food security, developing regions in the world are worse hit by scarcity of important legumes as a result of production constraints. Building on this observation, Daryanto *et al*. (2015) conducted a meta-analysis of literature data on the association between drought and legume production in different regions in the world. The study findings showed that different legume crops adopt differently to drought which then influences production and productivity. The response of the legumes to water shortage as a result of drought influenced eventual allocation of land to legume production. The study results seem to indicate that biophysical conditions such as drought that eventually impact on legume yield possibly affect farmers' decision to prefer production of one legume over another. However, the study did not clearly reveal how drought eventually influences farmers' decision to produce a given legume.

Agada (2015) acknowledged that despite Nigeria being the leading soybean producer in Sub-Saharan Africa, a large proportion of the Nigerian population still suffers from nutrition and health-related problems. To this effect, Agada (2015) sought to highlight factors constraining soybean production in Benue State. Consistent with Kadam and Suryawanshi (2011), Mbanya (2011), and

Singh *et al*. (2012), Agada (2015), findings indicated that socioeconomic, individuals and production related problems were the major constraints to soybean production. The study further reiterated that the ineffectiveness of the extension systems in the region slowed down improvement in soybean production and productivity.

Understanding farmer knowledge, perception, and behaviorare critical to understanding farming choices and decisions and in unraveling factors associated with differential adoption of agricultural technologies. Thus, Abtew *et al*. (2016) evaluated farmers' knowledge and perception of legume pests with the view of understanding production constraints that impact on legume production in the eastern region in Kenya. The study used a multivariate ordered probit regression to estimate the effect of a number of factors on farmers' knowledge and perception of legume crops. Results from the analysis indicated that education and location were statistically significant in explaining farmers' knowledge of legume pests. Lack of access to extension and training negated farmers' knowledge of legume pest management practices and thus influenced farmer perception of legume production. The results of the study seem to underscore the importance of knowledge and training in the technical aspects of legume production. This study built on this study by increasing its scope and focusing on technical, socioeconomic, institutional and household characteristics that facilitate or hinder soybean commercialization.

Kathage *et al*. (2016) analyzed the adoption of improved maize technology in Tanzania with the view of establishing reasons for low adoption rates. Employing parametric regression in an average treatment effect framework, the study found that there was no significant association between farmers' awareness or knowledge of hybrid seed and adoption. The findings were inconsistent with Garrett *et al*. (2013), Kadam and Suryawanshi (2011), Singh *et al*. (2012), Mbanya (2011), that access to credit is a major constraint to adoption of a crop technology. Kathage *et al*. (2016) also found no statistically significant association between access to credit and farmers' uptake of improved maize varieties. Furthermore, the study found evidence that access to extension information was only important in one region and not in another. As a result, Kathage *et al*. (2016) concluded by noting that expected returns on crop investment possibly explain adoption decision.

Soybean remains one of the world's major leguminous crops, produced for its high proteins content, oil, and its ability in soil-fertility improvement. To this end, Nyongesa *et al.* (2017) studied gender factor in soybean production in Kenya using various techniques. From the study, it was identified that market access was the most limiting constraint to soybean production. The technical, allocative and economic-efficiencies were also studied. The study revealed that farmers' economic-inefficiencies in soybean production arise from factors such as age, gender, input availability among others. It was added that increased use of identified factors was necessary to positively impact smallholders' efficiency thus resulting into higher soybean output and profitability. However, the authors failed to adopt appropriate methodological approach that combines production decision, marketing decision and extent of market participation as provided in the current study.

2.3 Determinants of Commercialization Decision and Extent of Commercialization

In Sub-Saharan countries, Kenya included, smallholder farmers find it cumbersome to participate in various output markets. This is because various barriers and constraints exist which frequently reduce their incentive for market participation. Based on this observation, a number of studies have been conducted to determine factors influencing farmers' decision to participate in output markets as wells as the extent of participation as discussed in this section. Starting with the definition, Lapar *et al*. (2003) and Pingali (1997) defined market participation as any market-oriented activity conducted to promote the sale of produce. Lapar *et al*. (2003) added that market participation represents produce sales as a fraction of total output. Makhura *et al*. (2001) posited that market participation can be referred to us as commercialization. In this regard, increased market participation involves the transition from subsistence farming to a market engagement network. This, in turn, involves frequent use of markets as well as market infrastructures for the purpose of trading agricultural products and services.

According to Pingali and Rosegrant (1995), agricultural commercialization usually involves a gradual transformation process from subsistence to fully commercialized agriculture with the main aim of realizing various welfare effects. This implied that agricultural commercialization involves farm households' decision to be fully commercialized when targeting output markets in their production decisions rather than being linked only to the volume of output they would probably

send to the market due to surplus production. Also, participation in agricultural market or commercialization of agriculture may also relate to the production of agricultural goods to meet particular output market demands as well as the supply of inputs for production to farmers (Makhura *et al.*, 2001). This, therefore, implies that commercialization operates both on the output and input aspects of agricultural markets (Gebremedhin & Moti, 2010). This normally happens when farmers participate or engage in the output markets, by trading the largest fraction of their produce to get income for purchasing inputs and other needs (Brian & Barret, 2014). They added that higher market participation by smallholder farmers can increase yield by providing information, money and another incentive for purchasing inputs.

Several studies have been carried out to analyze the various factors influencing smallholder farmer market participation and extent of market participation (commercialization) in developing countries. These studies have identified factors related to farm household characteristics, market factors and transaction costs, access to assets, natural factors, product characteristics, trust, and social capital, among other institutional services. Hlongwane *et al.* (2014) estimated a logistic regression model of market participation for small-scale maize farmers in South Africa. They found that market infrastructure and information, access to credit, gender, and marital status significantly stimulated, while the distance to output market, farming experience, and an external source of income significantly hindered smallholders' market participation. They further reiterated that distance to output market represents a proxy for transaction costs which negatively affect market participation. This implies that the greater the distance to the nearest output market the higher the transaction costs. Rahut *et al.* (2015) as examined roles of human capital, physical capital, and the transaction cost in market participation among subsistence farmers in the Himalayas. The results showed that ethnicity, land size, livestock assets, the gender of the household head, distances to market, education, and location are important determinants of market participation. Despite the contribution of studies by Hlongwane *et al.* (2014) and Rahut *et al.* (2015), these studies failed to apply the correct econometric model to capture the extent of participation in the output market.

Sharma *et al.* (2016) used descriptive statistics to examine the level of smallholder commercialization in Assam measured using the Household Commercialization Index (HCI). It

was revealed that the level of commercialization in Assam ranged from 63.3% to 74 %, which was significantly affected by access to market and higher farm size. According to Egbetokun *et al.* (2017), improvement of farmers' productivity remains an important activity to increase agricultural output in developing countries. However, this cannot be realized without markets that would effectively and efficiently join the increasingly specialized activities of hundreds of widely dispersed farmers into an integrated national economy. To this effect, Egbetokun *et al.* (2017) used descriptive statistics and probit regression to analyze investigates factors influencing market participation among maize farmers in Ogbomoso zone of Oyo state in Nigeria. They found that age, household size, gender, marital status, farming experience, membership of the association and educational level are the major factors significantly influencing farmers' decision to participates in the output market. Masuku *et al.* (2001) examined factors influencing the farmers' decision whether to participate or not to participate in maize markets in Swaziland. They found that the decision of whether to participate or not to participate in maize markets was influenced by income from off-farm sources. Alene *et al*. (2008) also added that non-farm income sources result in the contribution of more marketed output especially when such income is invested in farm technology and other farm improvements. Again, these studies failed to apply the correct econometric model to capture the fact that farmers' decision to participate in the output market, as well as the extent of market participation, is contingent on a decision to produce first.

Numerous constraints exist that restrict smallholder farmers' access to markets as well as preventing them from taking advantage of various market opportunities. Based on this observation, Kyaw *et al*. (2018) used the Heckman two-stage selection model to analyze the factors influencing smallholder rice farmers' decisions to participate in the agricultural market. According to Kyaw *et al.* (2018), a significant relationship was found between commercialization and factors such as household head age, household income, household size, education status, price of rice, distance to market, total produce of rice, membership of farmer organization, ownership of livestock, access to roads, market information and access to extension services. However, this study failed to consider the influence of technological development and other value-adding innovative factors on the extent of farmer participation in the agricultural market. In a similar study, Zamasiya *et al.* (2014) estimated a Heckman's Probit model with sample selection; first separating factors affecting a farmer's decision to participate in soybean markets or not, and secondly, the factors that affect

the intensity of a farmer's participation. The results presented that use of inoculants and improved soybean seed varieties positively and significantly correlated with farmers' participation in soybean markets. Other factors significantly influencing the extent of commercialization were ownership of radios, quantity produce, marital status, and education.

In another study, Seyoum *et al.* (2011) employed Robust Ordinary Least Squares(OLS) model toidentify factors determining the degree of commercialization or extent of market participation among potato growers Ethiopia. The found that farm size allocated to potato, access to market information and access to irrigation significantly affected the extent of market participation. Barrett (2008), studying the extent of market participation in staple grains in sub-Saharan Africa, found that smallholders' barriers to extent of participation in staple grains markets were mainly livestock, land, capital and improved technologies such as farm implements needed to generate surplus produce that increasingly influence the extent of market participation. However, the study failed to consider the fact that the degree of commercialization is dependent on the decision to participate in production and marketing. This study, therefore, addressed this gap by considering an econometric model that will capture all the decisions at once.

Kimei *et al.* (2017) employed Double Hurdle model to determine both factors influencing the decision to commercialize and level of commercialization (Household bean commercialization index) of beans among Smallholder Farmers in Rwanda. The results show that factors such as age, the quantity of beans produced, number of livelihoods a household head engages in, type of beans and market information influenced the decision to commercialize common beans in Rwanda. However, level of commercialization was influenced by a number of crops a household cultivates, the number of beans produced, market distance and number of livelihoods of a household.

In a related study, Mbitsemunda and Karangwa (2017) adopted a probit model to analyze the factors influencing market participation among bean growers and Tobit model to analyze the factors influencing to the extent of market participation in Nyanza District, Rwanda. This study was motivated by low market participation of bean Growers in Rwanda. Based on their results, market experience, bean quantity produced, and access to credit positively and significantly influenced the farmers' decision to participate in output market while the distance to nearest

market, access to off-farm activities, and age negatively and significantly influenced market participation decisions. On the other hand, marketing experience, quantity produced and selling price positively and significantly influenced the extent of market participation, whereas land size was found to have a negative significant influence on the quantity of bean marketed. Despite the contribution by Kimei *et al.* (2017) and Mbitsemunda and Karangwa (2017) to the understanding of determinants of market participation and level of commercialization, the studies employed a weak econometric approach which does not incorporate decision to produce as a contingent decision. The current study also built on these studies by increasing their scope to focus on technical, socioeconomic, institutional and household characteristics that influence soybean production decision, market participation decision, and extent of market participation at once.

Burke *et al.* (2015) argued that policies influencing farmers' market participation decisions may also induce additional farmers to become producers, in that previous double-hurdle model or Heckman's do not explicitly control for this possibility. In this regard, Burke *et al.* (2015) employed a triple-hurdle model to include an initial stage that includes production decision in addition to market participation decision and role that these producers choose to play (or not) in the marketplace. Based on their results, training, rural electrification, and improved grazing practices influence diary production decision and dairy market participation decision.Though the previous studies made significant contribution in understanding the factors affecting smallholder market participation and level commercialization, they did not explicitly analyze the determinants of soybean market participation and extent of commercialization The current study, therefore, built on this study by increasing its scope and focus to accommodate a double hurdle approach to model the influence of technical, product, institutional and household socioeconomic characteristics on soybean market participation decision, and degree of commercialization at once. This will help in minimizing the effect of unobserved characteristics and individual household heterogeneity and of the sampled households.

2.4 Determinants of Farmers' Choice of Output Market Outlet

A number of previous studies have identified factors influencing smallholder farmers' choice of marketing channel. These studies have estimated factors related to farm household characteristics, social capital, transaction costs, product characteristics, and access to assets to affect farmers'

choice market channel. Smallholder farmers face a myriad of challenges in choosing both input and output markets. This problem is profound in developing nations where smallholder farmers have a limited ability to supply farm products to the market. However, the ability of farmers to choose the marketing channel differs across time and space, and the type of marketing channel. Tsourgiannis *et al*. (2008) explored factors that influenced the choice of the market channel among sheep and goat farmers in Greece. Results from a logistic regression found that sales price, payment period and loyalty significantly influenced farmers' choice of marketing channel. In addition, Tsourgiannis *et al*. (2008) established that the volume of milk, farm income, as well as debt and the size of the flock influenced farmers' choice of marketing channels.

Masuku *et al.* (2001) also studied factors affecting the choice of a marketing channel by small-scale maize farmers in Swaziland. They that found transaction costs associated with transport costs and distance to market have an influence on market channel choice. In this regard, they noted that the longer the distance to the output market, the higher the transport cost and the less likelihood of selling through the formal channel. Ogunleye and Oladeji (2007) also identified the factors influencing the choice of cocoa marketing channel in Nigeria. Based on their results, mode of payment, time of payment, transportation cost and price of the product were found to be significantly affecting the choice of marketing channels. Chirwa (2009) also investigated the factors that determine farmers' choice of selling maize either through private traders, neighbors or local markets. Using a multinomial logit (MNL) model, the author found an educational level to be an important determinant of farmers' choice of market channel. However, the effect varied from one channel to another. Whereas the higher level of education had no statistically significant influence on the choice of marketing channel, the primary level of education influenced the choice of private trader over selling through neighbors or relatives. The study also found the area of land under maize to be positively associated with farmers' selling maize through private traders over the other channels. Whereas distance to the market and tarmac road and access to telephone services significantly influenced the choice of maize marketing channel. However, the association between effective market maize price and choice of the market channel was statistically insignificant.

Bardhan *et al*. (2012) used MNL to analyze the determinants of dairy farmers' choice of marketing channel in Uttarakhand, India. Results of the study were consistent with research findings by Tsourgiannis *et al*. (2008). Bardhan *et al*. (2012) found evidence that an increase in the volume of milk produced influenced the farmers to shift selling through cooperatives to the point of the first sale. Furthermore, Bardhan *et al*. (2012) found evidence that extension contact influenced farmer choice and intensity of participation in the market. However, the study found mixed results with respect to the influence of distance on the market. Whereas distance to the market was negatively associated with farmer participation in the output market in both hill and plain localities, the intensity of farmer participation in selected channels in plains was not statistically significant as compared to farmers in hills.

The emergence of new market channel usually influences farmers' choice of markets. Rao *et al*. (2013) found evidence that the emergence of supermarkets in Kenya influenced farmers' choice of market channels. The emergence of supermarkets guaranteed farmers steady access to the market, hence influencing their production decision especially with regard to technology use and quality standards. In addition, supermarkets offered superior output prices which influenced farmers to upgrade with production technology and efficiency in order to produce enough for the ready market. The study also seemed to suggest that farmers' choice of supermarkets was attributed to reduced marketing risk which allowed farmers to specialize and increase the scale of production. Results by Bardhan *et al*. (2012), Chirwa (2009), Ogunleye and Oladeji (2007), and Tsourgiannis *et al*. (2008), were further re-emphasized by Xaba and Masuku (2013) in a study on vegetable marketing in Swaziland. Xaba and Masuku (2013) found that the quantity of baby corn produced and education level significantly influenced farmers to either sell their products through different wholesale market channels. On the other hand, distance to the market, contractual agreement as well as farmer membership to farmer group influenced farmers' choice on the non-wholesale channel over the wholesale channel.

In another study, Umberger *et al*. (2015) investigated how buyer attributes influenced potato farmers' choice of the market channel in Indonesia. The study used the best-worst scaling experiment to determine the buyer attributes that influenced farmer decisions to sell to different traders. Results indicated that farmers had heterogeneous preferences for buyer attributes. Farmers

preferred selling to buyers who offered financial incentives such as credit and price premiums for quality sorting of potato. Farmers also preferred buyers who offered inputs, information, and commitment to the sale agreement. These findings were consistent with results of a study by Mmbando *et al*. (2016) on factors that influence farmers' choice of market channel for maize and pigeon pea output in Tanzania. Mmbando *et al*. (2016) found that access credit significantly influenced farmers' choice of market channel. Mmbando *et al*. (2016) findings that farmers' extension services positively and significantly influenced the choice of the market channel were also consistent with research findings by Bardhan *et al*. (2012). Furthermore, results from a MNL analysis showed that transaction costs, household wealth and social capital significantly affected farmers' choice of market channel.

Panda and Sreekumar (2012) studied determinants of choice of a marketing channel by vegetable farmers in India. According to their study, access to market information positively and significantly impacted the choice of both formal and informal market channels. They noted that there exists a significant relationship between road infrastructure and choice of formal market channel. Zivenge and Karavina (2012) also studied household characteristics influencing the choice of market channels. They reaffirmed that farmers with high levels of education are always risk-averse and good negotiators thus can search and understand production and marketing information needed to adjust production and marketing systems related to the different market channel. Martey *et al.* (2012) argued that product characteristics such perishability and price of a product have an influence on the choice of a marketing channel by smallholder farmers in Ghana. This implies that markets associated with higher prices provide farmers with an incentive to the sell. They added that output prices are normally determined the choice of the rural market compared to urban market; where perishable products require shorter output market channels to reduce the risk of damage.

In another study, Alemu *et al.* (2012) employed multinomial regression to analyse factor affecting farmers' participation in cooperatives or contracts channel in Ethiopia. They indicated that a positive relationship exists between the choice of contracts channel and the wealth of the household. In this regard, they presented that farmers with relatively higher assets or wealth such as land size normally possess a lower degree of risk aversion, thus more than willing to choose

new market channel opportunities. According to Umberger *et al*. (2015), farmers' participation in producer groups in terms of social capital or collective action helps them reduce transaction costs for choosing better market channels. Through farmer groups, farmers find it easy to access new technologies, obtain market information, and tap into high-value market channels. Despite the contribution of these studies on determinants of choice of marketing channel, no empirical evidence exists about why soybean producers choose a specific market channel to sell their produce. This study, therefore, addressed this research gap using multinomial logistic regression in order to inform important practical policy interventions needed to improve soybean farmers' market channel choice as well as their welfare. Again, though these studies made significant contribution in understanding the determinants of smallholder choice of marketing channels, they did not explicitly analyse the determinants of choice of main market channel by soybean farmers in Kenya. The current study, therefore, adopted correct methodological approach to unmask some determinants of choice of marketing channels resulting from unobserved individual household characteristics and heterogeneity of the sampled households.

2.5 Welfare Effect of Crop Production and Commercialization

The recent emphasis on the importance of soybean production and marketing is majorly focused on its nutritive value. However, it is critical to acknowledge the potential of soybean production and commercialization in contributing to improved household welfare. As such, there is a need to evaluate the potential impact of soybean production and marketing, by extension, the effect of crop commercialization on household income. Thus, research interest in the role of soybean in the nutritional welfare of households needs to be accompanied by an evaluation of its effect on income. However, there exists a dearth of research that has demonstrated the economic benefits of crop production and market participation in Sub-Saharan Africa.

Sanginga *et al*. (1999) using multivariate analysis, presented that adoption of soybean had a clear positive effect on household income generation and distribution, resource use, gender relations, human capital development, social equity, material welfare, and other social activities in the community. They further presented that soybean consumption, women's production of soybean, and income earned from participation in soybean markets, had a significant positive effect on both the short and long-term nutritional status indices. According to World Bank (2015), greater crop

production and market access for smallholder farmers remains one of the routes to raise farm incomes and escape poverty. Therefore, enhancing crop production and market participation of small-scale farmers is considered as a significant way to reduce poverty and food insecurity in developing countries.

Zhou *et al.* (2013) also noted that agricultural commercialization or market participation remains the only way to stimulate increased agricultural productivity that is essential for poverty eradication and ensure food security. Minten and Barrett (2008) argued that increased agricultural productivity resulting from agricultural market participation leads to increased incomes that motivate farmers to join the advanced market economy and become even fully commercialized. In this regard, commercialization helps in increasing commodity diversity in the market as well as specialization at both regional and farm levels (Zhou *et al.*, 2013).

Asfaw *et al*. (2011), also used multiple econometric techniques to provide a micro insight into the effect of improved pigeon pea production on household welfare in rural Tanzania. The study found that adoption of improved pigeon pea had a potential of improving household welfare by reducing poverty. Asfaw *et al*. (2011) established that improved pigeon directly led to higher incomes for adopting households as compared to non-adopter as a result of increased productivity. The higher incomes derived from pigeon production negatively impacted on poverty status for the adopting households. Weinhold *et al*. (2013) used panel dataset to examine the effect of soybean production on income, poverty, and inequality in Brazil. The study found the positive and statistically significant effect of soybean production on rural household income. They added that increased participation of farmers in soybean production increased median rural incomes which further translated into a reduction in rural poverty. However, the study found evidence that increased soybean production increase rural inequality, suggesting mixed effects of soy production on socioeconomic indicators. However, the study did not adequately disaggregate the impact of soybean production at the household level which possibly resulted in the biased generalization of the findings. This study solved this shortcoming by using household-level data which will allow disaggregation of welfare effects of soybean commercialization at household level.

In another study, Mmbando *et al*. (2013) evaluated the welfare effects of smallholder farmer participation in maize and pigeon pea in Tanzania. Using propensity score matching (PSM) and endogenous switching regression for impact estimation, the study found a positive and statistically significant association between participation in maize and pigeon pea market on household per capita consumption expenditure. Consistent with the findings by Asfaw *et al*. (2011), Mmbando *et al*. (2013) found that smallholder farmer participation in maize and pigeon pea market increased household income which, in turn, resulted in the improved purchase of food. In the end, there was improved food security and reduction in the level of household poverty.

Similarly to Franke *et al*. (2014) conducted an ex-ante assessment of grain-legume production systems with the view of establishing the potential benefits of legumes to farmers in Malawi. The study focused on the potential impact of soybean and groundnut production. Using cost-benefit analysis, the study found that legume production was relatively more profitable than the continuous production of maize. The study acknowledged that improved legume cultivation allowed resource-poor farmers to overcome input constraints. Cultivation of improved soybean crucially improved soil fertility which, in turn, resulted in higher legume and maize yields. Consequently, the study concluded by noting that grain legume had the potential of becoming not only food crops but also major cash crops for medium and high resource-endowed farmers. The study seems to suggest that legume crops have the potential of becoming the major source of income for farmers irrespective of their resource endowment. However, the article did not directly establish the link between legume commercialization and farm income, which this study estimated.

Rokani *et al*. (2014) explored the effect of Ugandan farmers' participation in entrepreneurship training on soybean productivity and its eventual influence on household income. The study determined the causal effects of farmer participation using PSM that was estimated in an average treatment effect framework. Further established was that farmer participation in soybean entrepreneurship training was positively associated with an increase in soybean productivity. Increased soybean productivity resulted in higher household income for the participating farmers as compared to non-participants. However, the results of the study probably suffered from omitted variable bias since the study's empirical approach ignored a number of socioeconomic and institutional factors that affect household income. Therefore, there is a high likelihood that the

effect of improved soybean productivity on household income was overstated. Thus, this study overcame this empirical shortcoming by including socioeconomic as well as institutional factors as important determinants of household income.

Verkaart *et al*. (2017), using three rounds of panel data and fixed effects instrumental variables specification, analyzed the impact of chickpea adoption on Ethiopian households' income and poverty. Results from the estimation indicated that adoption of chickpea had a statistically significant relationship with a number of household welfare indicators. Adoption of chickpea had a positive impact on both per capita household income and absolute household income. Furthermore, the study postulated that input use that is associated with improved chickpea cultivation significantly resulted in higher yields which allowed households to sell surplus produce. Furthermore, the authors noted that the return to improved chickpea was significantly higher, which contributed to a significant proportion of the total income of the adopting households. Ultimately, higher income from chickpea contributed immensely to poverty reduction in rural Ethiopia.

Carletto *et al.* (2017) also argued that smallholder transition from subsistence to commercial agriculture is key for economic and welfare growth. Based on their study on the impact of commercialization measured in terms of Household Crop Commercialization Index on nutritional outcomes, they provided that a positive and significant relationship exists between commercialization and nutritional status. Jari and Fraser (2009) also argued that welfare benefits resulting from agricultural commercialization normally occur both at the micro level and macro level through spill-over effects. They argued that at the micro or household level, commercialized crop production can eliminate credit constraint usually faced by smallholder farmers. This is due to continued contacts with the various market actor which leads to the development of trust as well as the creation of credit advances such as cash cropping, free training or extension services. On the other hand, at the macro level, commercialized crop production will create markets for both inputs and outputs thus attracting investors in remote areas to supply these goods and services traders and smallholder farmers need. This implies that commercialized crop production can also create private and public investment in human capital and infrastructure development that eventually improves productivity and incomes.

Mathenge *et al.* (2010) using household-level panel data, studied factors influencing commercialization and its implication on the welfare of the poor and marginalized households in Kenya. They found that a high percentage of households who exited poverty participated in agricultural markets by selling a high proportion of their crop production. Carletto *et al.* (2017) reaffirmed that if smallholder farmers increase their extent of participation in staple food crop markets by selling more commodities through increases in productivity and market efficiencies, their incomes will increase. Muriithi and Matz (2015) noted that crop commercialization plays an important role in achieving the overall goal of poverty alleviation, food security, and sustainable agriculture, through increased productivity and farm incomes, especially among smallholder farmers in developing countries. Based on this observation, Muriithi and Matz (2015) used fixed effect estimation to assess the welfare effects of vegetable commercialization in Kenya. Based on their results, a positive association exists between commercialization through the export market and income. Again, they argued that commercialization through the domestic market channel is positively and significantly associated with welfare measured in terms of income and asset holdings, depending on the model specification.

Camara (2017) using Endogenous Switching Model, identified potential welfare impact of smallholder farmers' participation in the cereals market in Guinea. He identified that smallholder farmers' participation in the cereals market significantly increases household income. Rahut *et al.* (2015) also identified a positive and significant association between the smallholders' market participation and the household income, as well as a negative and significant association between smallholders' market participation and poverty levels. They argued that through commercialization, smallholder farmers are able to increase their well-being by increasing income levels as well as reducing poverty levels considerably.

Consequently, the reviewed empirical literature shows the effect of agricultural commercialization or market participation analyzed using different approaches for different crops in Kenya and beyond (Camara, 2017; Carletto *et al.*, 2017; Mathenge *et al*., 2010; Muriithi & Matz, 2015; Rahut *et al.,* 2015;). Although these previous studies made a significant contribution in understanding the effect of smallholder market participation or level commercialization in Kenya and beyond, no

study has empirically analyzed welfare effect of soybean commercialization using a comprehensive household commercialization index in Kenya. Therefore, there was the need to study welfare effect of soybean commercialization in Kenya using a more comprehensive farm-level index and more robust and recent impact assessment econometric models, which this study estimated.

2.6 Theoretical Framework

Rural farm households engage in different activities in order to diversify their sources of income. Thus, such households allocate a stream of resources to the different activities based on the potential and actual contribution of the individual activity to household income. This implies to the welfare of a rural household which is best described as utility derived from different activities by the households given their income and the prices they face. Generally, it is accepted that marketing decisions can be analyzed within the utility maximization framework and agricultural commercialization theories. Specifically, the farmer has to make a decision to participate in soybean marketing given other alternative activity. In this context, the alternative activity can be on-farm or off-farm income generating or market-related activity. According to Barret (2008) the decision of a farmer whether or not to participate in agricultural markets are generally assumed to be derived from the maximization of utility or expected profit theories. Therefore, farmers' decision to participate in soybean marketing is based on random utility theory. Random utility theory uses some economic theory to motivate and evaluate human decision regarding binary choices (Rogers, 1995). According to the random utility theory, an individual decision or choice is anchored on the possibility that the choice or decision will maximize individual utility subject to resource constraints and utility of the available alternative (Caviglia-Harris, 2003).

In this regard, a household chooses to participate in soybean production if the expected utility that will be derived from soybean production exceed utility that would be derived from an alternative activity. Again, a household choose to participate in soybean markets if the expected utility derived from market participation exceed utility that would be derived from non-participation decision. In evaluating the expected utility, the farmer takes into account alternative and related investment as well as benefits and costs (Caviglia-Harris, 2003). Letting the expected utility of participating in soybean markets to be U_{ij} and the utility of not participating in soybean markets to be U_{ik}, the

farmer only participate in soybean marketing if the expected utility of participating in these activities is greater than the utility of not participating in the activity. Therefore, the probability that the farmer will participate in soybean markets can be expressed as:

$$\Pr(U_{ij} > U_{ik}) \tag{2.1}$$

The theory further states that these decisions depend on some observable and unobservable individual's socioeconomic, farm and institutional characteristics (Rogers, 1995):

$$y_i^* = \beta X_i + \varepsilon_i \qquad (i = 1,..., N) \tag{2.2}$$

$$y_i = 1 \text{ if } y_i^* > 0, \quad y_i = 0 \text{f } y_i^* \leq 0 \tag{2.3}$$

where y* represent latent variable which is unobserved, for instance, expected benefits or utility from smallholder farmers' decisions to participate in soybean output markets, X_i is a vector of the observed independent variables, $y_i = 1$, if a farmer participated in soybean output markets, and 0 otherwise, and ε_i is an error term which is independently distributed.

2.7 Conceptual Framework

The conceptual framework was operationalized as indicated in Figure 1, which shows the interaction of various factors that are considered to have an influence on farmers' decision to commercialize soybean as well as the extent of commercialization. A rural household engages in agricultural production to produce food for itself or sale. This study conceptualizes that participation in agricultural marketing is intended to boost agricultural productivity and livelihoods of rural households. Therefore, the contribution of soybean to household livelihood will only depend on whether it is first adopted, consumed it or sold it in a given market outlet. Therefore, the farmer's decision to participate in a given soybean market outlet will depend on the available land, costs of inputs as well as output prices (Bezu *et al*., 2014). Also, as often cited in the reviewed literature, participation in soybean market outlets can be explained by credit constraint, access to input and output market, land tenure and agro-ecological characteristics such as soil quality and rainfall (Bezu *et al*., 2014). In addition, participation in soybean market outlet is also influenced by household demographic characteristics such as farmers' marital status, age, gender, household size, education, income, and number of livestock units. Besides consumption, the household sells the surplus soybean in a given outlet which greatly impacts on household income viewed in terms of utility that is derived from participation in soybean marketing.

In turn, households' decision to be commercialized is also influenced by similar institutional, individual household demographic, and farm characteristics. However, if a household participate in soybean marketing at a given outlet his or her level of income is expected to change, due to the perceived benefits of such decisions or activity such as improved agricultural productivity, better prices, and healthy contracts. These, therefore, lead to changes in expected level of other livelihood measures.

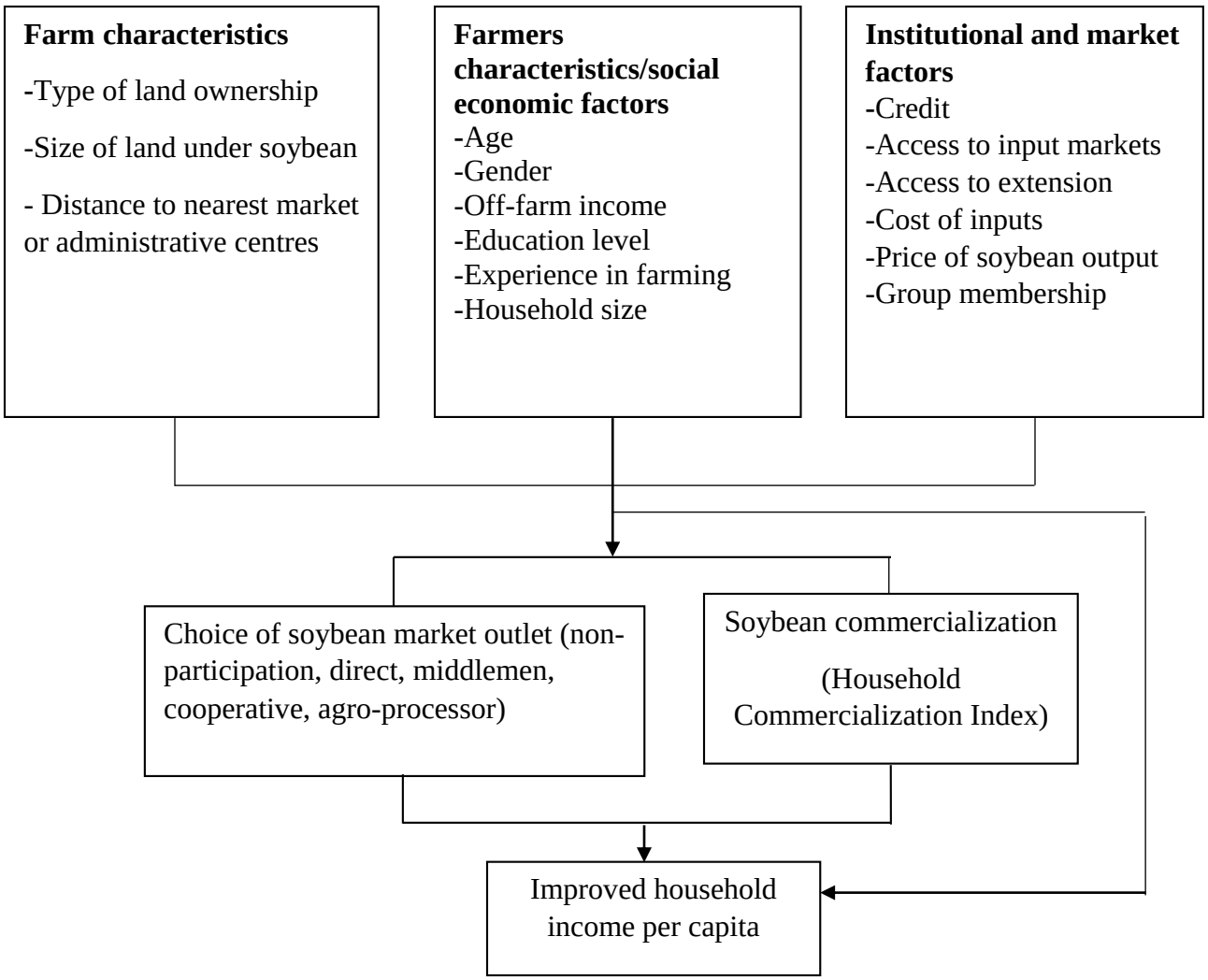

Figure 1: Conceptual Framework

CHAPTER THREE
METHODOLOGY

3.1 Introduction

This chapter presents information on the study area, research design, sampling design, sample size determination, sampling procedure, data collection methods, and analytical framework.

3.2 Study Area

This study was conducted in Butere Sub-County in Kakamega County. Kakamega County is one of the western counties in Kenya. The county has three regions, twelve sub-counties and sixty wards, two townships and several village units. The altitude of the Butere Sub-County ranges from 1240 meters to 2000 meters above the sea level and located between latitude 0° 16' 60.00" N and a longitude of 34° 45' 0.00" E. The sub-county occupies an area of approximately 1033.8km^2. Annual rainfall in the study area ranges from 1280.1 mm to 2214.1 mm per year while the temperature ranges from 18to 29degrees Celsius where January, February and March being the hottest months. Other months experiences relatively similar temperatures except for July and August which have relatively cold spells (KNBS, 2016). The choosing of Butere Sub-county in Kakamega County was motivated by the fact that it has the majority of the farmers practicing soybean production as part of their livelihood and, also where soybean commercialization is relatively low.

There are two main ecological zones in Kakamega county namely; upper medium and lower medium that covers central and northern parts of the county such as Ikolomani, Malava,Lurambi, Navakholo,and Shinyalu that practice intensive maize, tea, beans and horticulture production mainly on small-scale. The second ecological zone, the lower medium covers a major portion of the southern part of the county which includes, Butere, Kwhisero, Mumias East, Mumias West, and Matungu. In this zone, the main activity is sugar production with some farmers practicing maize, sweet potatoes, tea, groundnuts, and soybean.

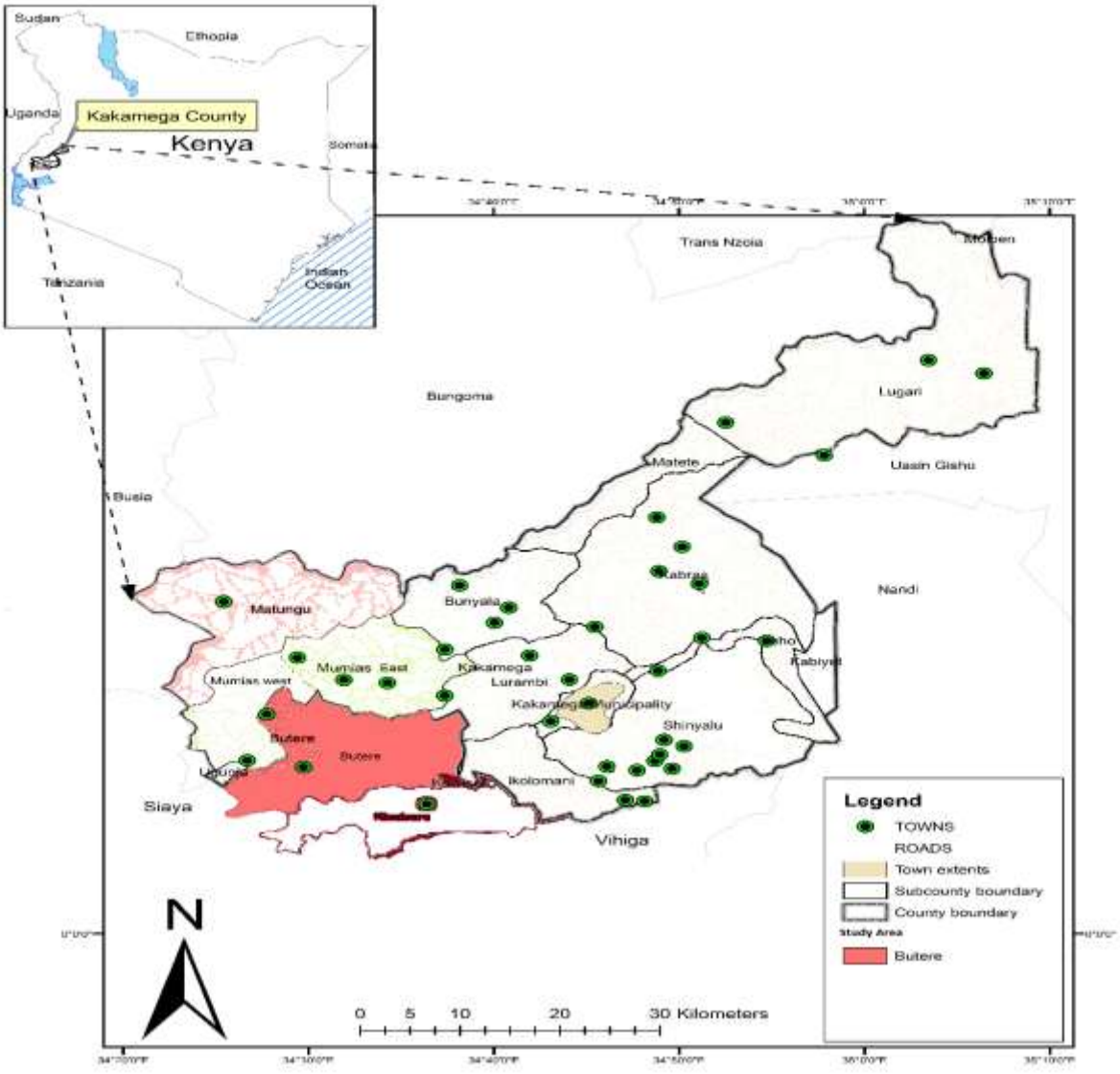

Figure 2: ArcGIS generated map of Kakamega County showing the study area

Source: World Resource Centre (2016)

3.3 Research Design

This study adopted exploratory research design in order to establish statistical conclusions on the factors influencing soybean commercialization, the choice of soybean market outlets as well as the effect of soybean commercialization on smallholder household income in Butere Sub-County in Kakamega county. This research design was appropriate since the study aims at exploring the aspects of soybean commercialization that has not been done by many researchers in the past.

3.4 Sample Size Determination

The sample size that adequately represents the population of smallholder farmers in Butere sub-county was determined using the Cochran's sample size determination formula because the population was unknown. The Cochran (1997) sample determination formula is given as:

$$n_0 = \frac{z^2 pq}{e^2} \tag{3.1}$$

where n_0 was the sample size, z was the critical value of the normal curve that cut off the area estimated at 1.96, e was the desired level of precision set at 6.9%, p was the estimated proportion of attributes present in the population and $q = 1 - p$. According to Cochran (1997), an error of less than 10% is usually acceptable. Therefore, assuming that $p = 0.5$, therefore, $q = 1 - 0.5 = 0.5$; $e = 0.069$ and $z = 1.96$, the sample size will be determined as follows:

$$n_0 = \frac{1.96^2 \times 0.5 \times 0.5}{0.069^2} = 201 \tag{3.2}$$

This will only represent soybean producers as sample unit in Butere Sub-County. The justification for the desired level of precision e set at 6.9%, is because all the tests of significance will be done a t 10% and below.

3.5 Sampling Design and Procedure

The target population of this study was small-scale farmers in Butere Sub-County in Kakamega County. Sampling unit was the smallholder soybean farmers in Butere Sub-County. The study used multi-stage sampling procedure to arrive at the required sample size of 201 soybean producers. The first step involved a purposive selection of Kakamega County because it is one of leading soybean producing counties in western region. The second stage involved purposive selection of Butere Sub-county because it is one of the leading soybean producing sub-counties in Kakamega County where commercialization efforts has remained low. Finally, 201 soybean producers were randomly sampled from two purposively selected wards in Butere Sub-County namely West Marama and Musanda. The selected sampled consisted of 114 randomly selected soybean households from West Marama because it represents a ward where most farmers are engaging in commercialization of soybean. Then, 87 from Musanda ward with least soybean commercialization activities. Random selection was done by counting and visiting every 5th household from the starting point.

3.6 Data Collection

This study used semi-structured questionnaires for interviews with the selected respondents to collect primary data (see appendix 1). The questionnaire was first pretested before the actual data collection using 15 farmers in sample households. The pretesting was done in Khwisero ward, in Butere Sub-county. This was done to enable the correction of mistakes, thus improving the quality, accuracy and reliability of the data collected. The information collected consisted of soybean production and commercialization activity in the year 2018. All respondents were household heads who were participating in soybean marketing in the season of the year. The questionnaires were administered by a group of trained enumerators through personal interviews. Informed consent were sought from all the participants for voluntary participation, and all respondent visited voluntarily agreed to be interviewed. All data that were obtained from the questionnaires were entered and analysed using the STATA and Statistical Package for Social Sciences (SPSS) computer software.

3.7 Analytical Framework

3.7.1 Determining factors influencing soybean commercialization among smallholder farmers in Butere Sub-County in Kakamega county

Policies that encourage market participation in the agricultural sector provide avenues for improving farmers' incomes and welfare. Therefore, generating knowledge about factors in fluencing soybean market participation would be an important input to formulating and implementing interventions that encourage soybean commercialization. Thus, this objective sought to establish the influence of different factors on smallholder farmers' decision to engage in soybean output market. In addition, the focus of this objective was to discern the drivers of extent of smallholder farmer participation in the soybean output market.

Previous empirical studies on market participation have characterized farmer decision to participate in output market as a two-step decision-making process. The first step is conceived as involving farmers' decision on whether or not to participate in the market. In the second step, farmers' who choose to participate in the market make the decision on the volume of the commodity to sell. The empirical estimation of the two-step decisions usually involves fitting a double hurdle model. Most of the empirical studies (Camara, 2017; Olwande *et al*., 2015;

Woldeyohanes *et al*., 2016) have applied double hurdle model to separate farmers who do not participate in the market from those who participate in the market in the first step. The first hurdle involves estimation of a probit model. The second step involves estimation of a truncated or censored tobit regression for the quantities sold in the market. The double hurdle is appropriate model for estimating two-step models when the targeted group of farmers who are all producers (Burke *et al*., 2015). Therefore, following Burke *et al*. (2015) approach, this study adopted double hurdle model in estimating factors influencing soybean commercialization among smallholder farmers in Butere Sub-County in Kakamega County. This is because the study targets only a group of farmers who are soybean producers. Non-producers of soybean were not included in the study.

The double hurdle model can be part of the triple hurdle model. But for the case of this study, the production decision by farmers was not considered (Burke *et al*., 2015). Therefore, the decision to participate in soybean markets was regarded as the initial condition for commercialization. Burke *et al*. (2015) noted that although production is an initial decision, it may be driven by a completely different structural process compared to the decision to participate in the market and the intensity of participation. Thus, this allows decision to participate in the market and the intensity of participation to be modelled using a two-step approach. Here, the first step examined the factors associated with whether or not to sell soybean in the market. The last step estimates the extent of market participation or commercialization measured using Household Commercialization Index (HCI) as follows;

$$HCI = \left(\frac{Total\ Value/Quanity\ of\ Soybean\ Sold}{Total\ Soybean\ production\ Value}\right) \quad (3.3)$$

The first step of Double Hurdle Model involved estimating a probit model. Following Tabe-Ojong*et al*. (2018), the decision whether or not to participate in soybean output market was estimated as:

$$p_i^* = \delta + \varphi X_i + \varepsilon_i \quad (3.4)$$

$$p_i = \begin{cases} 1 & \text{if } p_i^{p*} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Where p_i^* was underlying latent variable representing changes in net benefit or utility for participating in soybean output markets. $p_i = 1$was if a farmer participated in soybean output

market, and 0 otherwise (non-participation). φ was the vector of parameters to be estimated. X_1 was the vector of explanatory variables and ε_1 was independent identically distributed error terms.

In the last step, a truncated tobit regression was fitted to estimate the extent of market participation or commercialization as follows;

$$c = \beta_0 + \beta X_i + \mu_i \quad (3.5)$$

Equation 3.4 represented the probability of smallholder farmer's participation in the soybean output market which is a binary choice of whether to participate in the market or not. The probability of participation in the output market takes the value of 1 if the farmer participates or zero otherwise. c in equation 3.5 represented the extent of market participation (the quantity of soybean sold) measured in terms of Household commercialization index, which is the ratio of quantity sold to the total quantity produced by households.

Turning to other terms in Double Hurdle Model in equation 3.5, X represents a vector of explanatory variables, β represents parameters to be estimated and associated with the explanatory variables. δ, and β_0 are intercepts for equations 3.4, and 3.5, respectively. Lastly, ε_i and μ_i are stochastic disturbance terms.

The study expected that smallholder farmer participation in soybean marketing was enhanced or constrained by a number of factors. Infrastructural factors such as distance to seed dealer and output market were expected to constrain smallholder farmer participation in soybean marketing as well as extent of commercialization. Farm characteristics (land size, type of ownership, soil fertility) were also expected to have either a positive or a negative effect on farmer participation in output market.

Similarly, individual characteristics (age, farming experience, and education), household characteristics (household size, the gender of household head, savings and income) and institutional factors (credit, extension services, social capital, and network) were also expected to either constrain or encourage market participation. Equations 3.6 and 3.7 represent empirical

models for the first objective. Table 1 provides the description and expected signs of variables used in the bivariate probit model.

$$PIM = \beta_0 + \beta_1 Hsz + \beta_2 Gdm + \beta_3 Age + \beta_4 Educ + \beta_5 Exp + \beta_6 Incm + \beta_7 Land + \beta_8 Fert + \beta_9 Ext + \beta_{10} occ + \beta_{11} marr + \beta_{12} Cred + \beta_{13} Tanure + \beta_{14} Grm + \beta_{15} Mkti + \varepsilon_i \quad (3.6)$$

$$HCI = \beta_0 + \beta_1 Hsz + \beta_2 Gdm + \beta_3 Age + \beta_4 Educ + \beta_5 Exp + \beta_6 Incm + \beta_7 Land + \beta_8 Fert + \beta_9 Ext + \beta_{10} occ + \beta_{11} marr + \beta_{12} Cred + \beta_{13} Tanure + \beta_{14} Grm + \beta_{15} Mkti + \varepsilon_i \quad (3.7)$$

Table 1: Description of variables used in Double Hurdle Model

Variable	Description	Measurement	Expected sign	
Dependent Variable				
PIM	Participation in soybean marketing/commercialization (1=Yes, 0 otherwise)	Binary		
HCI	Household commercialization Index	Continuous		
Independent Variables				
Hsz	Household size	Continuous	±	
Gdm	Gender of the decision maker (1=Male, 0= Female)	Binary	±	
Age	Age of household head (Year)	Continuous	±	
Marr	Marital status of household head (1=Married, 0=Otherwise)	Binary		±
Educ	The education level of the decision maker (Years)	Continuous	±	
Exp	Soybean farming experience of the decision maker (Years)	Continuous		±
Incm	Natural log of other Household income (KES)	Continuous		±
Land	Total land size under soybean production (Acres)	Continuous		±
Fert	Soil fertility (1= Fertile 0= Not fertile)	Binary		±
Ext	Extension visitss(Number)	Continuous		±
Occ	Primary Occupation of head (1=Farming, 2 =Otherwise)	Binary		±
Cred	CreditAccess(1=Yes, 0=No)	Binary		±
Tanure	Land Tenure (1=Owned with title, 2= Owned without title)	Binary		±
Grm	Group membership (1=Yes, 0=No)	Binary		±
Mkti	Distance to nearest market (Walking minutes)	Continuous		±

3.7.2 Determining factors influencing the choice of soybean market outlets among smallholder farmers in Butere Sub-County in Kakamega county

Farmers decision to participate in a given market outlet is a choice decision and, therefore, depends on the alternatives at the disposal of the farmer. The study anticipated that farmers have more than two market outlets to choose from as when they decide to commercialize soybean production. The choice of market outlet depends on the decision to commercialize i.e when a farmer decide to be commercialized he or she has to decide on which market outlets to sell to. This implied that the standard binary econometric models are inapplicable in analyzing factors associated with farmer choice of market outlets. Thus, multinomial logistic regression model was used to analyze the factors that influence a farmer's choice of the market outlet. The MNL model allows modeling of response variables with more than two nominal choices (Greene, 2018; Gujarati, 2005). The multinomial logistic model not only focus on the farmers' choice to market soybean output or not but also farmer choice of market outlet. This study expected that farmers in the study area choose a market outlet out of the possible four market outlet. The market outlets would include direct (individual or institutional consumers), cooperatives, middlemen, or agro-processing companies.

This study classified farmers into five categories. The first category consisted of nonparticipants in soybean output market. The second and third categories comprised of farmers who sell directly to consumers and middlemen, respectively. The fourth category of farmers was farmers selling to cooperatives while the last category comprised of farmers selling to agro-processors. Given the marketing alternatives, the probability of farmers' choosing not to sell or sell to one of the four market channel conditioned on an explanatory variable was specified using the MNL model as follows:

$$P_{ij} = \frac{\exp(\beta_j x_i)}{1 + \sum_{j=1}^{5} \exp(\beta_j x_i)} \qquad \text{for j} = 0, 1, 2, 3, 4 \tag{3.8}$$

where x_i represents a vector of covariates that influence i^{th} farmer's choice of market outlet, β_j represents a parameter that corresponds to explanatory variables influencing farmer's choice of an alternative j .

The probability of the farmer not participating in the market was estimated as:

$$P_i|(j=0|x_i) = \frac{1}{1+\sum_{j=1}^{5}\exp(\beta_j x_i)} \quad (3.9)$$

and the probabilities of participating in either outlet 1, 2, 3 or 4 were estimated as:

$$P_i|(j=m|x_i) = \frac{\exp(\beta_j X_i)}{1+\sum_{j=1}^{5}\exp(\beta_j x_i)} \qquad \text{for m} > 0 \quad (3.10)$$

Equations 3.10 was simplified as:

$$P_i = \ln(p_i|1-p_i) = \alpha + \beta_i X_1 + \ldots \beta_n X_n + \varepsilon_i \quad (3.11)$$

where $\ln(p_i|1-p_i)$ = logit of different choice of market participation; P_i = nonmarket participation and $1-P_i$ = market participation either direct, middlemen, cooperative, or agro-processor. The term ε_i is an error term.

Equation 3.9 provides a model for market outlets that a farmer would choose. The five possibilities were set as the dependent variable, with nonmarket participation being the base category, taking the value of zero (0). Direct or consumer marketing was given a value of 1 while middlemen, cooperative and agro-processor were assigned values 1, 2, 3 and 4 respectively.

Table 2 provides the description of variables used in the MNL model and their prior expectations. Individual, as well as household characteristics such as age, gender, education, and household size, were expected to either have a positive or negative influence on farmer choice of market outlet. Institutional, social capital and network factors such as access to market information, group membership and the number of traders known to the farmers were expected to have a positive association with the choice of market outlet. Farm characteristic (quantity of output) and infrastructural factors (distance to the market and type of road) were expected to have a negative or positive effect on the choice of the marketoutlet. Equation 3.12 specifies an empirical model for the association of the dependent variable with a set of explanatory variables.

$$MKO = \beta_0 + \beta_1 Hsz + \beta_2 Gdm + \beta_3 Age + \beta_4 Educ + \beta_5 Exp + \beta_6 Qnty + \beta_7 Ext + \beta_8 Occ + \beta_9 Crd + \beta_{10} Dist + \beta_{11} Grm + \beta_{12} marr + \beta_{13} Inc + \beta_{14} \mathrm{Price} + \varepsilon_i \quad (3.12)$$

Table 2: Description of variables used in the Multinomial Logistic Regression Model

Variable	Description	Measurement	Expected sign
Dependent Variable			
MKO	Market outlet (0 = nonmarket participation, 1=Direct/consumers 2=Middlemen 3=Cooperative 4=Agro-processor)	Discrete	
Independent Variables			
Hsz	Household size (Numbers)	Continuous	±
Gdm	Gender of the household head (1=Male, 0 = Female)	Binary	±
Age	Age of household head (Years)	Continuous	±
Educ	Education level of the household head (Years)	Continuous	±
Exp	Soybean farming experience of the household head (Years)	Continuous	±
Qnty	Total soybean yield produced (Kgs)	Continuous	±
Ext	Extension visits (Number)	Continuous	±
Occ	Primary occupation of head (1=Farming, 0=Otherwise)	Continuous	±
Cred	Credit access(1=Yes, 0=No)	Binary	±
Dist	Distance to the nearest market (walking minutes)	Continuous	±
Grm	Group membership (1=Yes, 0 =No)	Binary	±
Marr	Marital status of head (1=Married, 0 =Otherwise)	Binary	±
Inc	Natural log of other annual income	Continuous	±
Price	Soybean price per unit (KES)	Continuous	±

3.7.3 Determining the effect of soybean commercialization on smallholder household income in Butere Sub-County in Kakamega County

The effect of farmer participation in soybean output market was ordinarily estimated using control function methods. Control function methods recognize that farmer decision to participate in soybean marketing, for instance, is non-random. Hence, control function methods account for endogeneity problem by allowing for self-selection into a decision. However, control function methods make strong functional form assumptions that are important for securing identification. For instance, the Propensity Score Matching (PSM) is popular for the estimation of causal effects. However, the PSM's strong unconfoundedness assumption implies that once observable characteristics are accounted for, farmer decision to participate in soybean market is random and uncorrelated with the outcome of interest (Abdulai & Huffman, 2014; Mmbando, 2014). Such an assumption may be untrue because of possible systematic differences in outcomes of soybean marketing and non-marketing farmers once the selection is based on unmeasured characteristics (Smith & Todd, 2005).

Farmers endogenously choose to either participate or not to participate in soybean market. Hence, it is likely that participation or nonparticipation in soybean market was influenced by unobserved farmer characteristics. It was crucial to note that the unobserved characteristics may be correlated with the outcome variable. Hence, this study applied endogenous switching regression (ESR) to estimate the effect of farmer decision on participation in soybean markets on income measured in terms of soybean gross margin as follows;

$$GM = TR - TVC \tag{3.13}$$

The ESR was estimated in two stages. The first stage involved estimation of the probit model which specified factors that influence farmer decision to participate in soybean output market. The second stage involved estimation of the effect of participation in soybean market on the outcome variables. This implied that there are two outcome equations for the decisions to participate and not to participate in soybean output markets. The two regression equations were specified as:

Decision to participate in soybean output market and its effect on household income:

$$\text{Regime 1 (Commercialized)}: \quad y_{1i} = x_{1i}\beta_1 + \varepsilon_{1i} \text{ if } P = 1 \tag{3.14}$$

$$\text{Regime 2 (Non-commercialized)}: \; y_{2i} = x_{2i}\beta_1 + \varepsilon_{2i} \text{ if } P = 0 \tag{3.15}$$

where, y_{1i} represents decisions to participate in soybean market, y_{2i} represents outcome variables for soybean marketers and non-marketer, respectively, x_{1i} and x_{2i} represents vectors of exogenous covariates with associated vector parameters β_1 and β_2 respectively, and ε_{1i} ε_{2i} are stochastic disturbance terms.

The *prior* expectation is that socio-economic characteristics either have a positive or negative effect on household income. Participation in soybean market was expected to be positively associated with household income since the farmer's decision to participate in marketing activity is assumed to be driven by the maximization of returns on investment. Institutional factors such as access to credit and information were expected to be positively and negatively associated with household income. Also, infrastructural and farm characteristics were also expected to be positively or negatively associated with household income. Equations 3.16 and 3.17 specifies empirical model for objective three and Table 3 provides the description of variables that will be used in the ESR model and their *priori* signs.

$$INCOME\_1 = \beta_0 + \beta_1 Hsz + \beta_2 Gdm + \beta_3 Age + \beta_4 Educ + \beta_5 Ofinc + \beta_6 Land + \beta_7 Exp + \beta_8 Dist + \beta_9 Grm + \beta_{10} ext + \beta_{11} Comer + \varepsilon_i \quad (3.16)$$

$$INCOME\_2 = \beta_0 + \beta_1 Hsz + \beta_2 Gdm + \beta_3 Age + \beta_4 Educ + \beta_5 Ofinc + \beta_6 Land + \beta_7 Exp + \beta_8 Dist + \beta_9 Grm + \beta_{10} ext + \beta_{11} Comer + \varepsilon_i \quad (3.17)$$

Table 3: Description of variables used in Endogenous Switching Regression Model

Variable	Description	Measurement	Expected sign
Dependent Variable			
INCOME_2	Natural log of annual household income per capita for soybean commercialized households	Continuous	
INCOME_2	Natural log of annual Household income per capita for non-commercialized soybean households	Continuous	
Independent Variables			
Hsz	Household size	Continuous	±
Comer	Participation in soybean marketing (1=Yes, 0 otherwise)	Binary	±
Gdm	Gender of the decision maker (1=Male, 0= Female)	Binary	±
Age	Age of household head (Years)	Continuous	±
Educ	Education level of the decision maker (Years)	Continuous	±
Ofinc	Natural log of other off-farm income	Continuous	±
Land	Total land size owned (Acres)	Continuous	±
Exp	Experience in soybeanYears)	Continuous	±
Dist	Distance to the nearest market (walking minutes)	Continuous	±
Ext	Extension Visit (Number)	Continuous	±
Grm	Group membership (1=Yes, 0= No)	Binary	±

CHAPTER FOUR
RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents the findings from analyzed data collected from smallholder soybean farmers in Butere Sub-County, Kenya. The chapter is subdivided into four sections mainly according to the objectives of the study. The first section discusses the descriptive statistics comprising of socio-economic, farm, market and institutional characteristics of soybean farmers in Butere sub-county. In the second section of the chapter, the empirical results of the double hurdle model on factors influencing soybean market participation and extent of commercialization among smallholder farmers in Butere Sub-County are discussed. Inferential statistics based on the analysis of multinomial logistic regression model on factors influencing the choice of main soybean market outlet among smallholder farmers in Butere Sub-County are discussed in the third section. The last section discusses the empirical results of endogenous switching regression model on the effect of soybean commercialization on smallholder household income in Butere Sub-County.

4.2 Descriptive Statistics

The selected sample consisted of 114 (56.72 %) commercialized soybean farmers/households i.e. those households who sold at least some of soybean output they had produced on their farms, and 87 (43.28 %) non-commercialized soybean farmers (Figure 3).

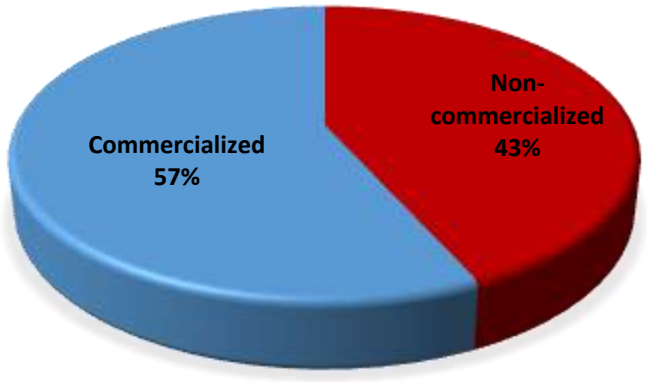

Figure 3: Distribution of sampled household based on soybean commercialization decision

Various groups of variables, specifically socio-economic, farm, market, as well as institutional characteristics of soybean farmers in Butere Sub-County, were analyzed. The socio-economic variables comprised of gender, household size, marital status, age, the main occupation of household head, level of education, experience in soybean production, the total cost of input used per acre, the total cost of labor used per acre, off-farm income, and remittance. Farm variables discussed comprised of total farm size, the fertility of land, land size allocated to soybean, land tenure and total soybean yield. Market variables comprised of distance to the nearest market and household commercialization index. Under institutional variables, the results for access to extension services, number of extension visits, access to credit, and household head membership in an agricultural group are discussed. These descriptive results are presented Table 4.

In terms of the gender of the household head, a high proportion of the soybean farmers who participated in soybean commercialization were male farmers (61.60%), and 70.11% of soybean farmers who did not commercialize were also male farmers. In other words, a higher percentage of soybean farmers interviewed were male farmers regardless of whether they commercialized soybean production or not. This implies that male farmers engaged in soybean production in Butere Sub-County than female farmers, and this is attributed to higher property ownership and access to land among male farmers. Among those households who commercialized and those who didn't commercialize soybean, 38.60%, and 29.89% were headed by female farmers, respectively. However, there was no significant relationship between soybean commercialization and gender of the household head.

Concerning the marital status of household head, 71.93 % of the soybean farmers who participated in soybean market were married while a significant number of farmers (81.39%) who didn’t commercialize soybean were also married. A higher percentage of soybean farmers interviewed were married regardless of whether they participated in the soybean market or not. However, a chi-square test shows an insignificant association between household marital status and soybean commercialization in Butere Sub-County in Kakamega county, Kenya.

Table 4: Descriptive characteristics of soybean farmers in Butere Sub-County

			Soybean Commercialization		
Variables	**Description**	**Overall (n=201)**	**Commercialized (n=114)**	**Non-commercialized (n=87)**	***Chi*^{2} value / t-value**
		Mean/ Percent	**Mean/ Percent**	**Mean/ Percent**	
Socio-economic					
Gender	Male	65.17	61.60	70.11	1.65
	Female	34.83	38.60	29.89	
Marital status	Married	76.12	71.93	81.39	2.54
	Not married	23.88	28.07	18.39	
Main occupation	Farming	77.61	80.70	73.56	3.74
	Businesses	8.96	8.77	9.20	
	Casual laborer	2.99	2.63	3.45	
	Salaried jobs	9.45	7.89	11.49	
	Others	1.00	0	2.30	
Age	Years	54.04 (0.85)	50.25 (1.11)	59.05 (1.12)	5.47***
Education Level	Years	8.89 (0.27)	10.39 (0.28)	6.92 (0.40)	-7.22***
Household size	Number	5.81 (0.19)	4.72 (0.17)	7.22 (0.33)	7.19***
Experience in soybean	Years	6.30 (0.43)	7.12 (0.62)	5.23 (0.52)	-2.22**
Cost of input per acre	KES	4300.84 (977.19)	4171.32 (627.86)	4470.55 (2109.80)	0.15
Cost of labor per acre	KES	19303.31 (3708.48)	13174.33 (627.86)	27334.39 (8404.78)	1.90*

Annual income	KES	46383.23 (4046.23)	50622.89 (5266.23)	40827.82 (6291.67)	-1.20
Remittances	KES	4329.85 (780.12)	5908.77 (1213.12)	2260.92 (804.12)	-2.34**
Farm					
Land fertility	High	23.38	28.75	19.30	5.16*
	Medium	68.16	19.30	59.77	
	Low	8.46	74.56	11.49	
Land tenure	Owned	94.53	6.14	93.10	2.66
	Leased	4.48	95.61	4.60	
	Communal	1.00	4.39	2.30	
Total land size	Acres	1.66 (0.10)	1.87 (0.15)	1.66 (0.16)	
Total land for soybean	Acres	0.31 (0.02)	0.37 (0.03)	0.23 (0.02)	-3.00***
Soybean yield per acre	Kg	213.92 (16.71)	259.87 (20.18)	153.73(26.95)	-3.22***
Institutional					
Extension access	Yes	62.19	76.32	56.32	22.35***
	No	37.81	23.68	43.68	
Credit Access	Yes	45.77	42.98	49.43	0.83
	No	54.23	57.02	50.57	
Group membership	Yes	61.69	66.30	57.80	1.53
	No	38.31	33.70	42.20	
Extension Visit	Number	4 (0.35)	5 (0.54)	2 (0.30)	-4.82***

Market					
Distance to nearest market	Walking minutes	24.30 (2.24)	16.35 (1.11)	34.71 (2.11)	8.20***
Household Commercialization Index	Units	0.38 (0.03)	0.68 (0.02)	0 (0)	-33.31***

Note: *, ** and *** is significant at 10%, 5% and 1% level, respectively. Standard errors are in parenthesis.

Table 4 also presents the results of a cross-tabulation of soybean commercialization and primary occupation of the household head. For the soybean farmers who commercialized their production, the higher percentage (80.70%) were engaging in food and livestock farming as their primary occupation. The remaining soybean farmers who commercialized their production were engaged in small businesses (8.77%), casual labor (2.63%) and salaried employment (7.89%) as their primary occupation. On the other hand, a large proportion (73.56%) of soybean farmers who did not commercialize their production they were also doing food and livestock farming as their primary occupation. The remaining soybean farmers who did not commercialize their production were engaged in small businesses (9.20%), casual labor (3.45%), salaried employment (11.49%), and other activities (2.3%) as their primary occupation. This shows that the majority of soybean farmers were practicing farming as their primary occupation compared to other occupations regardless of whether they participated in the soybean market or not. However, a chi-square test shows an insignificant association between primary occupation and soybean commercialization in Butere Sub-County.

The average age of the commercialized soybean farmers was 50 years while that of non-commercialized soybean farmers was 59 years. The t-test results indicated that there was a significant difference in the mean age of commercialized and non-commercialized soybean farmers at 1% level (Table 4). It shows that non-commercialized soybean farmers were significantly older than commercialized soybean farmers. This could be attributed to the unattractiveness of soybean production and marketing among the older farmers due to high labor requirement for production and marketing of soybean. Young farmers also tend to be less risk-averse and more willing to search and adopt new agricultural production technologies like improved soybean seeds which makes them produce surpluses for the markets compared to older farmers. Muricho (2015) also found that as farmers grow old, there is increased risk aversion and decreased long term investment in the farm.

It is also important to note that the level of education of household heads plays a significant role in enabling smallholder farmers to commercialize soybean production (Muricho, 2015). In this regard, the results in Table 4 revealed that commercialized soybean farmers had on average, significantly (t-value= -7.22, p> 0.000) more years of formal education (10 years) compared to

non-commercialized soybean farmers (7 years). This is an indication that farmers with higher levels of education are much more informed, thus able to effectively search and interpret information related to the importance of modern agricultural production and marketing technologies to produce a surplus for the market. Awotide *et al.* (2016) also posited that highly educated farmers are much more informed thus able to try new highly productive technologies to overcome production and marketing constraints.

The results also show that households commercializing soybean had an average number of 5 household members while those who did not had an average of 7 members. Non-commercialized soybean households had on average, significantly (t-value= 7.19, p> 0.000) large household sizes compared to commercialized soybean households. This could be attributed to the fact that large households have more mouths to feed thus require more food which lowers the amount of surplus available for the market. However, an average number of 5 household members among commercialized households implies that these households needs more many for various expenditures like payment of school fees hence might consider selling part of their soybean output for more income. This finding is consistent with Turaa *et al.* (2016) who asserted that larger family size lower marketed surplus than smaller family size. They added that larger family size requires a higher quantity of produce for consumption, and thus lessening the quantity available for sale.

The mean number of years in soybean production was found to be 7 years for those households commercializing soybean and 5 years for non-commercialized households. Households commercializing soybean had significantly more experience in soybean production than non-commercialized household at 5% level. This implies that longer periods in soybean production translate to high accumulated soybean production knowledge and experience obtained from years of experimentation which increases yields as well as a surplus for sale. This concurs with Masoku *et al.* (2010) who found a positive and significant association between maize commercialization and farming experience among smallholder farmers Swaziland.

The average cost of input used in soybean production per acre of land was found to be KES. 4,171 and KES. 4,470 for commercialized soybean households and non-commercialized soybean households, respectively. These results indicated that non-commercialized soybean households

were incurring the high cost of inputs compared to commercialized soybean households. However, the t-test results revealed that there was no statistically significant difference in the average cost of input used in soybean production per acre of land between commercialized soybean households and non-commercialized soybean households.

The average cost of labor used in soybean production per acre of land was found to be KES. 13,174 and KES. 27,334 for commercialized soybean households and non-commercialized soybean households, respectively. These results indicated that non-commercialized soybean households were incurring significantly (t-value= 1.90, p> 0.058) a higher cost of labor compared to commercialized soybean households 5. This could be attributed to the fact that commercialized soybean households tend to be technically and allocative efficient when it comes to the distribution of labor used in soybean production.

The mean value of annual income earned by commercialized soybean households was found to be KES 50,622 while that of non-commercialized soybean households was KES 40,823. These results indicated that commercialized soybean households earned more annual income compared to non-commercialized soybean households. However, the t-test results revealed that there was no statistically significant difference in the average annual income between commercialized soybean households and non-commercialized soybean households. Annual income consisted of income from on-farm employment, off-farm employment, small businesses and other incomes except for soybean income and remittances.

The average annual value of remittances received by commercialized soybean households was KES 5,909 while that of non-commercialized soybean households was KES 2,261. The mean difference in the average annual value of remittances received by the two groups of soybean farmers was found to be significant at 5%. Remittances comprised of the money received by soybean households in the last year from children, friends, and relatives either living within or outside Kenya. This money can be used to search and adopt required modern agricultural production technologies, inputs, and fertilizer for soybean farming to increase surpluses for the market. Also, remittances can be used to purchase other household food and non-food requirements which reduces household demand for soybean as well as increasing surpluses.

Therefore, commercialized soybean households who receive more remittances have a higher likelihood of participating in soybean commercialization. Again, commercialized farmers also seek these funds to invest in their farms since they have the capacity to repay it back. Coung (2009) found that remittances increase other food expenditures thus creating a surplus for markets.

Type of soil fertility plays an important role in enhancing the quantity of soybean produced. Descriptive statistics in Table 4 showed that commercialized soybean households had significantly (chi-square =5.16, p>0.077) more fertile plots than non-commercialized soybean households. Specifically, among commercialized soybean households, the proportion who had high, medium and low fertile plots was 19.30%, 74.56%, and 6.14%, respectively. Among, non-commercialized soybean households, the proportion who had high, medium and low fertile plots were 28.75%, 59.77%, and 11.49%, respectively. These results imply that a high positive association exists between the type of soil fertility and commercialization. High and medium soil fertility helps in enhancing the amount of soybean yields thus increasing the possibility of having surplus for markets. Muricho (2015) also found that there is a high positive correlation between soil fertility and agricultural commercialization probability and intensity.

From Table 4, the majority of those who engaged in soybean commercialization (95.61%) owned their plots, and the remaining 4.39% leased their plots. Similarly, the majority of non-commercialized soybean farmers (93.10%) owned their plots, and the remaining 4.60% and 2.30% had their plots as leased and communal, respectively. However, the chi-square results revealed that the type of land tenure or ownership had an insignificant relationship with soybean commercialization in Butere Sub-County. The amount of productive resources like the size of land owned theoretically plays an important role in agricultural commercialization process (Bellemare & Barrett, 2006). From the results in Table 4, commercialized soybean farmers recorded a bigger average farm size of about 1.87 acres compared to non-commercialized ones who had about 1.66 acres. This means that land is an important factor of agricultural production, farmers owning bigger farms might be allocating more land for production of soybean, thus experiencing higher production levels, and surplus level, hence have a higher likelihood of agricultural commercialization relative to those having smaller farms. However, t-test results revealed that there was no significant difference in the average land size owned between

commercialized soybean households and non-commercialized soybean households in Butere Sub-County.

The average land size allocated for soybean production for the commercialized soybean farmers was 0.37 acres with a standard error of 0.03 while the average land size allocated for soybean production for the non-commercialized soybean farmers was 0.23 acres with a standard deviation of 0.23. Statistically, there was a statistically significant difference related to average land size allocated for soybean production between those who commercialized and those who did not commercialize soybean production at 1% level of significance. This is an indication that commercialized farmers significantly allocated bigger farms size for soybean production compared to non-commercialized farmers. Greater allocation of land for soybean production increases production levels, as well as surplus level, hence increasing the likelihood of agricultural commercialization. Similarly, a study done by Abayneh and Tefera (2013) showed a positive significant correlation between land size allocated for maize and participation in the maize market.

Amount of soybean produced influences the amount available for home consumption as well as a surplus for sale. The results also revealed that commercialized soybean farmers harvested relatively higher soybean yields with an average of 259.87 Kg per acre of land compared to non-commercialized farmers with an average of 153.73 Kg per acre of land. There was a significant difference in the mean soybean yield per acre between those who commercialized and those who did not commercialize soybean production at 1% level. This implies that households who participate in the soybean market produce higher yields than those households who do not participate in the market. Households getting higher soybean yields could be in a better position to commercialize soybean due to higher surpluses than those households with lower yields at their disposal. These findings are also consistent with the findings of a study by Abayneh and Tefera (2013).

Results in Table 4 show that the majority of the surveyed respondents had access to extension services (62.19%), though commercialized soybean farmers had a significantly (Chi-square=22.35, $p>0.0001$) higher level of extension access (76.32%) compared to non-commercialized soybean farmers (56.32%). This could mean that better access to extension service

serves as a source of agricultural production and marketing information which enables soybean farmers to search and uptake new improved agricultural practices that enable them to harvest above their subsistence level and commercialize more intensively. This concurs with Kotchikpa and Wendkouni (2016) who posited that there exists a high positive relationship between extension excess and soybean commercialization. Similarly, majority of the surveyed respondents were members of agricultural groups (61.69%), though commercialized soybean farmers had a higher proportion of group membership (66.30%) compared to non-commercialized soybean farmers (57.80%). However, the chi-square results revealed that group membership had an insignificant relationship with soybean commercialization in Butere Sub-County.

Majority of the surveyed respondents did not have access to credit services (54.23%). A higher proportion of commercialized soybean households did not have access to credit (57.02%) compared to the non-commercialized soybean farmers (50.57%). However, the chi-square results revealed that access to credit had an insignificant relationship with soybean commercialization in Butere Sub-County. Overall, the results in Table 4 indicate that the average number of extension visit was 4 with a standard error of 0.35. The average number of extension visit for commercialized soybean households was 5 while that of non-commercialized soybean households was 2. Statistically, there was a significant difference related to an average number of extension visit between commercialized soybean households and non-commercialized soybean households at 1% level. In other words, commercialized soybean households significantly had a higher number of extension visits compared to non-commercialized soybean households. A high number of extension visits equips farmers with agricultural production and marketing information which enables them to try new improved agricultural practices to produce above their subsistence level and commercialize more intensively. Farmer groups enhance the interactions and information sharing between farmers and extension officers thus increasing their knowledge and skills to produce above their subsistence level (Ong'ayo *et al*., 2017).

The average walking distances to the nearest market center was 24 walking minutes. The results in Table 4 shows that there was a significant difference in the mean walking distances to the nearest market center between the groups (t-value= 8.20, p>0.0000). Distance to the nearest market center was used as a proxy for access to market information as well as access to possible output markets. On average, commercialized soybean farmers live closer to the market centers (16 walking

minutes) compared to non-commercialized soybean farmers (35 walking minutes). These results show that households living nearer to the market centers have better access to information and markets for output such as soybean yield; thus, they are more likely to participate intensively in the markets. This also indicates that farmers who live closer to the market centers have more access to market information, and thus they incur minimal transportation cost related to searching and interpretation of market information, as well as moving their products to the market. In turn, less transportation cost increases farmers' likelihood to commercialize intensively. Randela *et al.* (2008) also reiterated that a negative and significant relationship exists between distance to the nearest market as a proxy for access to market information and cotton commercialization.

Soybean Household Commercialization Index was calculated as; the total amount of soybean sold in the market from own production over the total amount of soybean produced on the farm. The average Household Commercialization Index of soybean in Butere Sub-County was 0.38. Overall, this implies that soybean households in Butere Sub-County were selling on average about 38% of the total value of soybean produced. Therefore, they were consuming more than 62% of the total value of all soybean produced. Low level of soybean commercialization was therefore evident in Butere Sub-County. For the non-commercialized soybean households, the average commercialization index as a measure of the extent of commercialization was 0. For commercialized soybean households, the average commercialization index as a measure of the extent of commercialization was 0.68. This implies that commercialized households were selling on average about 68% of the total value of soybean they produced. Therefore, commercialized soybean households were only consuming less than 32% of the total value of soybean produced. This is a clear indication of high soybean commercialization level among commercialized soybean households.

4.3 Factors Influencing Decision to Commercialize Soybean and Commercialization Intensity among Smallholder Farmers in Butere Sub-County

4.3.1 Preliminary diagnosis of multiple variables used in the Double Hurdle Model

Four tests were conducted to examine the suitability of multiple variables included in the double hurdle regression model. First, the variance inflation factor (VIF) was undertaken to determine whether multicollinearity existed between continuous variables, and the results are presented in Table 5. Multicollinearity is a state of very high inter-correlations among the independent variables. The results in Table 5 showed that VIF values of individual continuous variables range from 1.06 to 1.26 with mean VIF of 1.15. Thus, it was concluded that no collinearity existed between these independent continuous variables because VIF values were below the recommended value of 3.3 (Greene, 2018; Hair *et al.*, 2011; Knock & Lynn, 2012).

Table 5: Multicollinearity diagnosis results of the variance inflation factor (VIF)

Variable	**VIF**	**1/VIF**
Age of household head	1.18	0.844
Years of schooling of the household head	1.26	0.792
Household head experience in soybean	1.06	0.945
Household size	1.10	0.908
Natural logarithm other income	1.19	0.840
Total land under soybean	1.08	0.922
Number of extension visit	1.06	0.944
Distance to the nearest market	1.26	0.792
Mean VIF	1.15	

Second, the pairwise correlation test was conducted to determine whether multicollinearity existed between categorical variables, and the results are presented in Table 6. Similarly, the results confirmed that there was no multicollinearity among the categorical independent variables because the pairwise correlation coefficients were less than the recommended value of 0.75 in all cases (Greene, 2018).

Table 6: Pairwise correlation coefficients for categorical variables

Variable	Gender	Marital status	Primary occupation	Soil fertility	Land tenure	Credit Access	Group membership
Gender	1.0000						
Marital status	0.5703	1.0000					
Primary occupation	0.1421	0.1889	1.0000				
Soil fertility	-0.0721	-0.0025	-0.0083	1.0000			
Land tenure	-0.0537	0.1348	0.0807	-0.0055	1.0000		
Credit Access	0.1056	-0.0944	0.1534	-0.0079	-0.0454	1.0000	
Group membership	0.0040	0.0867	-0.0187	-0.0556	-0.0354	0.0872	1.0000

Third, the results of the Breusch- Pagan test showed that we could not reject the null hypothesis of constant variance or homoscedasticity (p= 0.1077) (see appendix 3). This implied that the model was free from heteroscedasticity problems. Lastly, a white test for heteroscedasticity was also used, and the results are presented in Table 7. However, the white test results indicated the presence of heteroscedasticity since a *chi*-square of 68.59 was significant at 5% level. Therefore, to counter this problem, robust standard errors were used in all the analyses.

Table 7: White test results for heteroscedasticity

Source	***chi*^{2}**	**Df**	**P-values**
Heteroscedasticity	68.59	44	0.0103
Skewness	19.46	8	0.0126
Kurtosis	3.15	1	0.0761
Total	**91.20**	**53**	**0.0009**

4.3.2 Double Hurdle Regression results for factors influencing soybean commercialization decision and intensity of commercialization

Even though the decision to commercialize soybean as well as the intensity of commercialization can be modeled independently, either by using a logit/probit and Tobit models, respectively, such estimations would result in inefficient and biased estimates of their parameters. This is because it ignores the potential correlation between the unobserved error terms of the two decisions; that is

the decision on the volume of soybean to sell is contingent on the initial decision to commercialize soybean. Such problems were addressed by running a double hurdle model with sample selection.

Double hurdle model was then used to determine factors influencing the decision to commercialize soybean (binary) and intensity commercialization (HCI) among smallholder farmers in Butere Sub-County. However, it is vital to first test whether the double hurdle model is preferred over the Tobit specification using the log-likelihood ratio (LR) statistic. In this regard, the appropriateness of the double hurdle model against a Tobit specification was checked using a likelihood ratio test. In this study, the LR test statistic was 146.45, and it was significant at 1% level (Appendix 3). Therefore, this test statistic showed that the double hurdle model was strongly preferred to Tobit specification. This indicates that two separate decision-making stages exist where soybean farmers make independent decisions regarding whether to commercialize or not and the intensity of commercialization. Also, the Tobit model is restrictive therefore unable to make any distinction between the two stages of commercialization decision-making the process.

The results of the maximum likelihood estimation for a double hurdle regression model, using craggit command, for the decision to commercialize soybean and intensity of soybean commercialization, are presented in Table 8. Also, the sigma constant was relatively high (0.179) and statistically significant at 1% level (Table 8). Sigma constant measures the correlation coefficient between the first tier model (Decision to commercialize soybean) and the second tier model (soybean commercialization intensity model). The significant value of sigma constant statistic is a clear indication of strong dependence between the two tiers, thus supporting the appropriateness of the double hurdle model approach over the Tobit specification (Wooldridge, 2010). Again, the log pseudo-likelihood for the fitted model was -22.164 and Wald chi^2 (15) of 74.12, (Prob > χ^2= 0.000), indicating that all parameters are jointly significant and all covariates included in the models explained the decision to commercialize soybean and intensity of soybean commercialization at 1% significance level.

Similar independent variables related to farmers' socioeconomic, institutional, farm and market characteristics were used to specify the estimated models. The model reported a set of variables significantly influencing soybean commercialization decision and intensity of commercialization.

Regarding household characteristics, the results in Table 8 showed that primary occupation of household head was negatively and significantly related to the soybean commercialization decision and intensity of soybean commercialization at 10% and 5% levels, respectively. By implication, those farmers engaging in crop and livestock farming as their primary occupation were more likely to commercialize soybean compared to those engaging in non-farm employment as a primary occupation. This could be attributed to the fact that household participation in off-farm duty as primary occupation often limit the time available for soybean production, and thus discouraging uptake of labour-intensive technologies that would result in higher soybean yield and surplus for markets.

Table 8: Double hurdle with selection estimation results for soybean commercialization decision and intensity of soybean commercialization, n = 201

	Tier 1 Commercialization decision model (Selection equation) 1=Commercialized, 0= Non-commercialized		Tier 2 Commercialization intensity model (HCI) (Outcome equation) Household Commercialization Index	
Variable Description	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
Socio-economic characteristics				
Gender of household head (1= Male, 0 = Female)	-0.164	0.417	0.054	0.085
Marital status of household head (1=Married, 0= Not married)	0.053	0.495	-0.094	0.010
Primary occupation of the head (1= Crop and livestock farming, 2= non-farm employment)	-0.638	0.355*	-0.174	0.075**
Age of household head (Years)	-0.055	0.014***	-0.010	0.003***
Years of schooling of household head (Years)	0.122	0.036***	0.029	0.009***
Household head experience in soybean (Years)	-0.017	0.024	-0.006	0.004
Household size (Number)	-0.277	0.062***	-0.081	0.014***
Natural logarithm of other income	-0.183	0.101*	-0.040	0.026
Farm Characteristics				
Total land under soybean production (Acres)	2.058	0.758***	0.340	0.121***

Soil fertility (1= Fertile, 0= Not fertile)	0.551	0.435	0.142	0.111
Land tenure (1=Owned with title, 0= Owned without title)	-1.289	0.468***	-0.141	0.142
Institutional Characteristics				
Table 8 Continues				
Credit Access (1 = Yes, 0 = No)	0.049	0.276	0.030	0.066
Group membership (1 = Yes, 0 = No)	-0.230	0.284	-0.003	0.068
Extension visit (Number)	0.154	0.043***	0.023	.006***
Market Characteristic				
Distance to nearest market (Walking minutes)	-0.040	0.009***	-0.011	0.002***
Constant	7.365	1.638***	1.85	0.394***
Sigma constant	0.178	0.012***		

Note: Log pseudo likelihood = -22.16402; Wald chi^2 (15): χ^2 = 74.12, Prob > χ^2= 0.000; Number of observation = 201; Likelihood-ratio test (LR) (16) = 146.45, Prob > chi^2 = 0.0000; *, ** and *** denote significant at 10%, 5% and 1% levels, respectively; Dependent variable for Selection model (first) is commercialization decision, 1=Soybean commercialization decision, 0=Non-commercialization decision; Dependent variable for outcome model (second) is Household commercialization index for soybean.

The results also showed that the intensity of soybean commercialization for household primarily engaging in off-farm jobs was significantly lower than that of farmers primarily engaging in crop and livestock farming by 17.4%, at 5% significance level, ceteris paribus. This finding could be attributed to the possibility that farmers working full time on their plots have time to search, interpret and implement agricultural information related to new soybean production technologies, thus able to harvest more surplus for markets, unlike those working off-farms. Wollni *et al*. (2010) also asserted that household participation in off-farm duties limits time available for farming activities and hence discourages adoption of conservational and labor-intensive agriculture technologies thus lowering yield and surplus for sale.

Age of the household head had a negative and significant influence on the decision to commercialize soybean as well as soybean commercialization intensity both at 1% significance level (Table 8). This implies that the older the farmer is, the lesser the likelihood of participating in the soybean market. In other words, holding other factors constant, a younger household head is more likely to be intensively commercialized compared to a similar household that is older. This implies that if the age of the household head increases by one year it reduces the intensity of soybean commercialization by 0.010, *ceteris paribus.* This could be attributed to the unattractiveness of soybean production and marketing among the older farmers due to high labor requirement for production and marketing of soybean. This also implies that younger farmers are less risk-averse and more innovative, and adaptable hence motivated to uptake and continue using new technologies like improved soybean seeds which makes them produce surpluses for the markets compared to older farmers. In related studies, Muricho (2015) and Onyeneke (2017) also found that younger farmers are more innovative and less risk-averse thus are more than willing to uptake and continue using new productive agricultural technologies like improved seed and fertilizer which makes them produce surpluses for the markets than older farmers.

Education of the household head also had a positive and significant effect on the soybean commercialization decision and intensity of soybean commercialization both at 1% significance level (Table 8). This implies that the more educated a household head is, the higher the likelihood of participating in soybean commercialization. A household head with one more year of formal education was about 0.122 more likely to commercialize soybean production compared to a similar

household with one year less of formal education. A one-year increase in years of schooling was likely to increase soybean commercialization intensity by 0.029, *ceteris paribus.* This could be attributed to the fact that education equips farmers with more agricultural information and skills that enable them to make commercialization decision accurately, hence increasing their participation in soybean market and in a more profitable way. Again, educated farmers are well informed and can search, consolidate and interpret extension information related to practicability and gains associated with new production technologies thereby increasing their production and surplus level. The positive effect of education level on commercialization was consistent with findings from other studies by Mottaleb *et al*. (2015) and Omiti *et al.* (2009).

Results showed household size was negatively and significantly related to the soybean commercialization decision and intensity of soybean commercialization both at 1% significance level. This implies that if the household size increases by one member it decreases the probability of participating in soybean commercialization. Also, a unit increase in household size decreases the intensity of soybean commercialization by 0.277, *ceteris paribus.* This implies that households with more members often have more mouths to feed thus associated with higher demand for food compared to households with fewer members. More mouths to feed in terms of larger household size thus require more food from available produce which lowers the amount of surplus available for the market. This finding is consistent with Turaa *et al.* (2016) who asserted that larger family size lower marketed surplus than smaller family size.

The total amount of annual household income from other sources other than soybean production had a negative and significant influence on soybean commercialization decision at 10% level. The results indicated that an increase in the total amount of annual household income from other sources reduces the probability of household soybean commercialization. A negative effect of the amount of income from other sources on commercialization decision could be attributed to the fact farmers invest less of such funds on activities to increase soybean yields and surplus. Also, they spend much income to expand other activities instead of soybean production. Again, an increased amount of total annual income reduces the farmers' incentive to commercialize its soybean production probably because they have alternative income sources. A similar finding was found by Muricho (2015) in his study on the determinants of agricultural commercialization and its

impacts on welfare among smallholder farmers in Kenya. However, the amount of income received from other sources had an insignificant effect on the intensity of soybean commercialization.

In this study, total land under soybean production was found to affect positively and significantly the decision to commercialize soybean and commercialization intensity both at 1% level. By implications, farmers who allocated a large piece of land for soybean production are more likely to participate in commercialization, compared those allocating small land sizes. Again, when all other factors are held constant, a unit increase in the total land under soybean production was found to increase the intensity of soybean commercialization by about 0.340 (Table 8). Greater allocation of land for soybean implies greater access to land, higher production levels, as well as higher surplus level, hence increasing the probability as well as intensity of soybean commercialization. Similarly, a study done by Abayneh and Tefera (2013) showed a positive and significant effect of land size allocated for maize and maize commercialization.

Land tenure had a negative and significant effect on commercialization decision. This implies that those farmers who own land without title deeds are less likely to participate in soybean commercialization compared to farmers who own land with title deeds. Better land tenancy provides long-term security which raises the probability that farmers will engage in long-term investment. Farmers owning land with title deed possess user rights and ownership which enable them to invest in long-term projects. These farmers can also use such user rights as a collateral for accessing credit facilities to increase their production as well as a surplus for markets. Kpadonou *et al.* (2017) presented similar findings that higher level of land ownership and user right security positively influence investments in the long-term agricultural project due to positive effect on-farm productivity, increased returns, and improved market surplus for the future benefit of smallholder farmer and his family. However, the type of land tenure had an insignificant effect on the intensity of soybean commercialization.

The results indicate that the number of extension visit had a positive and significant influence on soybean commercialization decision as well as the intensity of commercialization both at 1% significance level. By implication, the more the number of extension visits a farmer received, the higher the likelihood of soybean commercialization. Again, this implies that as the number of

extension visit increase by one unit, the intensity of commercialization increases by 0.023 when all other factors are held constant. This is because extension officers have many services they offer to soybean farmers in form of advice on proper crop management practices, crop marketing practices as well as the provision of other agricultural inputs, which help in informing farmers, building their capacity, and increasing their knowledge thereby reducing uncertainties associated with their decision-making process. Extension information, skills, and knowledge, therefore, enable farmers to adopt various agricultural technologies, thus ensuring greater yields and market surplus. Ong'ayo *et al*. (2017) also reported a similar finding that a high number of extension visits provides farmers with agricultural production and marketing information which enables them to try new improved agricultural practices to produce above their subsistence level and commercialize more intensively.

Distance to the nearest market center had a negative significant influence on both commercialization decision and intensity of commercialization at 1% level of significance. This implies that an increase in distance to the nearest input and output market reduces the likelihood of participating in soybean commercialization. Again, as the distance to the nearest input and output market increase by one unit, the likelihood of commercialization intensity reduces by 0.011when all other factors are held constant. This inverse relationship implies that, as the distance to the nearest input and output market center increases, there is a high likelihood of an increase in transportation costs, thereby lowering the probability of market participation. Also, households living nearer to the market centers have better access to market and market information, thus they are more likely to participate intensively in the markets. Those living near market centers also incur minimal transportation cost related to searching and interpretation of market information, as well as moving their products to the market thus increasing their likelihood to commercialize intensively. Randela *et al.* (2008) also found a negative and significant relationship between distance to the nearest market and cotton commercialization.

4.4 Factors Influencing the Choice of Soybean Main Market Outlet Among Smallholder Farmers in Butere Sub-County, Kakamega County

4.4.1 Descriptive statistics for variables used in the Multinomial Logistic regression for the choice of soybean main market outlet

The selected sample consisted of 114 (56.72 %) commercialized soybean households. Descriptive statistics of the commercialized soybean households presented in Figure 4 indicate that most of these households (70 of 114 or 61.40%) used consumer/direct marketing or sold their produce directly to consumers as their main outlet. About, 28.95% and 7.89% indicated middlemen/ brokers and cooperatives as their main soybean market outlets, respectively. Only 2 households (1.74%) sold to or reported agro-processors as their main soybean market outlet. Due to lower proportion, agro-processor as a market outlet was dropped to ensure robust econometric analysis and results.

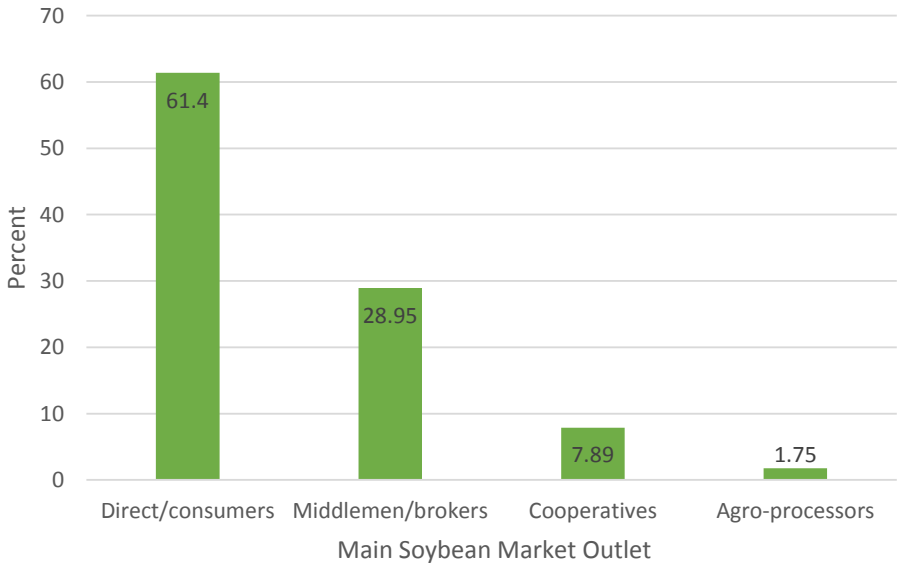

Figure 4: Choice of the main soybean market outlet

Table 9 provides some basic descriptive statistics of sampled commercialized soybean households in the study area, based on their choice of main market outlet. The average age of farmers that sell soybean directly to consumers was 51 years, while that of farmers who supply to middlemen/brokers and cooperatives was 49 years. This implies that these farmers are of middle age hence might have more assets, experience, and ability to reach and negotiate for a better price. Farmers that sell to consumers directly and cooperatives have the same average household size of

5 members compared to those that sell to brokers (4 members). Soybean producers selling to cooperatives have, on average, more years of formal schooling (12 years) than those selling to brokers (11 years) and directly to consumers (10 years). This implies that higher education levels increase participation in more complex market outlets. Farmers using direct market outlet had relatively more soybean farming experience (8 years) than those selling to cooperatives (7 years) and brokers (6 years), implying that they may have more ability to directly negotiate for better prices with consumers.

Descriptive statistics in Table 9 also show that soybean farmers participating in these three market outlets almost have the same amount of income from other sources. The amount of harvested yield also influences the choice or participation in a given market outlet (Alemu *et al.,* 2012). Soybean farmers selling to cooperatives harvested, on average higher soybean yields per acre (122 Kgs) than those selling to brokers (81 kgs per acre) and direct consumers (79 kgs/per acre). This implies that a higher level of production increases the desire for bulk selling in a more formal and complex market outlet.

Soybean farmers selling to cooperatives received on average 9 extension visits compared to those selling their output directly consumers (5 visits) and brokers (4 visits). More extension visits increase farmers' access to information and knowledge about market outlets and their price incentives thus encouraging participation in secure and complex outlets. Distance to the nearest main market is longer for farmers using cooperatives (19 walking minutes) as compared to farmers supplying to consumers directly (16.25 walking minutes) and brokers (16.45 walking minutes). The average selling price per kilogram of soybean was higher for the cooperative outlet (KES 95) than the consumer outlet (KES 84) and brokers (KES 82). Most of the soybean farmers participating in the three outlets had livestock and crop farming as their primary occupation. On average, the majority of soybean household heads participating in the three outlets were male and married farmers. Again, about 64% of soybean commercialized household heads had membership in agricultural groups. Overall, most of the soybean commercialized households (57%) do not have access to credit services. Descriptive results revealed significant differences in farmers' access credit across the market outlets at a 5% level. Soybean farmers selling to cooperatives had a significantly higher access to credit than (78%) than those supplying to brokers (55%) and directly

to consumers (33%). More complex outlets and high-value chains like cooperatives are usually characterized by numerous credit services and incentives to motivate farmers.

Table 9: Descriptive statistics for variables used in the Multinomial Logit Regression for the choice of soybean main market outlet

	Soybean Market Outlets								
Variable	**Direct/Consumers N= 70**		**Middlemen/Brokers N= 33**		**Cooperatives N=9**		**Total N=112**		**F/Chi2 – statistics**
	Mean/ Percent	**Std. dev**	**Mean/ Percent**	**Std. dev**	**Mean/ Percent**	**Std. dev**	**Mean/ Percent**	**Std. dev**	
Age of household head (Years)	51.11	11.40	48.75	13.35	48.78	11.72	50.23	11.98	0.607
Household size (Numbers)	5	1.95	4	1.58	5	1.56	5	1.84	0.186
Education level (Years)	9.94	3.27	10.85	2.39	11.78	3.35	10.36	3.072	0.133
Experience in soybean (Years)	7.62	7.08	5.97	6.30	7.44	6.24	7.13	6.78	0.510
Natural log of other annual income	10.41	1.17	9.92	1.18	10.45	1.34	10.28	1.20	0.135
Total soybean yield (Kg per acre)	78.51	61.58	81.21	81.84	121.78	139.32	82.79	79.40	0.306
Extension visits (Number)	5	6.10	4	3.02	9	9.28	5	5.8	0.048**
Distance to nearest market (Minutes)	16.25	12.34	16.45	10.47	19.67	13.31	16.58	11.83	0.720
Soybean price per unit (KES)	84.06	88.29	82.30	48.84	95.22	48.75	84.43	75.60	0.901
Gender of the household head									
Female	38.57		39.39		33.33		38.19		0.945
Male	61.43		60.61		66.67		61.61		
Marital status of the household head									
Not married	25.71		33.33		33.33		28.57		0.688

Married	74.29	66.67	66.67	71.43	
Primary occupation of the head					
Livestock/ Crop farming	81.43	78.79	77.78	80.36	0.932
Otherwise	18.57	21.21	22.22	19.64	
Credit access					
No	67.14	45.45	22.22	57.14	0.010**
Yes	32.86	54.55	77.78	42.86	
Membership in groups					
No	34.29	42.42	22.22	35.71	0.491
Yes	65.71	57.58	77.78	64.29	

Note:** denote significant at 5% level

4.4.2 Preliminary diagnosis of multiple variables used in the Multinomial Logistic Regression Model

Three tests were done to examine the suitability of multiple variables included in the Multinomial Logit Regression for the choice of soybean main market outlet. First, the pairwise correlation test was conducted to determine whether multicollinearity existed between categorical variables, and the results are presented in Table 10. The results confirmed that there was no correlation among the categorical independent variables because the pairwise correlation coefficients were less than the recommended value of 0.75 in all cases (Gujarati & Porter, 2009).

Table 10: Pairwise correlation coefficients for categorical variables used in Multinomial Logistic Regression

Variables	Gender	Marital Status	Primary Job	Credit Access	Group Membership
Gender	1.000				
Marital Status	0.517	1.000			
Primary Job	0.021	0.114	1.000		
Credit Access	0.201	-0.051	0.162	1.000	
Group Membership	-0.014	-0.018	-0.007	0.118	1.000

Second, the variance inflation factor (VIF) was conducted to determine whether multicollinearity existed between continuous variables included in the MNL model, and the results are presented in Table 11. The results showed that the VIF values of individual continuous variables range from 1.02 to 1.17 with a mean VIF of 1.11. Thus, it was concluded that no collinearity existed between these independent continuous variables because VIF values were below the recommended value of 10 (Greene, 2018; Hair *et al.*, 2011; Knock & Lynn, 2012).

Table 11: Multicollinearity diagnosis results for variables in the multinomial logit model

Variable	Variance inflation factor (VIF)
Age of household head (Years)	1.11
Household size (Numbers)	1.10
Education level (Years)	1.12
Experience in soybean (Years)	1.09
Natural log of other annual income	1.17
Total soybean yield (Kg)	1.02
Extension visits (Number)	1.13
Distance to nearest market (Minutes)	1.17
Soybean price per unit (KES)	1.11
Mean VIF	**1.11**

Third, a white test for heteroscedasticity was also used, and the results are presented in Table 12. The white test results indicated the presence of constant variance or homoscedasticity since a *chi*-square of 47.97 was not significant. However, the results of the Breusch- Pagan test confirmed the need to reject the null hypothesis of constant variance or homoscedasticity (chi2(1) = 6.90; Prob > chi2 = 0.0086) (see appendix 3). This implied that the model was not free from heteroscedasticity problems. Therefore, to counter this problem, robust standard errors were used in all the analyses.

Table 12: White test results for heteroscedasticity

Source	chi^2	Df	P-values
Heteroscedasticity	47.97	54	0.7050
Skewness	22.43	9	0.0076
Kurtosis	0.10	1	0.7540
Total	**70.49**	**64**	**0.2696**

4.4.3 Estimated coefficients of the variables in the MNL model for determinants of the choice of soybean main market outlet

The estimated results from the Multinomial Logistic Regression for soybean market outlet choice are presented in Table 13. The estimated likelihood ratio test (Chi2 (28) = 65.38, Prob>chi2

=0.0001) is statistically significant at 1% level, suggesting that the full Multinomial Logistic Regression adequately fits the data and improves the data fit, and hence outperforms, the alternative model. The results revealed that the estimated coefficients all have the a priori expected signs and show that age of household head, marital status, household size, education level, distance to nearest market, soybean price per unit, extension visits, and credit access all influence main market outlet choice made by soybean growers, but have different effect depending on the market outlet. The age of household head significantly reduces the likelihood that a soybean producer will sell through cooperatives relative to the direct market outlet, at a 10% level. These results suggest that as the age of household head increases, the probability that a farmer will sell to cooperative reduces. Intuitively, older farmers tend to have more wealth as well as transport assets such as vehicles, carts, and motorbikes that can be used to transport their produce to more distant markets or consumers directly instead of selling to cooperatives. Again, older heads of households are characterized by active supply participation and greater contacts and experience with consumers which allow for increased trade opportunities and trust between trading partners to be discovered at lower costs and through repeated exchange with the same party. Shiimi *et al.* (2012) also found that age increases the likelihood of selling products through the informal market channel. Age was insignificant for the broker category.

The marital status of the household head also affected the choice of the soybean market outlet. The coefficients for the marital status of the household head for the cooperative (-2.092) logits were statistically significant at the 10% level. This meant that the marital status of the household head significantly reduced the likelihood that a soybean farmer will sell through cooperatives relative to the direct market outlet. These results suggest that married household heads are less likely to sell through cooperatives relative to direct market outlets than unmarried households. In other words, farmers were less likely to sell through cooperative when they were married. This is because of a joint decision-making process associated with married households on ways of discovering trading opportunities such as direct marketing at a lower cost. However, the effect of household head marital status was insignificant for the broker category.

The coefficient (-0.313) for household size was statistically significant (at 5% level) for the broker category but insignificant for the cooperative category. The significant household size indicates

that an additional member of the household is likely to reduce the probability of selling through brokers compared to the direct market outlet. Household size is used as a proxy for household labour, and considering the labour demand for crop marketing for different outlets, labour can have an impact on the choice of market outlet. Intuitively, larger household size increases the labor force needed to supply and sell products to different consumers either in markets or at their doors rather than selling to brokers at the farm gate. Similarly, Komarek (2010) observed that household size had a significantly positive effect on participation in informal banana market channels.

Table 13: Multinomial Logistic Regression Results for the determinants of the choice of soybean main market outlet among smallholder farmers in Butere Sub-County, Kakamega County, n = 112

	Middlemen/Brokers vs Direct contrast			Cooperatives vs Direct contrast		
Variable	**Estimated Coefficient**	**Robust Std. Err**	**Z- Value**	**Estimated Coefficient**	**Robust Std. Err**	**Z- Value**
Gender of household head (0=Female, 1=Male)	0.138	0.648	0.21	1.260	1.357	0.93
Age of household head (Years)	-0.038	0.028	-1.36	-0.067	0.038	-1.72*
Marital status of the head (0= Not married, 1=Married)	-0.590	0.800	-0.74	-2.092	1.233	-1.70*
Household size (Numbers)	-0.313	0.140	-2.23**	-0.151	0.186	-0.81
Education level (Years)	0.164	0.086	1.90*	0.478	0.252	1.90*
Primary occupation of the head (0= Others, 1= Farming)	-0.154	0.644	-0.24	-0.844	1.783	-0.47
Experience in soybean (Years)	-0.056	0.043	-1.30	0.022	0.057	0.38
Natural log of other annual income	-0.380	0.243	-1.56	0.277	0.530	0.52
Total soybean yield (Kg)	-0.001	0.002	-0.43	0.004	0.005	0.84
Distance to nearest market (Minutes)	0.004	0.022	0.16	0.064	0.034	1.86*
Soybean price per unit (KES)	-0.004	0.003	-1.28	-0.008	0.003	-2.27**
Extension visits (Number)	0.033	0.048	0.68	0.190	0.056	3.39***
Credit access (0= No, 1=Yes)	1.301	0.528	2.47**	2.524	1.191	2.12**
Membership in groups (0= No, 1=Yes)	-0.409	0.628	-0.65	-0.216	0.914	-0.24
Constant	6.113	3.800	1.61	-6.060	8.801	-0.69

Note: Base category, Consumer/Direct market outlet; Number of observation =112; Log pseudo-likelihood = - 75.816; Wald chi^2/ Likelihood ratio test Chi^2 (28) = 65.38, Prob>chi^2 =0.0001; Pseudo R^2=0.2096; ***, ** and * denote significant at 1%, 5% and 10% level, respectively.

Education also influenced the choice of the soybean market outlet. The coefficients for years of schooling for brokers (0.164) and cooperatives (0.478) logits were all statistically significant at 10% level. The education level of household head positively increased the likelihood that soybean producers will sell to brokers and cooperatives relative to consumers directly. This meant that every additional year of education of the household head increases the likelihood of selling soybean produce through brokers and cooperatives as market outlets. Education level helps in enhancing marketing skills as well as mitigating transaction costs by increasing farmers' ability to obtain market information on complex and formal market outlets. Education ensures better skills and access to information thus increasing farmers' understanding of market dynamics, and therefore improve decisions on choice of complex and high-value market outlets such as brokers and cooperatives. Also, educated farmers tend to be risk-averse and good negotiators, thus able to gather and understand production and marketing information so that they can adjust marketing as well as production systems based on the demands of different market outlets. This result is consistent with the findings of Anteneh *et al*. (2011) and Zivenge and Karavina (2012) who revealed that education was positively associated with the choice of formal market channels.

The coefficient (0.064) for distance to the nearest market for the cooperative outlet was positive and statistically significant at 10% level, but insignificant for broker category. These results suggest that distance to the nearest market significantly increased the likelihood that a soybean farmer will sell to cooperatives rather than to consumers directly. Soybean producers who are located far away from market centers experience higher transaction costs, and so may opt to sell to village cooperatives at farm-gate, rather than selling to consumers in more distant market centers that increase transaction costs. Alemu *et al.* (2012) also revealed that crop producers located far from markets in Ethiopia often face higher transaction costs, and so opted to sell through cooperatives, in their villages or nearby villages.

The coefficients for soybean price per unit for the cooperative outlet logit was negative and statistically significant at 5% level, but insignificant for broker logit. Soybean price per unit as a proxy for price knowledge significantly reduced the likelihood that a farmer will sell to cooperatives relative to the direct market outlet. A higher soybean price provides an incentive for the selling outlet. Intuitively, households who perceive cooperative prices to be low and deductions

to be high are more likely to travel to rural markets to sell their products to consumers directly; they are less likely to sell to cooperatives at farm gate, in their village or nearby villages. Martey *et al*. (2012) also found a similarly significant relationship between output price and the choice of market outlet used by smallholder farmers in Ghana. They added that higher output price influenced the choice of rural-informal market outlets relative to urban-formal market outlets.

The coefficient for extension visits was positive and statistically significant at 1% level for the cooperative category logit, but insignificant for brokers category logit. A number of extension visits significantly increased the likelihood that a soybean producer will sell to cooperatives relative to the direct market outlet. An increase in the number of extension visit is expected to equip farmers with better skills and better information relevant to production and dynamic market price which, in turn, increases a producer's ability to choose formal contract market outlets such as cooperatives over informal spot markets. Alemu *et al.* (2012) also found that better access to extension services significantly increased the likelihood that smallholder farmers will choose the contractual cooperative market channels relative to the spot market in Ethiopia. Rahut *et al*. (2015) also added that many contacts with extension providers act as networks for disseminating production and marketing information, and this is expected to increase commercialization levels in complex and contract markets.

Access to credit also had an influence on the choice of soybean main market outlet. The coefficients for access to credit for brokers (1.301) and cooperatives (2.524) logits were all statistically significant at 15% level. Access to credit significantly increased the likelihood that a farmer will sell soybean produce to brokers and cooperatives relative to direct outlets. More complex, formal and high-value contractual chains like brokers and cooperatives are usually characterized by numerous credit services and incentive which bind and motivate farmers to sell in such outlets other than the direct market outlet. Also, better access to credit is expected to increase production, quality as well as marketing levels to meet the output requirements for formal and contract markets like cooperatives.

4.5 Effect of Soybean Commercialization on Smallholder Household Income in Butere Sub-County in Kakamega County

4.5.1 Preliminary diagnosis of multiple variables used in the Endogenous Switching Regression Model

Endogenous Switching Regression Model (ESR) is a two stage model, the first being Probit and second being joint selection and outcome equation estimated using Full Information Maximum Likelihood Method (FIML) (Greene, 2018). The probit model was estimated in the first stage of Endogenous Switching Regression to evaluate the determinants of soybean commercialization in Butere Sub-county, while in the second stage, selection and outcome equations were used to determine the effect of soybean commercialization on smallholder household income in Butere Sub-County. The study used log annual household income per capita as a measure of smallholder household income (Outcome variable). The annual household income was calculated as a summation of all annual incomes received in a household including income from the sale of soybean. This was divided by total household size and converted into natural logarithms to reduce variances and to eliminate the usage of large coefficients (Gujarati & Porter, 2009).

Diagnostic tests such as multicollinearity and heteroscedasticity tests were done to examine the suitability of multiple variables included in the first stage of the endogenous switching regression model that involved an estimation of the selection equation using the probit model. Variance inflation factor (VIF) was conducted to determine whether multicollinearity existed between explanatory variables included in the endogenous switching regression model, and the results are presented in Table 14.

Table 14 : Multicollinearity diagnosis results for variables in the ESR model

Variable	Variance inflation factor (VIF)
Age of household head (Years)	1.24
Sex of household head (1=Male, 0=Female)	1.20
Household size (Numbers)	1.13
Education level (Years)	1.37
Experience in soybean (Years)	1.07
Natural log of other annual income	1.19
Land owned (acres)	1.12
Extension visits (Number)	1.28
Group membership (1=yes, 0=No)	1.17
Distance to nearest market (Minutes)	1.26
Mean VIF	**1.20**

The results showed that the VIF values range from 1.07 to 1.37 with a mean VIF of 1.20. Thus, it was concluded that no collinearity existed between these independent variables because VIF values were below the recommended value of 10, 5 and 3.3 (Greene, 2018; Hair *et al.*, 2011; Knock & Lynn, 2012). Again, pairwise correlation test was conducted to determine whether multicollinearity existed between all the explanatory variables (See Appendix 5). Similarly, the results confirmed that there was no correlation among these independent variables because the pairwise correlation coefficients were less than the recommended value of 0.75 in all cases (Greene, 2018). Heteroscedasticity test was performed using Breuche-Pagan test and the test statistics were $chi^2(1) = 3.78$; Prob > chi2 = 0.1519) (see appendix 5). This implied that the model was free from heteroscedasticity problems. The results of both diagnostic tests indicated that the model was adequate.

The coefficient estimates from the first and second stages of the endogenous switching regression model are presented in Table 15 and 16, respectively. The diagnostic tests showed that the estimated coefficients of the instrument (distance to the market) were insignificant (p> 0.774 and p> 0.368) in both income equations but it was significant (p> 0.005) in the soybean commercialization decision equation (see appendix 4). Since the test statistic was insignificant, the

instrument was valid as it justifies that it does not directly affect household income (Smith & Todd, 2005). The estimated coefficient of correlation between the soybean commercialization equation and the per capita income function (rho_2) was negative and significantly (p> 0.003) different from zero as shown in Table 16. This justifies that both observed and unobserved factors affect soybean commercialization decisions and household income given the commercialization decision (Smith & Todd, 2005). The significance of the coefficient of correlation (rho_2) between the soybean commercialization equation and the per capita income of soybean households indicates that self-selection occurred in the soybean commercialization decision. This allows estimation with an endogenous switching regression model to control for the predicted probability of commercialization decision in order to correct for a possible selection bias associated with unobserved factors that might simultaneously affect the soybean commercialization and outcome decision (Mmbando*et al.*, 2015).

The likelihood ratio test (LR) for the joint independence equation was significant at 5% level ($chi^2(1) = 5.23$, $p> chi^2 = 0.0222$) as shown in Table 15 and 16. This suggests that the equations are jointly dependent hence providing evidence of endogeneity (Mmbando*et al.*, 2015). The probit model was estimated on commercialization decision (selection equation). Table 15 shows a log-likelihood ratio of -239.58 indicating how the model quickly converges. The Wald chi-square statistic (Wald $chi^2(9) = 89.40$, $p = 0.000$) shows that the model wholly and significantly fits the data well, and in that the soybean commercialization decision was attributed to the explanatory variables considered in the probit model. This also shows that the combination of explanatory variables meets the balance requirement.

4.5.2 Determinants of soybean commercialization decision

Table 15 also presents information about some of the factors influencing soybean commercialization, where the explained variable takes the value of one (1) if the farmer commercialized soybean production, and zero (0) if the farmer did not participate in soybean commercialization in Butere Sub-County.

Table 15: Probit model results for determinants of soybean commercialization decision

Variable	Coefficient	Standard Error	Z-value
Sex of household head (1=Male, 0=Female)	-0.098	0.280	-0.35
Age of household head (Years)	-0.044	0.012	-3.67***
Education level (Years)	0.095	0.034	2.85***
Household size (Numbers)	-0.223	0.059	-3.76***
Natural log of other annual income	-0.023	0.106	-0.21
Land owned (acres)	0.070	0.077	0.91
Experience in soybean (Years)	-0.008	0.021	-0.37
Group membership (1=yes, 0=No)	-0.086	0.272	-0.32
Extension visits (Number)	0.128	0.040	3.19***
Distance to nearest market (Minutes)	-0.029	0.108	-.271***
Constant	3.629	1.700	2.13**

Note: Number of observation = 201; Log likelihood =-4-239.583; Wald χ^2 (9) = 89.40, Prob > χ^2 = 0.000; Likelihood ratio test (LR) for the joint independence equation χ^2 (1) = 5.23, p> χ^2 = 0.0222); ***, ** and * denote significant at 1%, 5% and 10% levels, respectively.

The study found that the age of household head had a negative and significant influence on soybean commercialization at 1% level. This implies that older farmers are less likely to undertake soybean commercialization. This could be attributed to the high labor requirement for production and marketing of soybean thus making it unattractive investment among the older farmers compared to younger farmers. Muricho (2015) also found similar results. The education of the household head had a positive and significant effect on the soybean commercialization decision at 1% significance level (Table 15). More educated household heads are more likely to undertake soybean commercialization due to the high level of agricultural information which increases their production as well as surpluses. The positive effect of education level on commercialization was consistent with findings from other studies by Mottaleb *et al*. (2015) & Omiti *et al.* (2009).

Household size had a negative and significant influence on soybean commercialization decisions at 1% level. This shows that the larger the household size, the less likely that a farmer would

engage in soybean commercialization. Large households exert more pressure on household resources thus increasing the demand for food which lowers marketed surplus. Martey *et al*. (2012) and Omiti *et al*. (2009) also found that large household size increases household demand and consumption of food thus lowering marketed surplus. Extension visits had a positive and significant influence on soybean commercialization decisions at 1% significance level. By implication, the more the number of extension visits a farmer received, the higher the likelihood of soybean commercialization. Extension information, skills, and knowledge gained over time enable farmers to adopt various improved agricultural technologies, thus ensuring greater yields and marketed surplus (Ong'ayo *et al.,* 2017). Distance to the market measured in the nearest walking minutes had a negative and significant influence on soybean commercialization. Soybean farmers located farther from the markets are less likely to participate in the soybean market. As distant to markets increases, transportation cost also increase thus limiting active engagement in soybean commercialization (Randela *et al.,* 2008).

4.5.3 Estimates of endogenous switching regression for the effect of soybean commercialization on annual household income per capita

The determinants of annual household income per capita were analyzed in the second stage of the endogenous switching regression model, and results presented in Table 16. The joint (selection and outcome) equation was estimated using full information maximum-likelihood method (FIML). As noted earlier, the covariance coefficients of the two regimes were estimated in order to test for selection bias. The results show that the coefficients (rho_2) for the non-commercialized group was negative (-0.941) and statistically significant ($p<0.003$). This indicates that there exists sample selectivity as well as endogeneity. This also indicates that soybean farmers who commercialized their production differed from those who didn't commercialize and that soybean commercialization may not present the same effects on the non-commercialized farmers supposing they decide to do so. Since rho_2 was negative and significantly different from zero, the model justifies that households who choose to commercialize soybean receive higher annual income per capita than a random household from the sample would have received. In other words, the results confirm that soybean commercialization contributed to improved annual household income per capita. Differences in the annual household income per capita equation coefficients between the farm households who commercialized soybean and those that did not commercialize justify the presence

of sample heterogeneity. The annual household income function for farm households who commercialized soybean production was significantly different from that of farm households that did not participate in soybean commercialization. This implies that soybean households who choose to commercialize their production have more returns which further improves annual household income per capita compared to those who did not commercialize.

Household size had a negative and significant impact on annual income per capita of the farm households that commercialized soybean as well as those that did not commercialize soybean, both at 1% level. The elasticity coefficient shows that an additional member to a farm household increases annual income per capita by 21% for commercialized soybean households as compared to 9 % for non-commercialized soybean households, all factors held constant. Annual income per capita is a function of household size, therefore, larger household size reduces it. An increase in household size increases a household expenditure thus lowering household income. This finding asserts findings of a study by Opondo (2018) who reported that household size had a negative and significant influence on per capita income for the commercialized and non-commercialized cassava households. She presented that household expenditure increases as the household size grows subsequently lowering household income.

Table 16: Endogenous switching regression parameter estimates for the effect of soybean commercialization on log annual household income per capita

	FIML Endogenous switching regression Model			
	Soybean Commercialization Status			
	Commercialization		**Non-commercialization**	
Variables	**Coefficient**	**Standard Error**	**Coefficient**	**Standard Error**
Sex of household head (1=Male, 0=Female)	-0.055	0.140	0.078	0.145
Age of household head (Years)	-0.002	0.006	0.001	0.006
Education level (Years)	-0.002	0.024	-0.023	0.019
Household size (Numbers)	-0.218	0.042***	-0.096	0.024***
Natural log of other annual income	0.395	0.053***	0.857	0.047***
Land owned (acres)	0.055	0.042	-0.007	0.045
Experience in soybean (Years)	-0.010	0.009	-0.003	0.013
Group membership (1=yes, 0=No)	0.292	0.013**	-0.069	0.140
Extension visits (Number)	-0.011	0.014	0.043	0.024*
Constant	6.645	0.720***	0.216	0.674
/lns1	-0.468	0.066***		
/lns2	-0.466	0.095***		
/r1	0.015	0.388		
/r2	-1.749	0.581***		
sigma_1	0.626	0.041***		
sigma_2	0.628	0.060***		
rho_1	0.015	0.389		
rho_2	-0.941	0.066***		

Note: Number of observation = 201; Log likelihood =-4-239.583; Wald χ^2 (9) = 89.40, Prob > χ^2 = 0.000; Likelihood ratio test (LR) for the joint independence equation χ^2 (1) = 5.23, p> χ^2 = 0.0222); ***, ** and * denote significant at 1%, 5% and 10% levels, respectively.

The coefficients of the natural log of other annual income were positive and significant at 1% level for commercialized and non-commercialized groups. When other factors are held constant, the

elasticity coefficient shows that 1% increase in off-farm incomes increases annual income per capita by 39% and 4% for commercialized soybean households and non-commercialized soybean households, respectively. This implied that the effect of off-farm income on annual income per capita for a commercialized household was greater compared to the non-commercialized group. This implies that off-farm activities contribute significantly towards the income of soybean households.

The results also found that the coefficient of group membership was positive and significant at 5% level, for the commercialized group only. For commercialized soybean farmers, participation in groups increases their annual per capita income by 29%, other factors held constant. This implies that participation in groups impacts farm households' engagement in production and commercialization activities. Groups offer market opportunities such as access to output markets and fair prices through collective bargaining and resource pooling. Farmers also acquire new skills and knowledge by interacting freely with other group members thus increasing their production and commercialization activities which in turn increases their household incomes. Magreta *et al*. (2010) and Opondo (2018) also found a similarly positive and significant influence of group participation on household income.

The number of extension visits was found to be positive and significantly influencing annual income per capita for non-commercialized soybean households at 10% level. Farmers also acquire new skills and knowledge from extension visits which improve their production and commercialization activities thereby increasing their household incomes. The coefficient shows that a one-unit increase in extension visits increases annual income per capita by 4% for non-commercialized soybean households. Mmbando *et al.* (2015) also found a similarly positive and significant relationship between number of extension visits and household income.

Table 17 shows the results of expected soybean households' annual income (i.e. log annual household income per capita) under actual and counterfactual conditions.

Table 17: Average expected log annual household income per capita; treatment and heterogeneity effects

Sub-sample	Soybean Commercialization Decision		Treatment Effects
	To commercialize	Not to commercialize	
Commercialized Households	(a) 9.326	(c) 8.149	TT=1.176 (20.476)***
Non-commercialized Households	(d) 8.059	(b) 7.664	TU= 0.395 (4.199)***
Heterogeneity effects	BH1=1.267	BH2=0.485	TH= 0.781

Notes: (a) and (b) represent observed expected log annual household income per capita outcome for commercialized and non-commercialized soybean households, respectively; (c) and (d) represent counterfactual expected log annual household income per capita outcome for commercialized and non-commercialized soybean households, respectively.

TT = (a)-(c)=the effect of the treatment (i.e. soybean commercialization) on the treated (i.e. households that commercialized)

TU = (d)-(b)= the effect of the treatment (i.e. soybean commercialization) on the untreated (i.e. households that did not commercialize)

BH = the effect of base heterogeneity for households that commercialized and did not commercialize

TH = (TT-TU), i.e. transitional heterogeneity

*** denote significant at 1% levels.

The predicted log annual household income per capita from the ESR model was used to determine the mean log annual household income per capita gap between the commercialized and non-commercialized soybean households. The results present that the mean value of annual household income per capita for commercialized soybean farmers is statistically higher than had they not been commercializing soybean. The results also indicate that soybean commercialization increases annual household income per capita by about 117%. For non-commercialized farmers, the mean

annual household income per capita would have been increased by 40% had they commercialize soybean. These results imply that soybean commercialization increased household income measured in terms of annual household income per capita. The transitional heterogeneity effect of annual household income per capita for soybean is positive. This indicates that this income effect is bigger for the commercialized soybean household with respect to one that did not commercialize.

CHAPTER FIVE

CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter presents the summary of this study, conclusions, policy recommendations and areas of further research.

5.2 Conclusions

i. Smallholder farmers who are young, more educated, with few household members, more land under soybean production, frequently visited by extension officers and taking a shorter time walking to the market sell a larger portion of their sellable surplus of soybean in Butere Sub County, Kenya.

ii. Smallholder farmers in Butere Sub County, Kenya who are more educated and have access to credit choose to sell their soybean produce through brokers and cooperative market outlets in contrast to consumers.

iii. Soybean commercialization increases household incomes of smallholder farmers in Butere Sub County, Kenya.

5.4 Policy Recommendations

i. To increase soybean commercialization, extension officers in Butere Sub County should consider increasing the frequency of visits to young farmers and advise them on how to put more land under soy bean production for example by effectively intercropping it with other crops produced in the area.

ii. County government of Kakamega should enhance credit access to smallholder soybean farmers by promoting credit diversion and organizations that provide products which conveniently suit farmers selling their produce through brokers and cooperatives in Butere Sub County, Kenya.

iii. County government of Kakamega through their ministry of gender and social services should motivate farmers to form groups by reducing the length of procedure followed to register a group to encourage more farmers market participation hence improving household income in Butere Sub County, Kenya.

5.5 Areas of Further Research

While this study only aimed to determine factors influencing soybean commercialization, to determine factors influencing the choice of soybean main market outlet, and to evaluate the effect of soybean commercialization on smallholder household income in Butere Sub-County, Kakamega County, it may also be important for future research to evaluate the same in other soybean potential areas. The study also failed to look at the impact of soybean commercialization intensity as well as market outlets on income of smallholder, Therefore, future research should be more comprehensive in modeling the impact of soybean commercialization intensity as well as the effect of soybean market outlets on income of smallholder farmers, using robust non-linear panel data and dynamic panel data models to control for unobserved heterogeneity and endogeneity.

REFERENCES

Abayneh, Y., & Tefera, T. (2013). Factors influencing market participation decision and extent of participation of haricot bean farmers in Meskan District, Ethiopia. *International Journal of Management and Development Studies, 2*(8), 2320-0685.

Abdulai, A., & Huffman, W. (2014). The adoption and impact of soil and water conservation technology: An endogenous switching regression application. *Land Economics, 90*(1), 26-43.

Abtew, A., Niassy, S., Affognon, H., Subramanian, S., Kreiter, S., Garzia, G. T., & Martin, T. (2016). Farmers' Knowledge and Perception of Grain Legume Pests and their Management in the Eastern Province of Kenya. *Crop Protection, 87*, 90-97.

AGRA. (2017). Soya bean farming untapped. Retrieved from https://agra.org/soya-bean-farming-untapped/

Alemu, E., Mathijs, E., Maertens, M. Deckers, J., Gegziabher, K., Bauer, H., & Ghiwot, K. (2012). Vertical coordination in the local food chains: Evidence from farmers in Ethiopia. *International Journal of Economics and Finance Studies, 4*(1), 11-20.

Alene, D., Manyong, M., Omanya, G., Mignouna, D., Bokanga, M., & Odhiambo, G. (2008). Smallholder marketed surplus and input use under transactions costs: maize supply and fertilizer demand in Kenya. *Food Policy, 32*(4), 318-328.

Anteneh, A., Muradian, R., & Ruben, R. (2011). Factors affecting coffee farmers market outlet choice. The Case of Sidama Zone, Ethiopia. Paper presented at the EMNet 2011, 5th International Conference on Economics and Management of Networks, Limassol, Cyprus, December 1-3, 2011.

Asfaw, S., Kassie, M., Simtowe, F., & Lipper, L. (2012). Poverty Reduction Effects of Agricultural Technology Adoption: A Micro-evidence from Rural Tanzania. *Journal of Development Studies, 48*(9), 1288-1305.

Awotide, B., Karimov, A., & Diagne, A. (2016). Agricultural technology adoption, commercialization and smallholder rice farmers' welfare in rural Nigeria. *Agricultural and Food Economics, 4*(3), 1-24.

Barrett, C. (2008). Smallholder market participation: Concepts and evidence from Eastern and Southern Africa. *Food Policy, 33*(4), 299-317.

Bardhan, D., Sharma, M. L., & Saxena, R. (2012). Market Participation Behaviour of Smallholder Dairy Farmers in Uttarakhand: A Disaggregated Analysis. *Agricultural Economics Research Review, 25*(2), 243-254.

Bellemare M. F., & Barrett, C. B. (2006). An ordered Tobit model of market participation: evidence from Kenya and Ethiopia. *American Journal of Agricultural Economics, 88*(2), 324 – 337.

Bezu, S., Kassie, G. T., Shiferaw, B., & Ricker-Gilbert, J. (2014). Impact of Improved Maize Adoption on Welfare of Farm Households in Malawi: A Panel Data Analysis. *World Development, 59*, 120-131.

Brian, D., & Barrett, C. (2014). *Agricultural Factor Markets in Sub-Saharan Africa: An Updated View with Formal Tests for Market Failure*. Policy Research Working Paper (No. 7117). World Bank Group, Washington, DC, World Bank.

Burke, W., Myers, R., & Jayne, T. (2015). A Triple-Hurdle model of production and market participation in Kenya's Dairy Market. *American Journal of Agricultural. Economics, 97*(4), 1227–1246.

Camara, A. (2017). An analysis of welfare effect of market participation of smallholder farm households in Guinea. *The Economic Research Guardian, 7*(1), 2-23.

Carletto, C., Corral, P., & Guelfi, A. (2017). Agricultural commercialization and nutrition revisited: Empirical evidence from three African countries. *Food Policy, 67*(1), 106–118.

Caviglia, H. J. (2003). Sustainable Agricultural Practices in Rondônia, Brazil: Do Local Farmer Organizations Impact Adoption Rates? *Economic Development and Cultural Change, 52*(1), 23-50.

CGIAR, 2016. CGIAR Dryland cereals and legumes agri-food systems: Delivering development and dignity through diversification in drylands. Consultative Group of International Agricultural Research.

Chirwa, E. W. (2009). Determinants of Marketing Channels among Smallholder Maize Farmers in Malawi. Working Paper 3. The University of Malawi.

Cochran, W. G (1997). *Sampling Techniques, 3rd Edition*. Wiley Eastern: John Wiley & Sons, Inc.

Cuong, N. V. (2009). The impact of international and internal remittances on household welfare: evidence from Vietnam. *Asia-Pacific Development Journal, 16*, 59–74.

Daryanto, S., Wang, L., & Jacinthe, P. A. (2015). Global Synthesis of Drought Effects on Food Legume Production. *PloS one, 10*(6), 1-16.

Edwards, Y. D., & Allenby, G. M. (2003). Multivariate Analysis of Multiple Response Data. *Journal of Marketing Research, 40*(3), 321-334.

Egbetokun, A., Shittu, B., & Ayoade, M. (2017). Determinants of market participation among maize farmers in Ogbomoso Zone, Oyo State, Nigeria. *Cercetări Agronomice În Moldova, 1*(169): 109-118.

FAO. (2017). *FAO Statistics*. Retrieved 13 July 2017 from http://www.fao.org/faostat/en/#data/QC

FAOSTAT, 2019. Available at http://www.fao.org/faostat/en/#data/QC. (Accessed 12 January 2019).

Franke, A. C., Van Den Brand, G. J., & Giller, K. E. (2014). Which Farmers Benefit most from Sustainable Intensification? An Ex-ante Impact Assessment of Expanding Grain Legume Production in Malawi. *European Journal of Agronomy, 58*, 28-38.

Garrett, R. D., Lambin, E. F., & Naylor, R. L. (2013). Land Institutions and Supply Chain Configurations as Determinants of Soybean Planted Area and Yields in Brazil. *Land Use Policy, 31*, 385-396.

Gebremedhin, B., & Moti, J. (2012). Interdependence of smallholders' net market positions in Mixed-crop livestock systems of Ethiopian highlands. *Journal of Development and Agricultural Economic, 4*(7), 199-209.

Greene, W. H. (2018). *Econometric Analysis. 8th Edition.* Harlow, England: Pearson Education Limited.

Gujarati, D. (2005). *Basic Econometrics*. New Delhi, India: Tata McGraw-Hill.

Gujarati, D. N., & Porter, D.C (2009). *Basic Econometrics.* 5th edition, McGraw-Hill/Irwin, New York.

Hlongwane, J., Ledwada, L. & Belete, A. (2014). Analyzing the factors affecting the market participation of maize farmers: A case study of small-scale farmers in greater Giyani Local Municipality of the Moponi District, Limpopo Province. *African Journal of Agricultural Research, 9*(10), 895-899.

Houngnandan, P., Charlotte, Z., Dedehouanou, H., & Toukourou, F. (2015). Determinants of soybean [Glycine max (L.) Merrill] production system in Benin. *Journal of Experimental Biology and Agricultural Sciences, 3*(5), 430-439.

ICRISAT. (2013). *Bulletin of Tropical Legumes*. Retrieved 13 July 2017 from http://www.icrisat.org/TropicalLegumesII/pdfs/Flyer_A%20Bulletin%20of%20the%20Tropical%20October.pdf

Jari, B., & Fraser, G. (2009). An analysis of institutional and technical factors influencing agricultural marketing amongst smallholder farmers in the Kat River Valley, Eastern Cape Province, South Africa. *African Journal of Agricultural Research, 4*(11), 1129-1137.

Kadam, P. & Suryawanshi, S. D. (2011). Constraints and Suggestions of Soybean Growers in Adoption of Soybean Production Technology. *International Journal of Agricultural Engineering, 4*(2), 120-124.

Kathage, J., Kassie, M., Shiferaw, B., & Qaim, M. (2016). Big Constraints or Small Returns? Explaining Nonadoption of Hybrid Maize in Tanzania. *Applied Economic Perspectives and Policy, 38*(1), 113-131.

Kenya National Bureau of Statistics (KNBS) (2017). *Third quarter 2017 gross domestic product report.* Nairobi: Government of Kenya Printing Press

Kim, H., & Moschini, G. (2018). The Dynamics of Supply: U.S. Corn and Soybeans in the Biofuel Era. *Land Economics, 94*(4), 593–613.

Kimei, S., Mshenga, P., & Birachi, E. (2017). Factors influencing commercialization of beans among smallholder farmers in Rwanda. *IOSR Journal of Agriculture and Veterinary Science, 10*(8), 30-34.

Komarek, A. (2010). The determinants of the banana market commercialization in Western Uganda. *African Journal of Agricultural Research, 5*(9):775-784.

Kotchikpa, G. L., & Wendkouni, J. B. (2016). Factors influencing smallholder crop commercialization: Evidence from Cote d'Ivoire. *African Journal of Agricultural Research, 11*(41), 4128-4140.

Kpadonou, R., Owiyo, T., Barbier, B., Denton, F., Rutabingwa, F., & Kiema, A. (2017). Advancing climate-smart-agriculture in developing drylands: Joint analysis of the adoption of multiple on-farm soils and water conservation technologies in West Africa Sahel. *Land Use Policy, 61*, 196-207.

Kyaw, N., Soojung, A., & Lee, S. (2018). Analysis of the Factors Influencing Market Participation among Smallholder Rice Farmers in Magway Region, Central Dry Zone of Myanmar. *Sustainability, 10,* doi:10.3390/su10124441

Lapar, M., Holloway, G., & Ehui, S. (2003). Policy options promoting market participation among smallholder livestock producers: A case study from the Philippines. *Food Policy, 28*(3), 187-211.

Lu, M., & Yang, W. (2012). Multivariate Logistic Regression Analysis of Complex Survey Data with Application to BRFSS Data. *Journal of Data Science, 10*(2), 157-173.

Mahasi, J. M., Mukalama, J., Mursoy, R. C., Mbehero, P., & Vanlauwe, B. (2011). A Sustainable Approach to Increased Soybean Production in Western Kenya. *African Crop Science Conference Proceedings, 10,* 111-116.

Magreta, M., Magambo, T., & zingore, S. (2010). When the weak win. The role of farmer groups in influencing Agricultural policy outcome; a case of Nkhate irrigation scheme in Malawi. Poster presented at the Joint 3rd African Association of Agricultural Economists (AAAE) and 48th Agricultural Economists Association of South Africa (AEASA) Conference, Cape Town, South Africa, September 19-23, 2010.

Makhura, M., Kirsten, J., & Delgado, C. (2001). Transaction costs and smallholder articipation in the maize market in the Northern Province of South Africa. *Seventh Eastern and Southern Africa Regional Maize Conference*, 11-15 February, Pretoria, South Africa.

Mausch, K., Woldeyohanes, T., Heckelei, T., & Tabe-Ojong, M. P. (2018, April). A Triple Hurdle Model of the Impacts of Improved Chickpea Adoption on Smallholder Production and Commercialization in Ethiopia. In *92nd Annual Conference, April 16-18, 2018, Warwick University*, Coventry, UK (No. 273473). Agricultural Economics Society.

Masuku, M., Makura, T., & Rwelamira, J. (2001). Factors affecting market decisions in the maize supply chain among smallholder farmers in Swaziland. *Agrekon, 40*(4), 698-707.

Masuku M. B., Makhura M. T., & Rwelarmira J. K., (2010). *Factors affecting marketing decisions in the maize supply chain among smallholders in Swaziland.* Agrekon: Agricultural Economics Research, Policy and Practice: South Africa: Routledge.

Martey, E., Annin, K., Wiredu, A., & Attoh, C. (2012). Does access to market information determine the choice of marketing channel among smallholder yam farmers in the Brong

Ahafo region of Ghana? A Multinomial logit regression analysis. *Journal of Economics and Sustainable Development, 3*(12): 18-28.

Martey, M., Ramatu, M. A.-H., & Kuwornu, J. K. M. (2012). Commercialization of smallholder agriculture in Ghana: A Tobit regression analysis. African Journal of Agricultural Research, 7(14), 2131-2141

Mathenge, M., Place, F., Olwande, J., & Mithoefer, D. (2010). *Participation in Agricultural Markets among the Poor and Marginalized: Analysis of Factors Influencing Participation and Impacts on Income and Poverty in Kenya. A study Report*. Kenya: Tegemeo Institude of Agriculturasl Policy and Development and World Agroforestry Centre. Retrieved 13 December 2017 from https://www.researchgate.net/profile/john_olwande/publication/265151383_participation_in_agricultural_markets_among_the_poor_and_marginalized_analysis_of_factors_influencing_participation_and_impacts_on_income_and_poverty_in_kenya_study_report/links/54b7b29a0cf2bd04be33bcd5/participation-in-agricultural-markets-among-the-poor-and-marginalized-analysis-of-factors-influencing-participation-and-impacts-on-income-and-poverty-in-kenya-study-report.pdf

Mbanya, W. (2011). Assessment of the Constraints in Soybean Production: A Case of Northern Region, Ghana. *Journal of Developments in Sustainable Agriculture, 6*(2), 199-214.

Mbitsemunda, J., & Karangwa, A. (2017). Analysis of factors influencing market participation of smallholder bean farmers in Nyanza District of Southern Province, Rwanda. *Journal of Agricultural Science, 9*(11), 1916-9760.

Mhango, W. G., Snapp, S. S., & Phiri, G. Y. (2013). Opportunities and Constraints to Legume Diversification for Sustainable Maize Production on Smallholder Farms in Malawi. *Renewable Agriculture and Food Systems*, *28*(03), 234-244.

Mmbando, F. E. (2014). *Market participation, channel choice and impacts on household welfare: the case of smallholder farmers in Tanzania* (Doctoral dissertation). University of KwaZulu-Natal, Pietermaritzburg, South Africa.

Mmbando, F. E., Wale, E. Z., & Baiyegunhi, L. J. (2015). Welfare Impacts of Smallholder Farmers' Participation in Maize and Pigeonpea Markets in Tanzania. *Food Security, 7*(6), 1211-1224.

Mmbando, F. E., Wale, E., Baiyegunhi, L. J. S., & Darroch, M. A. G. (2016). The Choice of Marketing Channel by Maize and Pigeonpea Smallholder Farmers: Evidence from the Northern and Eastern Zones of Tanzania. *Agrekon, 55*(3), 254-277.

Minten, B., & Barrett, C. (2008). Agricultural technology, productivity, and poverty in Madagascar. *World Development, 36*(5), 797-822.

Mottaleb, K, Mohanty, S., & Nelson, A. (2015). Strengthening Market Linkages of Farm Households in Developing Countries. *Applied Economic Perspectives and Policy, 37*(2), 226-242.

Muricho, G. S. (2015). *Determinants of agricultural commercialization and its impacts on welfare among smallholder farmers in Kenya*. (Ph.D.), University of Nairobi, Nairobi.

Muriithi, B., & Matz, J. (2015). Welfare effects of vegetable commercialization: Evidence from smallholder producers in Kenya. *Food Policy, 50*, 80 - 91.

Muui, C. W., Muasya, R. M., & Kirubi, D. T. (2013). Baseline Survey on Factors affecting Sorghum Production and Use in Eastern Kenya. *African Journal of Food, Agriculture, Nutrition and Development, 13*(1), 7339-7353.

Nyongesa, D., Mabele, R., Esilaba, A., & Mutoni, C. (2017). The economics and gender factor in soya bean production and profitability in Kenya: A case of smallholder farms in Western Kenya. *International Journal of Agricultural Resources, Governance and Ecology, 13*(3), 211-240.

Odendo M, Onyango E., & Oucho P, 2013. Agronomic and financial analysis of maize- legumes production in western Kenya. Joint proceedings of the 27th Soil Science Society of East Africa and the 6th African Soil Science Society, 20-25 October, Nakuru, Kenya

OECD/FAO. (2016). *OECD-FAO Agricultural Outlook 2016-2025*. Paris: OECD Publishing.

Ogada, M. O. (2015). Constraints to increasing Soybean Production and Productivity in Benue State, Nigeria. *Asian Journal of Agricultural Extension, Economics & Sociology, 4*(4),: 277-284.

Ogunleye, K., & Oladeji, J. (2007). Choice of cocoa market channels among cocoa farmers in Ila local government area of Osun State, Nigeria. *Middle-East Journal of Scientific Research, 1*(1), 14-20.

Omiti, J., Otieno, D., Nyanamba, T., & McCulloug, E. (2009). Factors influencing the intensity of market participation by smallholder farmers: A case study of rural and peri-urban areas of Kenya. *African Journal of Agricultural and Resource Economics, 3*(1), 57–82.

Ong'ayo, A. H. (2017). Analysis of the effect of agricultural extension development initiatives on household? Agricultural food productivity and sufficiency among small-scale farmers: A case of Kilifi County, Kenya. *Journal of Agriculture and Food Science, 5*(2), 13-20.

Onyeneke, R. (2017). Determinants of adoption of improved technologies in rice production in Imo State, Nigeria. *African Journal of Agricultural Research, 12*(11), 888-896.

Opondo, F. A. (2018). Evaluation of cassava commercialization among smallholder farmers and its effect on household income: a case of Siaya and Kilifi counties. Doctor of Philosophy thesis, Egerton University.

Osmani, M., Islam, K., Ghosh, C., & Hossain, M. E. (2014). Commercialization of smallholder farmers and its welfare outcomes: Evidence from Durgapur Upazila of Rajshahi District, Bangladesh. *Journal of World Economic Research, 3*(6), 119-126.

Otsuka, K., Larson, D. F., & Hazell, P. B. R. (2013). *An African Green Revolution: Finding ways to Boost Productivity on Small Farms*. Dordrecht: Springer.

Panda, P., & Sreekumar, E. (2012). Marketing channel choice and marketing efficiency assessment in agribusiness. *Journal of International Food and Agribusiness Marketing, 24*(3), 213-230.

Pingali, P. (1997). From subsistence to commercial production systems: The transformation of Asian agriculture. *American Journal of agricultural Economics, 79*(2): 628-634.

Pingali, L., & Rosegrant, M. (1995). Agricultural commercialization and diversification: Process and polices. *Food Policy, 20* (3), 171–185.

Rahut, D., Ali, A., Castellanos, V., & Sahoo, P. (2015). Market participation by subsistence farmers: Evidence from the Himalayas. *The Journal of Animal and Plant Sciences, 25*(6), 1713-1721.

Randela, R., Alemu, Z. G., & Groenewald, J. A., (2008). Factors enhancing market participation by small-scale cotton farmers. *Agrekon, 47*(4), 234-267.

Rao, E. J., Brümmer, B., & Qaim, M. (2012). Farmer Participation in Supermarket Channels, Production Technology, and Efficiency: The Case of Vegetables in Kenya. *American Journal of Agricultural Economics, 94*(4), 891–912.

Rogers, E. (1995). *Diffusion of Innovations* (4th edition.). New York: The Free Press.

Rokani, L., Ekere, W., Walekwa, P., & Ebanyat, P. (2014). Effect of Entrepreneurship Training of Smallholder Farmers on Soybean Productivity and Household Incomes in Lango Sub-region, Northern Uganda. In *FourthRUFORUM Biennial Regional Conference* held in July, (pp. 21-25).

Russell, G. J., & Petersen, A. (2000). Analysis of Cross-Category dependence in Market Basket Selection. *Journal of Retailing, 76*(3): 367-392.

Sanginga, P., Adesina, A., Manyong, M., Otite, O., & Dashiell, K. (1999). *Social impact of soybean in Nigeria's southern Guinea savanna*. Nigeria: International Institute of Tropical Agriculture.

Seyoum, C., Lemma, T., & Karippai, R. (2011). Factors Determining the Degree of Commercialization of Smallholder Agriculture: The Case of Potato Growers in Kombolcha District, East Hararghe, Ethiopia. *Journal of Agricultural Development, 2*(1), 20-30.

Sharma, S., Jayanta, H., & Deka, N. (2016). Commercialization of small holder farming in Assam. *Economic Affairs, 61*(2), 335-337.

Shiimi, T., Taljaard, P.R., & Jordaan, H. (2012). Transaction costs and cattle farmers' choice of marketing channel in North-Central Namibia. *Agrekon, 51*(1), 42-58.

Singh, I., Singh, K. K., & Gautam, U. S. (2012). Constraints in Adoption of Soybean Production Ttechnology. *Indian Research Journal of Extension Education, 2*, 169-171.

Tsourgiannis, L., Eddison, J., & Warren, M. (2008). Factors affecting the Marketing Channel Choice of Sheep and Goat Farmers in the Region of East Macedonia in Greece regarding the Distribution of their Milk Production. *Small Ruminant Research, 79*(1), 87-97.

Turaa, E., Goshub, D., Demisiec, T., & Kenead, T. (2016). Determinants of market participation and intensity of marketed surplus of teff producers in Bacho and Dawo Districts of Oromia State, Ethiopia. *Journal of Economics and Sustainable Development, 7*(1)21, 13-24.

Umberger, W. J., Reardon, T., Stringer, R., & Mueller Loose, S. (2015). Market-Channel Choices of Indonesian Potato Farmers: A Best–Worst Scaling Experiment. *Bulletin of Indonesian Economic Studies, 51*(3): 461-477.

Van Ittersum K., Van Bussela L., Wolfa J., Grassinib P., Van Wart J., Guilpartb J., Claessensc L., de Grootd H., Wiebee K., Mason-D'Croze D., Yangb H., Boogaard H., Van Oort P., Van Loona M., Saito K., Adimo O., Adjei-Nsiah S., Agali A, Bala A., Chikowo R., Kaizzi K.,

Kouressy M., Makoi J., Ouattara K., Tesfaye K., & Cassman K., 2016. Can sub-Saharan Africa feed itself? *Proceedings of the National assembly of Sciences, 113* (52), 14964-14969.

Verkaart, S., Munyua, B. G., Mausch, K. & Michler, J. D. (2017). Welfare Impacts of Improved Chickpea Adoption: A Pathway for Rural Development in Ethiopia? *Food Policy, 66*, 50-61.

Wambugu P. W., & Muthamia, Z. K. (2009). *Country Report on the State of Plant GeneticResources for Food and Agriculture*. FAO. Retrieved 13 July 2017 from http://www.fao.org/docrep/013/i1500e/kenya.pdf

Weinhold, D., Killick, E., & Reis, E. J. (2013). Soybeans, Poverty and Inequality in the Brazilian Amazon. *World Development, 52*: 132-143.

Woldeyohanes, T., Heckelei, T., & Surry, Y. (2017). Effect of off-farm income on smallholder commercialization: Panel evidence from rural households in Ethiopia. *Agricultural Economics, 48*(2), 207-218.

Wollni, M., Lee, D., & Janice, T. (2010). Conservation agriculture, organic marketing, and collective action in the Honduran hillsides. *Agricultural Economics, 41*, 373-384.

Wooldridge J. (2010). *Econometric Analysis of Cross-Sectional and Panel Data.* Cambridge, Massachusetts: The MIT Press.

Wooldridge, J. M. (2013). *Introductory Econometrics: A modern approach.* Mason, OH: South-Western Cengage Learning.

Word Bank, 2015. *World Development Report 2015: Agriculture for Development*. Washington, D.C.

Xaba, B. G., & Masuku, M. B. (2012). Factors affecting the Choice of Marketing Channel by Vegetable Farmers in Swaziland. *Sustainable Agriculture Research, 2*(1), 112- 127.

Yen, B. T., Visser, S. M., Hoanh, C. T., & Stroosnijder, L. (2013). Constraints on Agricultural Production in the Northern Uplands of Vietnam. *Mountain Research and Development, 33*(4), 404-415.

Zamasiya, B., Mango, N., Nyikahadzoi, K., & Siziba, S. (2014). Determinants of soybean market participation by smallholder farmers in Zimbabwe. *Journal of Development and Agricultural Economics, 6*(2), 49-58.

Zivenge, E., & Karavina, C. (2012). Analysis of factors influencing market channel access by communal horticulture farmers in Chinamora District, Zimbabwe. *Journal of Development and Agricultural Economics, 4*(6), 47-150.

Zhou, S., Minde, I., & Mtigwe, B. (2013). Smallholder agricultural commercialization for income growth and poverty alleviation in Southern Africa: A review. *African Journal of Agricultural Research, 8*(22), 2599 -2608.

APPENDICES

Appendix I: Household Questionnaire

My name is Martha Wanjiru a master's student at Egerton University undertaking a research on the"**effect of soybean commercialization on income of smallholder farmers in Butere sub-county, Kenya."**The purpose of this study is purely academic and you have been identified as one of the respondents who will assist in this research. Your contribution will be highly appreciated and utmost discretion will be accorded to your responses even as they will be solely used for research purposes.

SECTION A: GENERAL INFORMATION

A1. Name of Enumerator…….

A2. County……..

A3. Ward…….

A4. Village…….

A5. Gps coordinate……..

A6. Region (0=rural, 1=urban)

A.7 Date of the interview……

SECTION B: FARMER, FARM AND INSTITUTIONAL CHARACTERISTICS

B8. Name of the respondent……..

B9. Relationship to the household head

1= household head [ ],2=spouse[ ]; 3=son[ ];4=daughter[ ]Others (specify)……..

B10. Gender of the household head ☐ male ☐ female

B11. Please indicate number of schooling years of the household head…….and years in agricultural training, if attended…….

B12. What is the occupation of the household?

1=casual laborer [ ]; 2=salaried employed [ ]; 3=off-farm activity [ ]; 4= soy bean farming [ ]; 5=other farming activities 5 = Others (specify)……

B13. Who makes the decision on farming?

1=Husband [ ] 2=wife[ ] 3=jointly[ ] 4=others (specify)……..

B14. Age of the household head.....

B15. Marital status of household head

Single ☐ married ☐

B16. What is the total number of household members at least in the last one year? ……

B17. What is the number and the value of livestock owned in last one year?

No.	Livestock list	Number owned in the last one year	Value in KES.
1	Cows		
2	Oxen		
3	Bulls		
4	Heifers		
5	Calves		
6	Goats		
7	Sheep		
8	Donkey		
9	Horses		
10	Poultry		
11	Others specify		

B18. What is the total size of land owned in acres?.....

B19. What is the walking distance in minutes from the farm to the nearest market center? ……

B20. What is the walking distance in minutes from the farm to the nearest administrative center? .....

B21. What is the walking distance in minutes from the farm to nearest source of seed dealer? ……..

B22. What is the walking distance in minutes from the farm to nearest formal credit service source?.....

B23. Indicate your group membership status in the table below

Are you a member of any agricultural group (0= Yes, 1=No)	Benefits acquired from membership	How many times do you meet in a year?	Do you hold any positions in the group? (0= yes, 1=no)	If yes, specify the position.

	1= marketing 2=production 3= inputs 4= credit/loans 5= others, specify			

B24. Do you have access to credit facilities?

0= No [ ] 1= Yes [ ]

B25. Have you ever borrowed money to use in the last one year?

0= No [ ] 1= Yes [ ]

B26. If yes (B25), state the amount borrowed? ……and how much allocated for farming?……

B27. Did you save money in the last 12 months?

0= No [ ] 1= Yes [ ]

B28. If yes (B27), how much did you save in the last 12 months?.....

B29. Kindly indicate any other off-farm income from other sources or income transfers in ksh…….

SECTION C: PRODUCTION CONSTRAINTS AND PLOT ALLOCATION CHARACTERISTICS

C29. Which variety of soybean do you produce _____

1. Local 2. Improved/hybrid 3. Both

C30. How long have you been growing soybean? .....

C31. What are the factors constraining soybean production?

Constraint	Do you experience the following constraints : 1 = Yes, 0 = No	Rank from the most important
High initial cost of production		
Inadequate capital		
Missing market		

Inadequate land		
Access to credit		
Pest and diseases		
High input Prices		
Inadequate extension services		
Unfavourable Weather		

C32. What size of the land is under soybean in acres for season 1 and 2 ? Fill in the table below

Season 1:

Land tenure sytem	Acres of total land	Amount occupied by soybean (acres)	Fertility of the land (1= high, 2= medium, 3= low)	Cost of leasing (Kshs/acre) If applicable
Owned				
Leased				
Communal				

Season 2:

Land tenure sytem	Acres of total land	Amount occupied by soybean (acres)	Fertility of the land (1= high, 2= medium, 3= low)	Cost of leasing (Kshs/acre)
Owned				
Leased				
Communal				

C33. Apart from soybean what are your other on-farm sources of income? Fill in the table below

Crop	Area (Acres) S1	Area (Acres) S2	Unit	Season One Output	Unit Price Season One	Season Two Output	Unit Price season one
Sugarcane							

Maize							
Sweet potatoes							
Tea							
Groundnuts							
Soybean							
Livestock							
others(specify)							

Unit codes: 1=kgs, 2=cart, 3=wheelbarrow, 4= bags, 5=others (specify)

SECTION D. INPUTS USE

D34 Did you use any inputs in soybean production last season?

0= No [ ] 1= Yes [ ]

D35. If yes in (D34), please indicate the quantity and cost used in soybean production in season 1 .

Input	Type (Optional)	Unit	Quantity	Cost/unit	Total cost
Manure					
Fertilizer					
Pesticides					
Fungicides					
Herbicide					
Others specify					

Unit codes: 1=kgs, 2=cart, 3=wheelbarrow, 4= bags, 5=others (specify)

D36. If yes in (D34), please indicate the quantity and cost used in soybean production in season 2.

Input	Type (Optional)	Unit	Quantity	Cost/unit	Total cost
Manure					
Fertilizer					
Pesticides					
Fungicides					

Herbicide					
Others specify					

Unit codes: 1=kgs, 2=cart, 3=wheelbarrow, 4= bags, 5=others (specify)

D37. Indicate in the table below any other additional cost on input purchase

Inputs	Mode of delivery to the farm	Price of transportation	Time spent to deliver the inputs to the farm
Manure			
Fertilizer			
Pesticide			
Fungicide			
Herbicide			
Others specify			

SECTION E. LABOUR USE

E38. Indicate in the table below, family labor used in soybean production in the last two seasons for male and female.
Season 1: Family labour for male

Activity	No. male	Days worked	Hours worked	Total man-days worked (8 hours per day)	Price per man-day	Total cost
First tillage						
Second tillage						
Planting						
First weeding						
Second weeding						
Third weeding						

Pesticide application						
Fertilizer application (Top dressing)						
Fungicide application						
Harvesting						
Herbicide application						
Others activities						

Season 1: Family labour for female

Activity	No. female	Days worked	Hours worked	Total man-days worked (8 hours per day)	Price per man-day	Total cost
First tillage						
Second tillage						
Planting						
First weeding						
Second weeding						
Third weeding						
Pesticide application						
Fertilizer application (Top dressing)						
Fungicide application						
Harvesting						
Herbicide application						
Others activities						

Season 2: Family labour for male

Activity	No. male	Days worked	Hours worked	Total man-days worked	Price per man-day	Total cost

				(8 hours per day)		
First tillage						
Second tillage						
Planting						
First weeding						
Second weeding						
Third weeding						
Pesticide application						
Fertilizer application (Top dressing)						
Fungicide application						
Harvesting						
Herbicide application						
Others activities						

Season 2: Family labour for female

Activity	No. female	Days worked	Hours worked	Total man-days worked (8 hours per day)	Price per man-day	Total cost
First tillage						
Second tillage						
Planting						
First weeding						
Second weeding						
Third weeding						
Pesticide application						
Fertilizer application (Top dressing)						

Fungicide application						
Harvesting						
Herbicide application						
Others activities						

E39. Indicate in the table below, hired labour used in soybean production in the last two seasons for male and female.

Season 2: Hired labour for male

Activity	No. male	Days worked	Hours worked	Total man-days worked (8 hours per day)	Price per man-day	Total cost
First tillage						
Second tillage						
Planting						
First weeding						
Second weeding						
Third weeding						
Pesticide application						
Fertilizer application (Top dressing)						
Fungicide application						
Harvesting						
Herbicide application						
Others activities						

Season 2: Hired labour for female

Activity	No. female	Days worked	Hours worked	Total man-days worked (8 hours per day)	Price per man-day	Total cost
First tillage						

Second tillage						
Planting						
First weeding						
Second weeding						
Third weeding						
Pesticide application						
Fertilizer application (Top dressing)						
Fungicide application						
Harvesting						
Herbicide application						
Others activities						

Season 2: Hired labour for male

Activity	No. male	Days worked	Hours worked	Total man-days worked (8 hours per day)	Price per man-day	Total cost
First tillage						
Second tillage						
Planting						
First weeding						
Second weeding						
Third weeding						
Pesticide application						
Fertilizer application (Top dressing)						
Fungicide application						
Harvesting						

Herbicide application						
Others activities						

Season 2: Hired labour for female

Activity	No. female	Days worked	Hours worked	Total man-days worked (8 hours per day)	Price per man-day	Total cost
First tillage						
Second tillage						
Planting						
First weeding						
Second weeding						
Third weeding						
Pesticide application						
Fertilizer application (Top dressing)						
Fungicide application						
Harvesting						
Herbicide application						
Others activities						

SECTION F. SOYBEAN YIELDS AND MARKETING

F40. Indicate in the table below the amount of soybean produced in the last two seasons.

Number of seasons	Total Yield	Unit	Consumed	Sold	Gift and donations	Price per unit (KES)	Total revenue (KES)
Sn1							
Sn2							

Unit codes:1=kgs, 2 =debe, 3=gorgoro, 4=90kg bag

F41. Indicate in the table below the amount of other crops produced in the last two seasons

Number of seasons	Total Yield	Unit	Consumed	Sold	Gift and donations	Price per unit (KES)	Total revenue (KES)
Season 1							
Season 2							

Crop type: 1 = maize 2 = Sorghum, 3 = millet, 4 =Sugarcane, 5 = Sweet potatoes, 6 = Tea, 7 = Bean 8 = other specify. **Unit codes**:1=kgs, 2 =debe, 3=gorgoro, 4=90kg bag

F42: Indicate the amount of soybean sold, price, channel, requirement, reasons and payment mode in the table below

Season	Market Participation (1= Yes, 0=No)	Main marketing channel	Quantity sold	Price/unit	Revenue	Mode of payment	Main reason for channel choice	Main requirement of channel
Season 1								
Season 2								

Marketing Channel: 0 = non market participation, 1= Direct, 2= Middlemen, 3= Cooperatives, 4 = Agro-processors, 5= Others specify

Unit codes: 1=kgs, 2=bags/sacs,3=gorogoros,4=others (specify)

Mode of payment: 0 = cash, 1= Mpesa 2= Bank

Reason codes: 1=high price, 2=reliable (will always purchase), 3=accessibility, 4=only available channel, 5=don't know where to sell, 6=low quality not fit for other channels,7=others(specify)

Requirement codes: 1=quality,2=transport,3=group member,4=contract,5=variety(specify)

F43. Are you satisfied with your current main channel?

Yes ☐ No ☐

F44. If no (F43), why?

1= Relatively low prices, 2=distance to the market, 3=unreliable market, 4=information asymmetry, 5=high cost of transaction,6=others (specify)

F45: If yes in (F39), why?

1= Relatively high price, 2=distance to the market, 3=reliable market,4=information asymmetry, 5=low cost of the transaction, 6=others (specify)

SECTION H: EXTENSION

H46.Did you have any extension contact in the last 12 months?

Yes ☐ No ☐

H47. If yes (H46), Indicate the source, topic, number of visit and level of satisfaction with extension services.

Provider extension /agent 1. Government 2. NGOs 3. Lead farmer 4. Others (specify)	What was the extension service about? 1. Crop production 2. Dairy production 3. Marketing 4. Others (specify)	Number of times of the visits	Level of satisfaction 1. Very satisfied 2. Satisfied 3. Neutral 4. Dissatisfied 5. Very dissatisfied

Appendix II: Output for Descriptive Results, t-test and Chi^2-Test

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.0   Copyright 1985-2015 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

30-student Stata lab perpetual license:
       Serial number:  401406221642
         Licensed to:  Kenya
                       The Real Africa

Notes:
      1.  Unicode is supported; see help unicode_advice.

. use "C:\Users\user\Desktop\Workings\Shiro\Merged 2.dta"

. *******EFFECT OF SOYBEAN COMMERCIALIZATION ON INCOME OF SMALLHOLDER FARMERS IN BUTERE SUB-COUNTY, KAKAMEG
> A COUNTY, KENYA***********

.
. *******This do file presents the analysis of data collected from smallholder soybean farmers in Butere Su
> b-County, Kenya.**********

.
. ********Descriptive statistics**********

.
. use "C:\Users\user\Desktop\Workings\Shiro\Merged 2.dta"

.
. ******Categorical socio-economic variables*******

.
. tab DID_YOU_PARTICIPATE_IN_THE_SOYBE

    Did you |
participate |
     in the |
    soybean |
market last |
     season |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         87       43.28       43.28
        Yes |        114       56.72      100.00
------------+-----------------------------------
      Total |        201      100.00

.
. tab SEX_HHHEAD DID_YOU_PARTICIPATE_IN_THE_SOYBE, chi2

           | Did you participate
 Gender of | in the soybean market
 household |      last season
      head |        No        Yes |     Total
-----------+---------------------+----------
    Female |        26         44 |        70
      Male |        61         70 |       131
-----------+---------------------+----------
     Total |        87        114 |       201

          Pearson chi2(1) =   1.6498   Pr = 0.199

.
. tab SEX_HHHEAD if DID_YOU_PARTICIPATE_IN_THE_SOYBE==1

  Gender of |
  household |
       head |      Freq.     Percent        Cum.
------------+-----------------------------------
     Female |         44       38.60       38.60
       Male |         70       61.40      100.00
------------+-----------------------------------
      Total |        114      100.00

.
. tab SEX_HHHEAD if DID_YOU_PARTICIPATE_IN_THE_SOYBE==0

  Gender of |
  household |
```

```
      head |      Freq.     Percent        Cum.
-----------+-----------------------------------
    Female |         26       29.89       29.89
      Male |         61       70.11      100.00
-----------+-----------------------------------
     Total |         87      100.00

. 
. tab M_status DID_YOU_PARTICIPATE_IN_THE_SOYBE, chi2

           | Did you participate
           | in the soybean market
           |      last season
  M_status |        No        Yes |     Total
-----------+----------------------+----------
 otherwise |        16         32 |        48
   Married |        71         82 |       153
-----------+----------------------+----------
     Total |        87        114 |       201

          Pearson chi2(1) =   2.5432   Pr = 0.111

. 
. tab M_status

  M_status |      Freq.     Percent        Cum.
-----------+-----------------------------------
 otherwise |         48       23.88       23.88
   Married |        153       76.12      100.00
-----------+-----------------------------------
     Total |        201      100.00

. 
. tab M_status if DID_YOU_PARTICIPATE_IN_THE_SOYBE==1

  M_status |      Freq.     Percent        Cum.
-----------+-----------------------------------
 otherwise |         32       28.07       28.07
   Married |         82       71.93      100.00
-----------+-----------------------------------
     Total |        114      100.00

. 
. tab M_status if DID_YOU_PARTICIPATE_IN_THE_SOYBE==0

  M_status |      Freq.     Percent        Cum.
-----------+-----------------------------------
 otherwise |         16       18.39       18.39
   Married |         71       81.61      100.00
-----------+-----------------------------------
     Total |         87      100.00

. 
. tab PRIMARY_OCCUPATION DID_YOU_PARTICIPATE_IN_THE_SOYBE, chi2

                | Did you participate
        Primary | in the soybean market
occupation of the |      last season
  household head |        No        Yes |     Total
-----------------+----------------------+----------
        Farming |        64         92 |       156
 Business person |         8         10 |        18
  Casual Laborer |         3          3 |         6
Salaried Employee |        10          9 |        19
        Student |         1          0 |         1
 Others (Specify) |         1          0 |         1
-----------------+----------------------+----------
          Total |        87        114 |       201

          Pearson chi2(5) =   3.7411   Pr = 0.587

. 
. tab PRIMARY_OCCUPATION

        Primary |
occupation of the |
  household head |      Freq.     Percent        Cum.
-----------------+-----------------------------------
        Farming |        156       77.61       77.61
 Business person |         18        8.96       86.57
  Casual Laborer |          6        2.99       89.55
Salaried Employee |         19        9.45       99.00
        Student |          1        0.50       99.50
 Others (Specify) |          1        0.50      100.00
-----------------+-----------------------------------
          Total |        201      100.00

.
```

```
. tab PRIMARY_OCCUPATION if DID_YOU_PARTICIPATE_IN_THE_SOYBE ==0

        Primary |
occupation of the |
  household head |      Freq.     Percent        Cum.
-----------------+-----------------------------------
        Farming |         64       73.56       73.56
 Business person |          8        9.20       82.76
  Casual Laborer |          3        3.45       86.21
Salaried Employee |         10       11.49       97.70
        Student |          1        1.15       98.85
 Others (Specify) |          1        1.15      100.00
-----------------+-----------------------------------
          Total |         87      100.00

.
. tab PRIMARY_OCCUPATION if DID_YOU_PARTICIPATE_IN_THE_SOYBE ==1

        Primary |
occupation of the |
  household head |      Freq.     Percent        Cum.
-----------------+-----------------------------------
        Farming |         92       80.70       80.70
 Business person |         10        8.77       89.47
  Casual Laborer |          3        2.63       92.11
Salaried Employee |          9        7.89      100.00
-----------------+-----------------------------------
          Total |        114      100.00

.
. ttest AGE_HHHEAD , by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    59.04598    1.122765    10.47245    56.81399    61.27796
     Yes |     114    50.25439    1.112226    11.87532    48.05087    52.45791
---------+--------------------------------------------------------------------
combined |     201      54.0597    .8519992    12.07917    52.37965    55.73976
---------+--------------------------------------------------------------------
    diff |            8.791591    1.607303                5.622059    11.96112
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =   5.4698
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000

.
. ttest YEARS_SCHOOL , by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87     6.91954    .4008873    3.739228    6.122603    7.716478
     Yes |     114    10.38596    .2869919    3.064235    9.817382    10.95455
---------+--------------------------------------------------------------------
combined |     201    8.885572    .2665594    3.779132    8.359945      9.4112
---------+--------------------------------------------------------------------
    diff |           -3.466425     .480115               -4.413191   -2.519659
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -7.2200
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000

.
. ttest HH_SIZE, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    7.218391    .3267563     3.04778    6.568821    7.867961
     Yes |     114     4.72807    .1714279    1.830349    4.388441      5.0677
---------+--------------------------------------------------------------------
combined |     201     5.80597    .1920971    2.723447    5.427175    6.184766
---------+--------------------------------------------------------------------
    diff |            2.490321    .3462783                1.807475    3.173166
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =   7.1917
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
```

```
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000

.
. ttest EXPERIENCE_IN_SOYBEAN_PRODUCTION, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    5.229885    .5233651    4.881625     4.18947      6.2703
     Yes |     114    7.122807    .6297175    6.723543    5.875223    8.370391
---------+--------------------------------------------------------------------
combined |     201    6.303483    .4271416    6.055777    5.461204    7.145761
---------+--------------------------------------------------------------------
    diff |           -1.892922    .8537794               -3.576538   -.2093061
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.2171
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0139         Pr(|T| > |t|) = 0.0277          Pr(T > t) = 0.9861

.
. ttest Total_Cost_of_input_per_acre, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    4470.554    2109.799     19678.9    276.4121    8664.697
     Yes |     114    4171.322    627.8634    6703.747    2927.412    5415.233
---------+--------------------------------------------------------------------
combined |     201    4300.841    977.1894    13854.05    2373.925    6227.757
---------+--------------------------------------------------------------------
    diff |            299.232    1977.089               -3599.501    4197.965
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =   0.1513
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.5601         Pr(|T| > |t|) = 0.8799          Pr(T > t) = 0.4399

.
. ttest Total_Cost_of_labour_per_acre , by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    27334.39    8404.783    78394.6    10626.23    44042.54
     Yes |     114    13174.33    1058.689    11303.7    11076.87    15271.78
---------+--------------------------------------------------------------------
combined |     201    19303.31    3708.48    52576.77    11990.57    26616.05
---------+--------------------------------------------------------------------
    diff |            14160.06    7436.135               -503.676    28823.79
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =   1.9042
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9708         Pr(|T| > |t|) = 0.0583          Pr(T > t) = 0.0292

.
. ttest TOTAL_OTHERS, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    40827.82     6291.67    58684.79    28320.39    53335.24
     Yes |     114    50622.89    5266.137    56226.96    40189.73    61056.06
---------+--------------------------------------------------------------------
combined |     201    46383.23    4046.238    57365.33    38404.47      54362
---------+--------------------------------------------------------------------
    diff |           -9795.079    8157.485               -25881.29    6291.128
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -1.2007
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1156         Pr(|T| > |t|) = 0.2313          Pr(T > t) = 0.8844

.
. ttest TOTAL_INCOME_FROM_REMITTANCES, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
```

```
   Group |     Obs       Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    2260.92    804.1232    7500.362    662.3755    3859.464
     Yes |     114    5908.772    1213.518    12956.83    3504.574     8312.97
---------+--------------------------------------------------------------------
combined |     201    4329.851    780.1158    11060.05    2791.543    5868.158
---------+--------------------------------------------------------------------
    diff |           -3647.852    1557.127               -6718.439    -577.266
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -2.3427
Ho: diff = 0                                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0101         Pr(|T| > |t|) = 0.0201          Pr(T > t) = 0.9899

.
. tab  TYPE_OF_LAND_OWNERSHIP_TENURE DID_YOU_PARTICIPATE_IN_THE_SOYB, chi2

   Type of |  Did you participate
      land | in the soybean market
 ownership |      last season
 (tenure) |        No        Yes |     Total
----------+----------------------+----------
     Owned |        81        109 |       190
    Leased |         4          5 |         9
 Communal |         2          0 |         2
----------+----------------------+----------
     Total |        87        114 |       201

          Pearson chi2(2) =   2.6585   Pr = 0.265

.
. tab TYPE_OF_LAND_OWNERSHIP_TENURE

    Type of |
       land |
 ownership |
  (tenure) |      Freq.     Percent        Cum.
----------+-----------------------------------
     Owned |        190       94.53       94.53
    Leased |          9        4.48       99.00
 Communal |          2        1.00      100.00
----------+-----------------------------------
     Total |        201      100.00

.
. tab  TYPE_OF_LAND_OWNERSHIP_TENURE if DID_YOU_PARTICIPATE_IN_THE_SOYB==1

    Type of |
       land |
 ownership |
  (tenure) |      Freq.     Percent        Cum.
----------+-----------------------------------
     Owned |        109       95.61       95.61
    Leased |          5        4.39      100.00
----------+-----------------------------------
     Total |        114      100.00

.
. tab  TYPE_OF_LAND_OWNERSHIP_TENURE if DID_YOU_PARTICIPATE_IN_THE_SOYB==0

    Type of |
       land |
 ownership |
  (tenure) |      Freq.     Percent        Cum.
----------+-----------------------------------
     Owned |         81       93.10       93.10
    Leased |          4        4.60       97.70
 Communal |          2        2.30      100.00
----------+-----------------------------------
     Total |         87      100.00

.
. tab FERTILITY_OF_THE_LAND DID_YOU_PARTICIPATE_IN_THE_SOYB, chi2

          |  Did you participate
 Fertility | in the soybean market
   of the |      last season
     land |        No        Yes |     Total
----------+----------------------+----------
     high |        25         22 |        47
   medium |        52         85 |       137
      low |        10          7 |        17
----------+----------------------+----------
     Total |        87        114 |       201

          Pearson chi2(2) =   5.1356   Pr = 0.077
```

```
.
. tab FERTILITY_OF_THE_LAND

  Fertility |
of the land |      Freq.     Percent        Cum.
------------+-----------------------------------
       high |         47       23.38       23.38
     medium |        137       68.16       91.54
        low |         17        8.46      100.00
------------+-----------------------------------
      Total |        201      100.00

.
. tab FERTILITY_OF_THE_LAND if DID_YOU_PARTICIPATE_IN_THE_SOYB==1

  Fertility |
of the land |      Freq.     Percent        Cum.
------------+-----------------------------------
       high |         22       19.30       19.30
     medium |         85       74.56       93.86
        low |          7        6.14      100.00
------------+-----------------------------------
      Total |        114      100.00

.
. tab FERTILITY_OF_THE_LAND if DID_YOU_PARTICIPATE_IN_THE_SOYB==0

  Fertility |
of the land |      Freq.     Percent        Cum.
------------+-----------------------------------
       high |         25       28.74       28.74
     medium |         52       59.77       88.51
        low |         10       11.49      100.00
------------+-----------------------------------
      Total |         87      100.00

.
. ttest ACREAGE_SIZE_IN_ACRES , by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    1.618678    .1562497      1.4574    1.308064    1.929292
     Yes |     114    1.689912    .1397662    1.492295     1.41301    1.966814
---------+--------------------------------------------------------------------
combined |     201     1.65908    .1039709    1.474043     1.45406      1.8641
---------+--------------------------------------------------------------------
    diff |          -.0712341    .2103097               -.4859557    .3434875
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -0.3387
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.3676         Pr(|T| > |t|) = 0.7352          Pr(T > t) = 0.6324

.
. ttest TOTAL_LAND_SIZE , by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87     1.66477    .1582012    1.475603    1.350276    1.979264
     Yes |     114    1.870833    .1470733    1.570313    1.579455    2.162212
---------+--------------------------------------------------------------------
combined |     201    1.781642    .1078968    1.529701    1.568881    1.994403
---------+--------------------------------------------------------------------
    diff |          -.2060632    .2178244               -.6356035    .2234771
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -0.9460
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1726         Pr(|T| > |t|) = 0.3453          Pr(T > t) = 0.8274

.
. ttest LAND_SOYBEAN , by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87     .231069    .022201    .2070771    .1869348    .2752031
     Yes |     114    .3734211    .026106    .2787355    .3217004    .4251417
---------+--------------------------------------------------------------------
```

```
combined |     201    .311806    .018303    .2594901    .2757143    .3478976
---------+--------------------------------------------------------------------
    diff |          -.1423521    .0356322               -.2126172   -.0720869
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -3.9950
Ho: diff = 0                                 degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0001          Pr(T > t) = 1.0000


. ttest TOTAL_YIELD_LAST_SEASON, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    23.34483    2.915869    27.19742    17.54827    29.14138
     Yes |     114    83.17544    7.385035    78.85059    68.54435    97.80652
---------+--------------------------------------------------------------------
combined |     201    57.27861    4.842845    68.65917    47.72902    66.82819
---------+--------------------------------------------------------------------
    diff |          -59.83061     8.83335               -77.24959   -42.41163
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -6.7733
Ho: diff = 0                                 degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000

.
. ttest TOTAL_SOYBEAN_YIELD_PER_ACRE, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    153.7285    26.95101    251.3823    100.1517    207.3053
     Yes |     114    259.8694    20.17645    215.4255    219.8962    299.8425
---------+--------------------------------------------------------------------
combined |     201    213.9278     16.7173    237.0087     180.963    246.8926
---------+--------------------------------------------------------------------
    diff |          -106.1409    32.97761               -171.1713   -41.11045
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -3.2186
Ho: diff = 0                                 degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0008         Pr(|T| > |t|) = 0.0015          Pr(T > t) = 0.9992

.
. tab EXTENSION DID_YOU_PARTICIPATE_IN_THE_SOYB, chi2

   Did you |
  have any |
 extension |  Did you participate
contact in | in the soybean market
  the last |      last season
 12 month |        No        Yes |     Total
-----------+----------------------+----------
        No |        49         27 |        76
       Yes |        38         87 |       125
-----------+----------------------+----------
     Total |        87        114 |       201

          Pearson chi2(1) =  22.3529   Pr = 0.000

.
. tab EXTENSION if DID_YOU_PARTICIPATE_IN_THE_SOYB==0

    Did you |
   have any |
  extension |
 contact in |
the last 12 |
      month |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         49       56.32       56.32
        Yes |         38       43.68      100.00
------------+-----------------------------------
      Total |         87      100.00

.
. tab EXTENSION if DID_YOU_PARTICIPATE_IN_THE_SOYB==1

    Did you |
   have any |
```

```
  extension |
 contact in |
the last 12 |
     month |      Freq.     Percent        Cum.
-----------+-----------------------------------
        No |         27       23.68       23.68
       Yes |         87       76.32      100.00
-----------+-----------------------------------
     Total |        114      100.00

.
. tab EXTENSION

    Did you |
   have any |
  extension |
 contact in |
the last 12 |
     month |      Freq.     Percent        Cum.
-----------+-----------------------------------
        No |         76       37.81       37.81
       Yes |        125       62.19      100.00
-----------+-----------------------------------
     Total |        201      100.00

.
. tab DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC DID_YOU_PARTICIPATE_IN_THE_SOYB, chi2

    Do you |
     have |  Did you participate
 access to | in the soybean market
   credit |      last season
facilities |        No        Yes |     Total
-----------+----------------------+----------
        No |        44         65 |       109
       Yes |        43         49 |        92
-----------+----------------------+----------
     Total |        87        114 |       201

          Pearson chi2(1) =   0.8252   Pr = 0.364

.
. tab DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC  if DID_YOU_PARTICIPATE_IN_THE_SOYB==0

Do you have |
  access to |
    credit |
 facilities |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         44       50.57       50.57
        Yes |         43       49.43      100.00
------------+-----------------------------------
      Total |         87      100.00

.
. tab DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC  if DID_YOU_PARTICIPATE_IN_THE_SOYB==1

Do you have |
  access to |
    credit |
 facilities |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         65       57.02       57.02
        Yes |         49       42.98      100.00
------------+-----------------------------------
      Total |        114      100.00

.
. tab DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC

Do you have |
  access to |
    credit |
 facilities |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |        109       54.23       54.23
        Yes |         92       45.77      100.00
------------+-----------------------------------
      Total |        201      100.00

.
. tab ARE_YOU_A_MEMBER_OF_ANY_AGRICULT DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC, chi2

 Are you a |
 member of |
       any | Do you have access to
agricultur |   credit facilities
```

```
 al group |        No        Yes |     Total
-----------+----------------------+----------
        No |        46         31 |        77
       Yes |        63         61 |       124
-----------+----------------------+----------
     Total |       109         92 |       201

          Pearson chi2(1) =   1.5275   Pr = 0.216

.
. tab ARE_YOU_A_MEMBER_OF_ANY_AGRICULT if DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC==1

  Are you a |
  member of |
       any |
agricultura |
   l group |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         31       33.70       33.70
        Yes |         61       66.30      100.00
------------+-----------------------------------
      Total |         92      100.00

.
. tab ARE_YOU_A_MEMBER_OF_ANY_AGRICULT if DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC==0

  Are you a |
  member of |
       any |
agricultura |
   l group |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         46       42.20       42.20
        Yes |         63       57.80      100.00
------------+-----------------------------------
      Total |        109      100.00

.
. tab ARE_YOU_A_MEMBER_OF_ANY_AGRICULT

  Are you a |
  member of |
       any |
agricultura |
   l group |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         77       38.31       38.31
        Yes |        124       61.69      100.00
------------+-----------------------------------
      Total |        201      100.00

.
. ttest EXTENSION_VISITS, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    1.666667    .2973044    2.773071    1.075645    2.257688
     Yes |     114    4.912281    .5422044    5.789159    3.838076    5.986485
---------+--------------------------------------------------------------------
combined |     201    3.507463    .3515509    4.984094    2.814241    4.200685
---------+--------------------------------------------------------------------
    diff |           -3.245614     .673076               -4.572891   -1.918337
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =  -4.8221
Ho: diff = 0                     degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000

.
. ttest DISTANCE_TO_NEAREST_MARKET, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87    34.71264    2.110228    19.6829    30.51765    38.90764
     Yes |     114    16.35965    1.110362   11.85542    14.15982    18.55948
---------+--------------------------------------------------------------------
combined |     201    24.30348     1.27958   18.14118    21.78028    26.82668
---------+--------------------------------------------------------------------
    diff |            18.35299    2.238424               13.93892    22.76707
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t =   8.1991
Ho: diff = 0                     degrees of freedom =      199
```

```
    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000

.
. ttest HCI, by(DID_YOU_PARTICIPATE_IN_THE_SOYBE)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      No |      87           0           0           0           0           0
     Yes |     114    .6832819    .0179107    .1912343    .6477975    .7187662
---------+--------------------------------------------------------------------
combined |     201    .387533    .0259973    .3685755     .336269     .438797
---------+--------------------------------------------------------------------
    diff |          -.6832819    .0205147               -.7237359   -.6428278
------------------------------------------------------------------------------
    diff = mean(No) - mean(Yes)                                   t = -33.3070
Ho: diff = 0                                 degrees of freedom =      199

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000

.
. sum HCI

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         HCI |       201    .387533    .3685755         0          1

.
```

Appendix III: Objective One Output, Double Hurdle Regression Results

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.0   Copyright 1985-2015 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

30-student Stata lab perpetual license:
       Serial number:  401406221642
         Licensed to:  Kenya
                       The Real Africa

Notes:
      1.  Unicode is supported; see help unicode_advice.

. use "C:\Users\user\Desktop\SHIRO_DESCRIPTIVE\Merged 2.dta"

. reg DID_YOU_PARTICIPATE_IN_THE_SOYBE AGE_HHHEAD YEARS_SCHOOL EXPERIENCE_IN_SOYBEAN_PRODUCTION HH_SIZE log
> _otherincome LAND_SOYBEAN EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET

      Source |       SS           df       MS      Number of obs   =       201
-------------+----------------------------------   F(8, 192)       =     28.60
       Model |  26.8277413         8  3.35346766   Prob > F        =    0.0000
    Residual |  22.5155423       192   .11726845   R-squared       =    0.5437
-------------+----------------------------------   Adj R-squared   =    0.5247
       Total |  49.3432836       200  .246716418   Root MSE        =    .34244

------------------------------------------------------------------------------------------------------
DID_YOU_PARTICIPATE_IN_THE_SOYBE |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------------------+--------------------------------------------------------------------
                    AGE_HHHEAD |  -.0086304   .0021818    -3.96   0.000    -.0129338   -.0043271
                  YEARS_SCHOOL |   .0250328   .0071989     3.48   0.001     .0108337    .0392318
EXPERIENCE_IN_SOYBEAN_PRODUCTION |   .0024022   .0041131     0.58   0.560    -.0057104    .0105148
                     HH_SIZE |  -.0546572   .0093287    -5.86   0.000    -.0730571   -.0362573
                log_otherincome |  -.0383445    .020156    -1.90   0.059    -.0781001    .0014111
                  LAND_SOYBEAN |   .2636848   .0971912     2.71   0.007     .0719852    .4553845
                EXTENSION_VISITS |    .022385   .0049986     4.48   0.000     .0125258    .0322442
      DISTANCE_TO_NEAREST_MARKET |  -.0080181   .0015001    -5.34   0.000    -.0109769   -.0050592
                           _cons |   1.535015   .2801206     5.48   0.000     .9825057    2.087524
------------------------------------------------------------------------------------------------------

.
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
DISTANCE_T~T |      1.26    0.791704
YEARS_SCHOOL |      1.26    0.792203
log_otheri~e |      1.19    0.839607
  AGE_HHHEAD |      1.18    0.844201
     HH_SIZE |      1.10    0.908387
LAND_SOYBEAN |      1.08    0.921840
EXTENSION_~S |      1.06    0.944682
EXPERIENCE~N |      1.06    0.945102
-------------+----------------------
    Mean VIF |      1.15

.
. *****testing heteroskedasticity*****

.
. estat imtest

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |      68.59     44    0.0103
            Skewness |      19.46      8    0.0126
            Kurtosis |       3.15      1    0.0761
---------------------+-----------------------------
               Total |      91.20     53    0.0009
---------------------------------------------------

.
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of DID_YOU_PARTICIPATE_IN_THE_SOYBE
```

```
        chi2(1)      =    2.59
        Prob > chi2  =   0.1077

.
. ****for categorical variables****

.
. *****pairwise correlation***

.
. pwcorr SEX_HHHEAD M_status pri_occupation soil_fertility tenure DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC ARE_YOU_
> A_MEMBER_OF_ANY_AGRICULT

             | SEX_HH~D M_status pri_oc~n soil_f~y   tenure DO_YOU~C ARE_YO~T
-------------+------------------------------------------------------------------
  SEX_HHHEAD |   1.0000
    M_status |   0.5703   1.0000
pri_occupa~n |   0.1421   0.1889   1.0000
soil_ferti~y |  -0.0721  -0.0025  -0.0083   1.0000
      tenure |  -0.0537   0.1348   0.0807  -0.0055   1.0000
DO_YOU_HAV~C |   0.1056  -0.0944   0.1534  -0.0079  -0.0454   1.0000
ARE_YOU_A_~T |   0.0040   0.0867  -0.0187  -0.0556  -0.0354   0.0872   1.0000

.
. **************DOUBLE HURDLE MODEL_craggit estimation*****************************

.
. xi: craggit DID_YOU_PARTICIPATE_IN_THE_SOYBE SEX_HHHEAD M_status pri_occupation AGE_HHHEAD YEARS_SCHOOL E
> XPERIENCE_IN_SOYBEAN_PRODUCTION  HH_SIZE log_otherincome LAND_SOYBEAN soil_fertility tenure DO_YOU_HAVE_A
> CCESS_TO_CREDIT_FAC ARE_YOU_A_MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET, second(
> HCI SEX_HHHEAD M_status pri_occupation AGE_HHHEAD YEARS_SCHOOL EXPERIENCE_IN_SOYBEAN_PRODUCTION  HH_SIZE
> log_otherincome LAND_SOYBEAN soil_fertility tenure DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC ARE_YOU_A_MEMBER_OF_A
> NY_AGRICULT EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET) vce(robust)

Estimating Cragg's tobit alternative
Assumes conditional independence

initial:       log pseudolikelihood =     -<inf>  (could not be evaluated)
feasible:      log pseudolikelihood = -166.33447
rescale:       log pseudolikelihood = -166.33447
rescale eq:    log pseudolikelihood = -145.42113
Iteration 0:   log pseudolikelihood = -145.42113  (not concave)
Iteration 1:   log pseudolikelihood = -104.48308
Iteration 2:   log pseudolikelihood = -95.985006
Iteration 3:   log pseudolikelihood = -28.902787
Iteration 4:   log pseudolikelihood = -22.441073
Iteration 5:   log pseudolikelihood = -22.167009
Iteration 6:   log pseudolikelihood =  -22.16402
Iteration 7:   log pseudolikelihood =  -22.16402

                                      Number of obs     =        201
                                      Wald chi2(15)     =      74.12
Log pseudolikelihood =  -22.16402     Prob > chi2       =     0.0000

------------------------------------------------------------------------------------------
                         |               Robust
                         |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------------+----------------------------------------------------------------
Tier1                    |
              SEX_HHHEAD |  -.1642665   .4174136    -0.39   0.694    -.982382    .6538491
                M_status |   .0528499   .4951259     0.11   0.915    -.917579    1.023279
           pri_occupation |  -.6381654   .3554908    -1.80   0.073   -1.334915    .0585839
              AGE_HHHEAD |  -.0547991   .0139157    -3.94   0.000   -.0820734   -.0275248
            YEARS_SCHOOL |   .1215269   .0359264     3.38   0.001    .0511125    .1919413
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.0167978   .0240633    -0.70   0.485    -.063961    .0303654
                 HH_SIZE |  -.2772298   .0620617    -4.47   0.000   -.3988686   -.1555911
           log_otherincome |  -.1828624   .1006484    -1.82   0.069   -.3801297    .014405
              LAND_SOYBEAN |   2.058894   .7584139     2.71   0.007     .57243    3.545358
           soil_fertility |   .5507368   .4352093     1.27   0.206   -.3022578    1.403731
                   tenure |  -1.28881   .4675775    -2.76   0.006   -2.205245   -.3723753
DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC |   .0493087   .2758088     0.18   0.858   -.4912666    .5898839
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.230388   .2837373    -0.81   0.417    -.786503    .3257269
          EXTENSION_VISITS |   .1538349   .0431443     3.57   0.000    .0692736    .2383963
   DISTANCE_TO_NEAREST_MARKET |  -.0402895   .0091044    -4.43   0.000   -.0581337   -.0224453
                    _cons |   7.36539   1.638108     4.50   0.000    4.154757    10.57602
-------------------------+----------------------------------------------------------------
Tier2                    |
              SEX_HHHEAD |   .0536632   .0849581     0.63   0.528   -.1139422    .2212685
                M_status |  -.0944726   .0998796    -0.95   0.345    -.291515    .1025698
           pri_occupation |   -.173748   .0747346    -2.32   0.021   -.3211843   -.0263117
              AGE_HHHEAD |   -.010439   .0028386    -3.68   0.000    -.016039   -.004839
            YEARS_SCHOOL |   .0291057    .009169     3.17   0.002    .0110172    .0471942
EXPERIENCE_IN_SOYBEAN_PRODUCTION |   -.006229   .0041183    -1.51   0.132   -.0143536    .0018956
                 HH_SIZE |  -.0813581   .0142049    -5.73   0.000   -.1093816   -.0533347
           log_otherincome |  -.0395246    .026781    -1.48   0.142   -.0923582    .0133089
              LAND_SOYBEAN |   .3596352   .1208225     2.98   0.003    .1212765    .5979939
```

```
              soil_fertility |   .1420508    .111385     1.28   0.204    -.0776896    .3617912
                  tenure |  -.1416463   .1418894    -1.00   0.319    -.4215657    .138273
DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC |   .0307122   .0661323     0.46   0.643    -.0997537    .161178
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.0035604   .0686234    -0.05   0.959    -.1389407    .1318199
          EXTENSION_VISITS |   .0230229   .0067774     3.40   0.001     .0096524    .0363934
    DISTANCE_TO_NEAREST_MARKET |  -.0114235   .0026752    -4.27   0.000    -.0167012   -.0061458
                  _cons |   1.851021   .3942377     4.70   0.000     1.073269    2.628774
-------------------------------+----------------------------------------------------------------
sigma                          |
                  _cons |   .1785016   .0127126    14.04   0.000     .1535854    .2034179
------------------------------------------------------------------------------------------------

.
. predict commercialization, eq(Tier1)

.
. predict commercialization_extent, eq(Tier2)

.
. predict cons, eq(sigma)

.
. generate commercialization0= 1 - normal( commercialization )

.
. generate commercialization1= normal( commercialization )

.
. xi:  tobit HCI SEX_HHHEAD M_status pri_occupation AGE_HHHEAD YEARS_SCHOOL EXPERIENCE_IN_SOYBEAN_PRODUCTIO
> N  HH_SIZE log_otherincome LAND_SOYBEAN soil_fertility tenure DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC ARE_YOU_A_
> MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET, ll(0) vce(robust)

Tobit regression                                Number of obs     =        201
                                                F(  15,    186)   =      12.31
                                                Prob > F          =     0.0000
Log pseudolikelihood = -95.390606               Pseudo R2         =     0.4576

------------------------------------------------------------------------------------------------
                               |               Robust
                           HCI |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------------------+----------------------------------------------------------------
                    SEX_HHHEAD |   .0536632   .0849581     0.63   0.528    -.1139422    .2212685
                      M_status |  -.0944726   .0998796    -0.95   0.345    -.291515     .1025698
                  pri_occupation |  -.173748   .0747346    -2.32   0.021    -.3211843   -.0263117
                     AGE_HHHEAD |  -.010439   .0028386    -3.68   0.000    -.016039    -.004839
                   YEARS_SCHOOL |   .0291057   .009169     3.17   0.002     .0110172    .0471942
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.006229   .0041183    -1.51   0.132    -.0143536    .0018956
                        HH_SIZE |  -.0813581   .0142049    -5.73   0.000    -.1093816   -.0533347
                log_otherincome |  -.0395246   .026781    -1.48   0.142    -.0923582    .0133089
                  LAND_SOYBEAN |   .3596352   .1208225     2.98   0.003     .1212765    .5979939
                soil_fertility |   .1420508    .111385     1.28   0.204    -.0776896    .3617912
                      tenure |  -.1416463   .1418894    -1.00   0.319    -.4215657    .138273
DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC |   .0307122   .0661323     0.46   0.643    -.0997537    .161178
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.0035604   .0686234    -0.05   0.959    -.1389407    .1318199
              EXTENSION_VISITS |   .0230229   .0067774     3.40   0.001     .0096524    .0363934
    DISTANCE_TO_NEAREST_MARKET |  -.0114235   .0026752    -4.27   0.000    -.0167012   -.0061458
                      _cons |   1.851021   .3942377     4.70   0.000     1.073269    2.628774
-------------------------------+----------------------------------------------------------------
                      /sigma |   .3787441   .0258926                      .3276632    .4298249
------------------------------------------------------------------------------------------------
          87  left-censored observations at HCI <= 0
         114     uncensored observations
           0 right-censored observations

.
. estimates store tobit

.
. xi:probit DID_YOU_PARTICIPATE_IN_THE_SOYBE SEX_HHHEAD M_status pri_occupation AGE_HHHEAD YEARS_SCHOOL EXP
> ERIENCE_IN_SOYBEAN_PRODUCTION  HH_SIZE log_otherincome LAND_SOYBEAN soil_fertility tenure DO_YOU_HAVE_ACC
> ESS_TO_CREDIT_FAC ARE_YOU_A_MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET, vce(robus
> t)

Iteration 0:   log pseudolikelihood = -137.50366
Iteration 1:   log pseudolikelihood =  -58.83294
Iteration 2:   log pseudolikelihood = -56.994833
Iteration 3:   log pseudolikelihood = -56.973444
Iteration 4:   log pseudolikelihood = -56.973433
Iteration 5:   log pseudolikelihood = -56.973433

Probit regression                               Number of obs     =        201
                                                Wald chi2(15)     =      74.12
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -56.973433               Pseudo R2         =     0.5857

------------------------------------------------------------------------------------------------
                               |               Robust
```

```
DID_YOU_PARTICIPATE_IN_THE_SOYBE |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------------+------------------------------------------------------------------
                SEX_HHHEAD |  -.1642665   .4174136    -0.39   0.694     -.982382    .6538491
                  M_status |   .0528499   .4951258     0.11   0.915    -.917579    1.023279
             pri_occupation |  -.6381654   .3554908    -1.80   0.073    -1.334915    .0585838
                AGE_HHHEAD |  -.0547991   .0139157    -3.94   0.000    -.0820734   -.0275248
              YEARS_SCHOOL |   .1215269   .0359264     3.38   0.001     .0511125    .1919413
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.0167978   .0240633    -0.70   0.485    -.063961    .0303654
                  HH_SIZE |  -.2772298   .0620617    -4.47   0.000    -.3988686   -.1555911
            log_otherincome |  -.1828624   .1006484    -1.82   0.069    -.3801297    .014405
              LAND_SOYBEAN |   2.058894   .7584139     2.71   0.007      .57243    3.545358
             soil_fertility |   .5507368   .4352093     1.27   0.206    -.3022578    1.403731
                    tenure |  -1.28881   .4675775    -2.76   0.006    -2.205245   -.3723753
DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC |   .0493087   .2758088     0.18   0.858    -.4912666    .5898839
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.230388   .2837373    -0.81   0.417    -.786503    .3257269
          EXTENSION_VISITS |   .1538349   .0431443     3.57   0.000     .0692736    .2383963
  DISTANCE_TO_NEAREST_MARKET |  -.0402895   .0091044    -4.43   0.000    -.0581337   -.0224453
                   _cons |    7.36539   1.638108     4.50   0.000     4.154757    10.57602
-------------------------------------------------------------------------------------------------
Note: 0 failures and 2 successes completely determined.

.
. estimates store probit

.
. xi:  truncreg HCI SEX_HHHEAD M_status pri_occupation AGE_HHHEAD YEARS_SCHOOL EXPERIENCE_IN_SOYBEAN_PRODUC
> TION  HH_SIZE log_otherincome LAND_SOYBEAN soil_fertility tenure DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC ARE_YOU
> _A_MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET, ll(0) vce(robust)
(note: 87 obs. truncated)

Fitting full model:

Iteration 0:   log pseudolikelihood =  34.214247
Iteration 1:   log pseudolikelihood =  34.774095
Iteration 2:   log pseudolikelihood =  34.809345
Iteration 3:   log pseudolikelihood =  34.809413
Iteration 4:   log pseudolikelihood =  34.809413

Truncated regression
Limit:       lower =          0              Number of obs     =        114
             upper =       +inf              Wald chi2(15)     =      27.97
Log pseudolikelihood =  34.809413              Prob > chi2       =     0.0217

-------------------------------------------------------------------------------------------------
                                |               Robust
                            HCI |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------------------+----------------------------------------------------------------
                SEX_HHHEAD |   .0681718   .0444744     1.53   0.125    -.0189965     .15534
                  M_status |  -.0673683   .0487126    -1.38   0.167    -.1628432    .0281066
             pri_occupation |  -.0250578   .0413268    -0.61   0.544    -.1060569    .0559412
                AGE_HHHEAD |  -.0002773   .0017189    -0.16   0.872    -.0036463    .0030916
              YEARS_SCHOOL |  -.0038641   .0066945    -0.58   0.564    -.016985    .0092568
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.0047277   .0024999    -1.89   0.059    -.0096273     .000172
                  HH_SIZE |  -.0190173   .0110564    -1.72   0.085    -.0406876    .0026529
            log_otherincome |  -.0111379   .0149978    -0.74   0.458    -.0405329    .0182572
              LAND_SOYBEAN |   .101686   .0506296     2.01   0.045     .0024539    .2009182
             soil_fertility |   .038802   .0902235     0.43   0.667    -.1380327    .2156368
                    tenure |   .1274416   .0942585     1.35   0.176    -.0573017    .312185
DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC |  -.0097017   .0402124    -0.24   0.809    -.0885165    .0691132
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |   .023415   .0401645     0.58   0.560    -.0553059    .1021359
          EXTENSION_VISITS |   .0017498   .0039679     0.44   0.659    -.0060271    .0095268
  DISTANCE_TO_NEAREST_MARKET |  -.0012256   .0015843    -0.77   0.439    -.0043308    .0018795
                   _cons |   .8723605   .2825865     3.09   0.002     .3185011    1.42622
--------------------------------+----------------------------------------------------------------
                     /sigma |   .1785016   .0127369    14.01   0.000     .1535377    .2034655
-------------------------------------------------------------------------------------------------

.
. estimates store truncreg

.
. lrtest ( tobit ) ( probit truncreg ), stats dir force

Likelihood-ratio test                                 LR chi2(16) =    146.45
                                                  Prob > chi2 =    0.0000

Assumption: (tobit) nested in (probit, truncreg)

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |        Obs  ll(null)  ll(model)      df         AIC        BIC
-------------+---------------------------------------------------------------
       tobit |        201 -175.8514  -95.39061      17    224.7812   280.9374
      probit |        201 -137.5037  -56.97343      16    145.9469   198.7997
   truncreg |        114         .   34.80941      17   -35.61883   10.89655
-----------------------------------------------------------------------------
```

```
Note: N=Obs used in calculating BIC; see [R] BIC note.

-------------------------------------------------------------------------------
       name | command     depvar       npar  title
------------+------------------------------------------------------------------
      tobit | tobit       HCI            17
     probit | probit      DID_YOU_PA~E   16
  truncreg | truncreg    HCI            17
```

.

Appendix IV: Objective Two Output, Multinomial Logistic Regression Results

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.0   Copyright 1985-2015 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

30-student Stata lab perpetual license:
       Serial number:  401406221642
         Licensed to:  Kenya
                       The Real Africa

Notes:
      1.  Unicode is supported; see help unicode_advice.

. use "C:\Users\user\Desktop\SHIRO_DESCRIPTIVE\Objective two\Merged 2.dta"

. tab MARKET_OUTLET_USED_IN_THE_LAST_S if DID_YOU_PARTICIPATE_IN_THE_SOYBE==1

  Market outlet used in the last |
                          season |      Freq.     Percent        Cum.
---------------------------------+-----------------------------------
                          Direct |         70       61.40       61.40
                       Middlemen |         33       28.95       90.35
                   Cooperatives |          9        7.89       98.25
                  Agro-processors |          2        1.75      100.00
---------------------------------+-----------------------------------
                           Total |        114      100.00

.
. tab DID_YOU_PARTICIPATE_IN_THE_SOYBE

    Did you |
participate |
     in the |
    soybean |
market last |
     season |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         87       43.28       43.28
        Yes |        114       56.72      100.00
------------+-----------------------------------
      Total |        201      100.00

.
. drop if MARKET_OUTLET_USED_IN_THE_LAST_S==4
(2 observations deleted)

. drop if  DID_YOU_PARTICIPATE_IN_THE_SOYBE==0
(87 observations deleted)

.
.
. tab PRICE_PER_UNIT

  Price per |
       Unit |      Freq.     Percent        Cum.
------------+-----------------------------------
         16 |          1        0.89        0.89
```

```
        20 |       2       1.79        2.68
        25 |       1       0.89        3.57
        30 |       4       3.57        7.14
        35 |       1       0.89        8.04
        40 |       9       8.04       16.07
  49.44279 |       2       1.79       17.86
        50 |      18      16.07       33.93
        55 |       2       1.79       35.71
        60 |       9       8.04       43.75
        62 |       1       0.89       44.64
        70 |       7       6.25       50.89
        75 |       9       8.04       58.93
        80 |       7       6.25       65.18
        90 |       4       3.57       68.75
       100 |      19      16.96       85.71
       125 |       3       2.68       88.39
       140 |       1       0.89       89.29
       150 |       6       5.36       94.64
       175 |       2       1.79       96.43
       200 |       2       1.79       98.21
       250 |       1       0.89       99.11
       750 |       1       0.89      100.00
-----------+-----------------------------------
     Total |     112      100.00

. ed PRICE_PER_UNIT

. tab EXTENSION_VISITS

  Number of |
      times |
    visited |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |         27       24.11       24.11
          1 |          4        3.57       27.68
          2 |         12       10.71       38.39
          3 |         17       15.18       53.57
          4 |         13       11.61       65.18
          5 |          9        8.04       73.21
          6 |          3        2.68       75.89
          7 |          1        0.89       76.79
          8 |          9        8.04       84.82
          9 |          1        0.89       85.71
         10 |          4        3.57       89.29
         12 |          5        4.46       93.75
         20 |          3        2.68       96.43
         23 |          1        0.89       97.32
         24 |          1        0.89       98.21
         25 |          1        0.89       99.11
         30 |          1        0.89      100.00
------------+-----------------------------------
      Total |        112      100.00

. oneway AGE_HHHEAD  MARKET_OUTLET_USED_IN_THE_LAST_S, tabulate

     Market |
outlet used |
in the last |  Summary of Age of household head
     season |        Mean   Std. Dev.       Freq.
------------+------------------------------------
     Direct |   51.114286   11.404986          70
  Middlemen |   48.757576   13.355781          33
  Cooperati |   48.777778    11.72367           9
------------+------------------------------------
      Total |   50.232143   11.978947         112

                    Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups       145.26241      2   72.6312049      0.50     0.6069
 Within groups      15782.7019    109    144.79543
------------------------------------------------------------------------
    Total           15927.9643    111   143.495174

Bartlett's test for equal variances:  chi2(2) =   1.1107  Prob>chi2 = 0.574

. oneway HH_SIZE  MARKET_OUTLET_USED_IN_THE_LAST_S, tabulate

     Market |
outlet used |     Summary of Total number of
in the last |          household members
     season |        Mean   Std. Dev.       Freq.
------------+------------------------------------
     Direct |   4.9142857   1.9541324          70
  Middlemen |   4.2727273   1.5865199          33
  Cooperati |   5.2222222   1.5634719           9
------------+------------------------------------
```

```
     Total |       4.75   1.8380366        112

                    Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------------
Between groups      11.4132756      2   5.70663781      1.71     0.1855
 Within groups      363.586724    109   3.33565802
------------------------------------------------------------------------------
    Total               375    111   3.37837838

Bartlett's test for equal variances:  chi2(2) =   2.1103  Prob>chi2 = 0.348

. oneway YEARS_SCHOOL  MARKET_OUTLET_USED_IN_THE_LAST_S, tabulate

     Market |
outlet used |    Summary of Number of years in
in the last | school of HH head (complete years)
    season |        Mean   Std. Dev.       Freq.
------------+------------------------------------
    Direct |   9.9428571   3.2699143          70
 Middlemen |   10.848485   2.3864358          33
 Cooperati |   11.777778     3.34581           9
------------+------------------------------------
     Total |   10.357143    3.072274         112

                    Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------------
Between groups      38.1448773      2   19.0724387      2.06     0.1325
 Within groups      1009.56941    109   9.26210466
------------------------------------------------------------------------------
    Total          1047.71429    111   9.43886744

Bartlett's test for equal variances:  chi2(2) =   4.0300  Prob>chi2 = 0.133

. oneway EXPERIENCE_IN_SOYBEAN_PRODUCTION MARKET_OUTLET_USED_IN_THE_LAST_S, tabulate

     Market |
outlet used |  Summary of Experience in Soybean
in the last |             production
    season |        Mean   Std. Dev.       Freq.
------------+------------------------------------
    Direct |   7.6285714   7.0816592          70
 Middlemen |    5.969697   6.3022062          33
 Cooperati |   7.4444444   6.2472216           9
------------+------------------------------------
     Total |       7.125    6.782496         112

                    Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------------
Between groups      62.7152237      2   31.3576118      0.68     0.5099
 Within groups      5043.53478    109   46.2709613
------------------------------------------------------------------------------
    Total            5106.25    111   46.0022523

Bartlett's test for equal variances:  chi2(2) =   0.6760  Prob>chi2 = 0.713

. oneway log_otherincome MARKET_OUTLET_USED_IN_THE_LAST_S, tabulate

     Market |   Summary of Natural logarithm of
outlet used |     other household income less
in the last |              remittance
    season |        Mean   Std. Dev.       Freq.
------------+------------------------------------
    Direct |   10.417753   1.1747661          70
 Middlemen |   9.9240938   1.1823885          33
 Cooperati |   10.453635   1.3446635           9
------------+------------------------------------
     Total |   10.275183    1.201368         112

                    Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------------
Between groups      5.77713233      2   2.88856616      2.04     0.1351
 Within groups      154.427513    109   1.41676618
------------------------------------------------------------------------------
    Total          160.204646    111    1.4432851

Bartlett's test for equal variances:  chi2(2) =   0.2775  Prob>chi2 = 0.870

. oneway TOTAL_YIELD_LAST_SEASON MARKET_OUTLET_USED_IN_THE_LAST_S, tabulate

     Market |
outlet used |
in the last | Summary of Total yield last season
    season |        Mean   Std. Dev.       Freq.
```

```
------------+-----------------------------------
    Direct | 78.514286  67.577651         70
 Middlemen |  81.212121  81.848472         33
 Cooperati |  121.77778  139.31868          9
------------+-----------------------------------
     Total | 82.785714  79.400817        112

                    Analysis of Variance
    Source              SS         df      MS           F     Prob > F
------------------------------------------------------------------------
Between groups      15042.3007      2   7521.15036      1.20     0.3060
 Within groups      684756.056    109   6282.16566
------------------------------------------------------------------------
    Total           699798.357    111    6304.4897

Bartlett's test for equal variances:  chi2(2) =  10.6376  Prob>chi2 = 0.005

. oneway EXTENSION_VISITS MARKET_OUTLET_USED_IN_THE_LAST_S, tabulate

     Market |
outlet used |
in the last | Summary of Number of times visited
    season |        Mean   Std. Dev.       Freq.
------------+-----------------------------------
    Direct |  4.7571429   6.1063894         70
 Middlemen |  3.8787879   3.0285885         33
 Cooperati |  9.2222222   9.2840963          9
------------+-----------------------------------
     Total |  4.8571429   5.8199055        112

                    Analysis of Variance
    Source              SS         df      MS           F     Prob > F
------------------------------------------------------------------------
Between groups       203.77215      2   101.886075      3.12     0.0480
 Within groups      3555.94214    109   32.6233223
------------------------------------------------------------------------
    Total           3759.71429    111   33.8712999

Bartlett's test for equal variances:  chi2(2) =  22.9896  Prob>chi2 = 0.000

. oneway DISTANCE_TO_NEAREST_MARKET MARKET_OUTLET_USED_IN_THE_LAST_S, tabulate

     Market |
outlet used |   Summary of Distance to nearest
in the last |     market (in walking minutes)
    season |        Mean   Std. Dev.       Freq.
------------+-----------------------------------
    Direct |  16.257143   12.347253         70
 Middlemen |  16.454545   10.473451         33
 Cooperati |  19.666667   13.313527          9
------------+-----------------------------------
     Total |  16.589286   11.832581        112

                    Analysis of Variance
    Source              SS         df      MS           F     Prob > F
------------------------------------------------------------------------
Between groups      93.5538961      2   46.7769481      0.33     0.7196
 Within groups      15447.5532    109   141.720672
------------------------------------------------------------------------
    Total           15541.1071    111   140.009974

Bartlett's test for equal variances:  chi2(2) =   1.3343  Prob>chi2 = 0.513

. oneway PRICE_PER_UNIT MARKET_OUTLET_USED_IN_THE_LAST_S, tabulate

     Market |
outlet used |
in the last |      Summary of Price per Unit
    season |        Mean   Std. Dev.       Freq.
------------+-----------------------------------
    Direct |  84.055508   88.292267         70
 Middlemen |   82.30303   48.836772         33
 Cooperati |  95.222222   48.751353          9
------------+-----------------------------------
     Total |  84.436478   75.601675        112

                    Analysis of Variance
    Source              SS         df      MS           F     Prob > F
------------------------------------------------------------------------
Between groups      1207.35293      2   603.676465      0.10     0.9014
 Within groups      633225.711    109   5809.41019
------------------------------------------------------------------------
    Total           634433.064    111   5715.61319

Bartlett's test for equal variances:  chi2(2) =  14.9454  Prob>chi2 = 0.001

. tab SEX_HHHEAD chi2
```

```
variable chi2 not found
r(111);

. tab SEX_HHHEAD, chi2
option chi2 not allowed
r(198);

. tab SEX_HHHEAD MARKET_OUTLET_USED_IN_THE_LAST_S, chi2

 Gender of |  Market outlet used in the last
 household |              season
     head |    Direct  Middlemen  Cooperati |     Total
-----------+---------------------------------+----------
    Female |       27         13         3 |        43
      Male |       43         20         6 |        69
-----------+---------------------------------+----------
     Total |       70         33         9 |       112

          Pearson chi2(2) =   0.1123   Pr = 0.945

. tab SEX_HHHEAD if MARKET_OUTLET_USED_IN_THE_LAST_S ==1

 Gender of |
 household |
     head |      Freq.     Percent        Cum.
-----------+-----------------------------------
    Female |         27       38.57       38.57
      Male |         43       61.43      100.00
-----------+-----------------------------------
     Total |         70      100.00

. tab SEX_HHHEAD if MARKET_OUTLET_USED_IN_THE_LAST_S ==2

 Gender of |
 household |
     head |      Freq.     Percent        Cum.
-----------+-----------------------------------
    Female |         13       39.39       39.39
      Male |         20       60.61      100.00
-----------+-----------------------------------
     Total |         33      100.00

. tab SEX_HHHEAD if MARKET_OUTLET_USED_IN_THE_LAST_S ==3

 Gender of |
 household |
     head |      Freq.     Percent        Cum.
-----------+-----------------------------------
    Female |          3       33.33       33.33
      Male |          6       66.67      100.00
-----------+-----------------------------------
     Total |          9      100.00

. tab SEX_HHHEAD

 Gender of |
 household |
     head |      Freq.     Percent        Cum.
-----------+-----------------------------------
    Female |         43       38.39       38.39
      Male |         69       61.61      100.00
-----------+-----------------------------------
     Total |        112      100.00

. tab M_status MARKET_OUTLET_USED_IN_THE_LAST_S, chi2

   Marital |
 status of |  Market outlet used in the last
 household |              season
     head |    Direct  Middlemen  Cooperati |     Total
-----------+---------------------------------+----------
 otherwise |       18         11         3 |        32
   Married |       52         22         6 |        80
-----------+---------------------------------+----------
     Total |       70         33         9 |       112

          Pearson chi2(2) =   0.7467   Pr = 0.688

.
. tab M_status if MARKET_OUTLET_USED_IN_THE_LAST_S ==1

    Marital |
  status of |
  household |
      head |      Freq.     Percent        Cum.
-----------+-----------------------------------
  otherwise |         18       25.71       25.71
```

```
   Married |         52       74.29      100.00
-----------+-----------------------------------
     Total |         70      100.00

.
. tab M_status if MARKET_OUTLET_USED_IN_THE_LAST_S ==2

   Marital |
  status of |
  household |
      head |      Freq.     Percent        Cum.
-----------+-----------------------------------
 otherwise |         11       33.33       33.33
   Married |         22       66.67      100.00
-----------+-----------------------------------
     Total |         33      100.00

.
. tab M_status if MARKET_OUTLET_USED_IN_THE_LAST_S ==3

   Marital |
  status of |
  household |
      head |      Freq.     Percent        Cum.
-----------+-----------------------------------
 otherwise |          3       33.33       33.33
   Married |          6       66.67      100.00
-----------+-----------------------------------
     Total |          9      100.00

.
. tab M_status

   Marital |
  status of |
  household |
      head |      Freq.     Percent        Cum.
-----------+-----------------------------------
 otherwise |         32       28.57       28.57
   Married |         80       71.43      100.00
-----------+-----------------------------------
     Total |        112      100.00

. tab pri_occupation  MARKET_OUTLET_USED_IN_THE_LAST_S, chi2

                    | Market outlet used in the last
                    |             season
      pri_occupation |    Direct  Middlemen  Cooperati |     Total
---------------------+---------------------------------+----------
Farming(crop/livestoc |        57        26         7 |        90
              Others |        13         7         2 |        22
---------------------+---------------------------------+----------
               Total |        70        33         9 |       112

          Pearson chi2(2) =   0.1403   Pr = 0.932

.
. tab pri_occupation  if MARKET_OUTLET_USED_IN_THE_LAST_S ==1

        pri_occupation |      Freq.     Percent        Cum.
-----------------------+-----------------------------------
Farming(crop/livestock) |         57       81.43       81.43
                Others |         13       18.57      100.00
-----------------------+-----------------------------------
                 Total |         70      100.00

.
. tab pri_occupation  if MARKET_OUTLET_USED_IN_THE_LAST_S ==2

        pri_occupation |      Freq.     Percent        Cum.
-----------------------+-----------------------------------
Farming(crop/livestock) |         26       78.79       78.79
                Others |          7       21.21      100.00
-----------------------+-----------------------------------
                 Total |         33      100.00

.
. tab pri_occupation  if MARKET_OUTLET_USED_IN_THE_LAST_S ==3

        pri_occupation |      Freq.     Percent        Cum.
-----------------------+-----------------------------------
Farming(crop/livestock) |          7       77.78       77.78
                Others |          2       22.22      100.00
-----------------------+-----------------------------------
                 Total |          9      100.00

.
```

```
. tab pri_occupation

        pri_occupation |      Freq.     Percent        Cum.
-----------------------+-----------------------------------
Farming(crop/livestock) |         90       80.36       80.36
                Others |         22       19.64      100.00
-----------------------+-----------------------------------
                 Total |        112      100.00

. tab DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC MARKET_OUTLET_USED_IN_THE_LAST_S, chi2

    Do you |
     have |
 access to |  Market outlet used in the last
   credit |              season
facilities |    Direct  Middlemen  Cooperati |     Total
-----------+---------------------------------+----------
        No |       47        15         2 |        64
       Yes |       23        18         7 |        48
-----------+---------------------------------+----------
     Total |       70        33         9 |       112

          Pearson chi2(2) =   9.1807   Pr = 0.010

.
. tab DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC if MARKET_OUTLET_USED_IN_THE_LAST_S ==1

Do you have |
  access to |
    credit |
 facilities |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         47       67.14       67.14
        Yes |         23       32.86      100.00
------------+-----------------------------------
      Total |         70      100.00

.
. tab DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC if MARKET_OUTLET_USED_IN_THE_LAST_S ==2

Do you have |
  access to |
    credit |
 facilities |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         15       45.45       45.45
        Yes |         18       54.55      100.00
------------+-----------------------------------
      Total |         33      100.00

.
. tab DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC if MARKET_OUTLET_USED_IN_THE_LAST_S ==3

Do you have |
  access to |
    credit |
 facilities |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |          2       22.22       22.22
        Yes |          7       77.78      100.00
------------+-----------------------------------
      Total |          9      100.00

.
. tab DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC

Do you have |
  access to |
    credit |
 facilities |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         64       57.14       57.14
        Yes |         48       42.86      100.00
------------+-----------------------------------
      Total |        112      100.00

. tab ARE_YOU_A_MEMBER_OF_ANY_AGRICULT MARKET_OUTLET_USED_IN_THE_LAST_S, chi2

 Are you a |
 member of |
       any |  Market outlet used in the last
agricultur |              season
 al group |    Direct  Middlemen  Cooperati |     Total
-----------+---------------------------------+----------
        No |       24        14         2 |        40
       Yes |       46        19         7 |        72
-----------+---------------------------------+----------
```

```
    Total |       70        33         9 |      112

          Pearson chi2(2) =   1.4229   Pr = 0.491

.
. tab ARE_YOU_A_MEMBER_OF_ANY_AGRICULT if MARKET_OUTLET_USED_IN_THE_LAST_S ==1

  Are you a |
  member of |
      any |
agricultura |
   l group |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         24       34.29       34.29
        Yes |         46       65.71      100.00
------------+-----------------------------------
      Total |         70      100.00

.
. tab ARE_YOU_A_MEMBER_OF_ANY_AGRICULT if MARKET_OUTLET_USED_IN_THE_LAST_S ==2

  Are you a |
  member of |
      any |
agricultura |
   l group |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         14       42.42       42.42
        Yes |         19       57.58      100.00
------------+-----------------------------------
      Total |         33      100.00

.
. tab ARE_YOU_A_MEMBER_OF_ANY_AGRICULT if MARKET_OUTLET_USED_IN_THE_LAST_S ==3

  Are you a |
  member of |
      any |
agricultura |
   l group |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |          2       22.22       22.22
        Yes |          7       77.78      100.00
------------+-----------------------------------
      Total |          9      100.00

.
. tab ARE_YOU_A_MEMBER_OF_ANY_AGRICULT

  Are you a |
  member of |
      any |
agricultura |
   l group |      Freq.     Percent        Cum.
------------+-----------------------------------
         No |         40       35.71       35.71
        Yes |         72       64.29      100.00
------------+-----------------------------------
      Total |        112      100.00

. pwcorr SEX_HHHEAD M_status pri_occupation DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC ARE_YOU_A_MEMBER_OF_ANY_AGRICULT

             | SEX_HH~D M_status pri_oc~n DO_YOU~C ARE_YO~T
-------------+---------------------------------------------
  SEX_HHHEAD |   1.0000
    M_status |   0.5167   1.0000
pri_occupa~n |   0.0206   0.1137   1.0000
DO_YOU_HAV~C |   0.2014  -0.0513   0.1622   1.0000
ARE_YOU_A_~T |  -0.0137  -0.0177  -0.0067   0.1183   1.0000

. reg MARKET_OUTLET AGE_HHHEAD HH_SIZE YEARS_SCHOOL EXPERIENCE_IN_SOYBEAN_PRODUCTION log_otherincome TOTAL_YIELD_LA
> ST_SEASON EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET PRICE_PER_UNIT

      Source |       SS           df       MS      Number of obs   =       112
-------------+----------------------------------   F(9, 102)       =      1.48
       Model |  5.28585759         9   .58731751   Prob > F        =    0.1655
    Residual |  40.4909281       102  .396969884   R-squared       =    0.1155
-------------+----------------------------------   Adj R-squared   =    0.0374
       Total |  45.7767857       111  .412403475   Root MSE        =    .63006

------------------------------------------------------------------------------------------
                MARKET_OUTLET |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------------------+-----------------------------------------------------------
                 AGE_HHHEAD |  -.0061721   .0052528    -1.18   0.243    -.0165911    .0042469
                    HH_SIZE |  -.0401031   .0341904    -1.17   0.244    -.1079197    .0277134
               YEARS_SCHOOL |   .0462796   .0206231     2.24   0.027     .0053738    .0871854
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.006611   .0092098    -0.72   0.475    -.0248785    .0116565
```

```
            log_otherincome |   -.090125   .0537921    -1.68   0.097    -.1968213    .0165714
  TOTAL_YIELD_LAST_SEASON |   .0007635   .0007612     1.00   0.318    -.0007464    .0022733
         EXTENSION_VISITS |   .0192276   .0109099     1.76   0.081    -.0024121    .0408673
DISTANCE_TO_NEAREST_MARKET |   .0054621   .0054701     1.00   0.320    -.0053879    .0163121
          PRICE_PER_UNIT |  -.0002731   .0008332    -0.33   0.744    -.0019257    .0013794
                   _cons |   2.225567    .730719     3.05   0.003     .7761896    3.674945
------------------------------------------------------------------------------------------------

. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
DISTANCE_T~T |      1.17    0.853646
log_otheri~e |      1.17    0.856340
EXTENSION_~S |      1.13    0.887079
YEARS_SCHOOL |      1.12    0.890857
PRICE_PER_~T |      1.11    0.901409
 AGE_HHHEAD |      1.11    0.903256
    HH_SIZE |      1.10    0.905560
EXPERIENCE~N |      1.09    0.916554
TOTAL_YIEL~N |      1.02    0.978965
-------------+----------------------
    Mean VIF |      1.11

. linktest

      Source |       SS           df       MS      Number of obs   =       112
-------------+----------------------------------   F(2, 109)       =      7.39
       Model |  5.46495204         2  2.73247602   Prob > F        =    0.0010
    Residual |  40.3118337       109  .369833336   R-squared       =    0.1194
-------------+----------------------------------   Adj R-squared   =    0.1032
       Total |  45.7767857       111  .412403475   Root MSE        =    .60814

------------------------------------------------------------------------------
MARKET_OUT~T |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _hat |  -1.040098   2.943562    -0.35   0.725    -6.874142    4.793946
     _hatsq |   .7077467   1.017043     0.70   0.488      -1.308    2.723493
      _cons |   1.436616   2.100811     0.68   0.496    -2.727124    5.600355
------------------------------------------------------------------------------

. estat imtest

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |      47.97     54    0.7050
            Skewness |      22.43      9    0.0076
            Kurtosis |       0.10      1    0.7540
---------------------+-----------------------------
               Total |      70.49     64    0.2696
---------------------------------------------------

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of MARKET_OUTLET

         chi2(1)      =     6.90
         Prob > chi2  =   0.0086

. ovtest

Ramsey RESET test using powers of the fitted values of MARKET_OUTLET
       Ho:  model has no omitted variables
                 F(3, 99) =      0.65
                 Prob > F =      0.5837

. mlogit MARKET_OUTLET SEX_HHHEAD AGE_HHHEAD M_status HH_SIZE YEARS_SCHOOL pri_occupation EXPERIENCE_IN_SOYBEAN_PRO
> DUCTION log_otherincome TOTAL_YIELD_LAST_SEASON DISTANCE_TO_NEAREST_MARKET PRICE_PER_UNIT EXTENSION_VISITS DO_YOU
> _HAVE_ACCESS_TO_CREDIT_FAC ARE_YOU_A_MEMBER_OF_ANY_AGRICULT, baseoutcome(1) vce(robust)

Iteration 0:   log pseudolikelihood = -95.917436
Iteration 1:   log pseudolikelihood = -79.359639
Iteration 2:   log pseudolikelihood = -76.080171
Iteration 3:   log pseudolikelihood = -75.819696
Iteration 4:   log pseudolikelihood = -75.816277
Iteration 5:   log pseudolikelihood = -75.816277

Multinomial logistic regression                 Number of obs     =        112
                                                Wald chi2(28)     =      65.38
                                                Prob > chi2       =     0.0001
Log pseudolikelihood = -75.816277               Pseudo R2         =     0.2096
```

```
------------------------------------------------------------------------------
                             |               Robust
MARKET_OUTLET |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------------------+------------------------------------------------
Direct                       | (base outcome)
-----------------------------+------------------------------------------------
Middlemen_brockers           |
                SEX_HHHEAD |   .1376774   .6483352     0.21   0.832    -1.133036    1.408391
                AGE_HHHEAD |   -.037525   .0276869    -1.36   0.175    -.0917904    .0167404
                  M_status |  -.5896111   .7969634    -0.74   0.459    -2.151631    .9724085
                   HH_SIZE |  -.3131436   .1404847    -2.23   0.026    -.5884886   -.0377987
              YEARS_SCHOOL |   .1637531   .0863008     1.90   0.058    -.0053933    .3328996
             pri_occupation |  -.1544492   .6437161    -0.24   0.810     -1.41611    1.107211
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.0562006    .043138    -1.30   0.193    -.1407494    .0283483
           log_otherincome |  -.3792906    .243051    -1.56   0.119    -.8556618    .0970806
      TOTAL_YIELD_LAST_SEASON |  -.001062   .0024802    -0.43   0.669    -.0059231    .0037991
   DISTANCE_TO_NEAREST_MARKET |   .0035483   .0218079     0.16   0.871    -.0391945     .046291
            PRICE_PER_UNIT |  -.0040482   .0031548    -1.28   0.199    -.0102314     .002135
          EXTENSION_VISITS |   .0327444   .0482756     0.68   0.498    -.0618741    .1273629
DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC |   1.301448   .5275373     2.47   0.014     .2674943    2.335402
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.4087011   .6284361    -0.65   0.515    -1.640413    .8230111
                    _cons |   6.113676   3.799334     1.61   0.108    -1.332883    13.56023
-----------------------------+------------------------------------------------
Cooperatives                 |
                SEX_HHHEAD |   1.260355   1.356858     0.93   0.353    -1.399038    3.919748
                AGE_HHHEAD |  -.0666637   .0388211    -1.72   0.086    -.1427516    .0094241
                  M_status |  -2.092396   1.233397    -1.70   0.090     -4.50981    .3250174
                   HH_SIZE |   -.151463   .1863811    -0.81   0.416    -.5167633    .2138373
              YEARS_SCHOOL |   .4782951   .2523291     1.90   0.058    -.0162608    .9728509
             pri_occupation |  -.8439988   1.783981    -0.47   0.636    -4.340537     2.65254
EXPERIENCE_IN_SOYBEAN_PRODUCTION |   .0215294    .057272     0.38   0.707    -.0907218    .1337805
           log_otherincome |   .2766107   .5297058     0.52   0.602    -.7615936    1.314815
      TOTAL_YIELD_LAST_SEASON |   .0041834   .0049641     0.84   0.399    -.005546    .0139128
   DISTANCE_TO_NEAREST_MARKET |   .0636646   .0341519     1.86   0.062    -.0032719    .1306011
            PRICE_PER_UNIT |  -.0077704   .0034295    -2.27   0.023    -.014492   -.0010488
          EXTENSION_VISITS |   .1897277   .0560449     3.39   0.001     .0798818    .2995736
DO_YOU_HAVE_ACCESS_TO_CREDIT_FAC |   2.524409   1.191003     2.12   0.034     .1900847    4.858732
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.2161604   .9138163    -0.24   0.813    -2.007207    1.574887
                    _cons |  -6.060363   8.800619    -0.69   0.491    -23.30926    11.18853
------------------------------------------------------------------------------

.
```

Appendix V: Objective Three Output, Endogenous Switching Regression Results

```
___  ____  ____  ____  ____ (R)
/__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.0   Copyright 1985-2015 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

30-student Stata lab perpetual license:
       Serial number:  401406221642
         Licensed to:  Kenya
                       The Real Africa

Notes:
     1.  Unicode is supported; see help unicode_advice.

. use "C:\Users\user\Desktop\SHIRO_DESCRIPTIVE\Objective one\merged 22.dta"

. movestay LNTHINCOME_PER_CAPITA SEX_HHHEAD AGE_HHHEAD YEARS_SCHOOL HH_SIZE log_otherincome ACREAGE_SIZE_IN_ACRES
EXPERI
> ENCE_IN_SOYBEAN_PRODUCTION ARE_YOU_A_MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS,
select(DID_YOU_PARTICIPATE_IN_THE_SOYBE=
>   DISTANCE_TO_NEAREST_MARKET)

Fitting initial values .....
Iteration 0:   log likelihood = -241.72613  (not concave)
Iteration 1:   log likelihood =  -241.5865  (not concave)
Iteration 2:   log likelihood = -241.52362  (not concave)
Iteration 3:   log likelihood = -241.47737  (not concave)
Iteration 4:   log likelihood = -241.43529  (not concave)
Iteration 5:   log likelihood = -241.39234  (not concave)
Iteration 6:   log likelihood = -241.35397  (not concave)
Iteration 7:   log likelihood = -241.31779  (not concave)
Iteration 8:   log likelihood = -241.28181  (not concave)
Iteration 9:   log likelihood = -241.24732  (not concave)
Iteration 10:  log likelihood = -241.21459  (not concave)
Iteration 11:  log likelihood = -241.18312  (not concave)
Iteration 12:  log likelihood = -241.15306  (not concave)
Iteration 13:  log likelihood = -241.12442  (not concave)
Iteration 14:  log likelihood =  -241.0971  (not concave)
Iteration 15:  log likelihood =   -241.071  (not concave)
Iteration 16:  log likelihood = -241.04607  (not concave)
Iteration 17:  log likelihood = -241.02219
Iteration 18:  log likelihood = -240.13772  (backed up)
Iteration 19:  log likelihood =  -239.5887
Iteration 20:  log likelihood = -239.58281
Iteration 21:  log likelihood = -239.58281

Endogenous switching regression model           Number of obs   =        201
                                                Wald chi2(9)    =      89.40
Log likelihood = -239.58281                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------------------------
                               |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------------------+----------------------------------------------------------------
LNTHINCOME_PER_CAPITA_1        |
                  SEX_HHHEAD |  -.0551234    .13957    -0.39   0.693    -.3286755    .2184287
                  AGE_HHHEAD |   -.00538   .0063304    -0.85   0.395    -.0177874    .0070273
                YEARS_SCHOOL |  -.0021744   .0238373    -0.09   0.927    -.0488947    .0445458
                     HH_SIZE |  -.2181993   .0423577    -5.15   0.000    -.3012189   -.1351796
              log_otherincome |   .3947681    .053119     7.43   0.000     .2906567    .4988795
          ACREAGE_SIZE_IN_ACRES |   .0546728   .0419681     1.30   0.193    -.0275833    .1369288
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.0097783   .0090332    -1.08   0.279     -.027483    .0079265
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |   .2923587   .1396156     2.09   0.036     .0187171    .5660003
             EXTENSION_VISITS |  -.0107835   .0136245    -0.79   0.429     -.037487     .01592
                        _cons |   6.457305   .7202823     8.96   0.000     5.045577    7.869032
-------------------------------+----------------------------------------------------------------
LNTHINCOME_PER_CAPITA_0        |
                  SEX_HHHEAD |   .0780587   .1448668     0.54   0.590     -.205875    .3619923
                  AGE_HHHEAD |   .0012752   .0062856     0.20   0.839    -.0110443    .0135947
```

```
                YEARS_SCHOOL |  -.0230144   .0193841    -1.19   0.235    -.0610066    .0149777
                   HH_SIZE |   -.096004   .0244391    -3.93   0.000    -.1439038   -.0481042
              log_otherincome |   .8571316   .0471899    18.16   0.000     .7646412     .949622
        ACREAGE_SIZE_IN_ACRES |  -.0075703   .0453189    -0.17   0.867    -.0963936    .0812531
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.0031794   .0130474    -0.24   0.807    -.0287518    .0223929
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.0687242   .1403035    -0.49   0.624     -.343714    .2062656
            EXTENSION_VISITS |   .0433016   .0239059    -1.81   0.070    -.0901562     .103553
                    _cons |   .2159709   .6740483     0.32   0.749    -1.105139    1.537081
-----------------------------------+----------------------------------------------------------------
DID_YOU_PARTICIPATE_IN_THE_SOYBE |
                SEX_HHHEAD |  -.0982025   .2800235    -0.35   0.726    -.6470385    .4506336
                AGE_HHHEAD |  -.0435483   .0118782    -3.67   0.000    -.0668292   -.0202674
                YEARS_SCHOOL |   .0954721   .0335021     2.85   0.004     .0298093    .1611349
                   HH_SIZE |  -.2226048    .059128    -3.76   0.000    -.3384936    -.106716
              log_otherincome |  -.0226525   .1061463    -0.21   0.831    -.2306955    .1853904
        ACREAGE_SIZE_IN_ACRES |   .0696223   .0765708     0.91   0.363    -.0804536    .2196983
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.0078528   .0213564    -0.37   0.713    -.0497105     .034005
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.0862562   .2722102    -0.32   0.751    -.6197785     .447266
            EXTENSION_VISITS |   .1282918   .0401928     3.19   0.001     .0495153    .2070683
    DISTANCE_TO_NEAREST_MARKET |  -.0293145   .0108015    -2.71   0.007     -.050485    -.008144
                    _cons |   3.629047   1.700366     2.13   0.033     .2963916    6.961702
-----------------------------------+----------------------------------------------------------------
      /lns1 |  -.4682897   .0662518    -7.07   0.000    -.5981409   -.3384385
      /lns2 |  -.4659235    .095254    -4.89   0.000    -.6526179   -.2792292
       /r1 |   .0149924   .3878915     0.04   0.969    -.7452608    .7752457
       /r2 |  -1.749185   .5806481    -3.01   0.003    -2.887234   -.6111358
-------------+----------------------------------------------------------------
    sigma_1 |   .6260721   .0414784                 .5498329    .7128826
    sigma_2 |   .6275553   .0597771                 .5206809    .7563665
      rho_1 |   .0149913   .3878043                -.6323131    .6499694
      rho_2 |  -.9412827   .0661862                -.9938075   -.5449261
------------------------------------------------------------------------------
LR test of indep. eqns. :           chi2(1) =     5.23   Prob > chi2 = 0.0222
------------------------------------------------------------------------------

. mspredict xx, yc1_1

.
. mspredict xy, yc1_2

.
. mspredict yx, yc2_2

.
. mspredict yy, yc2_1

.
. summarize xx xy yx yy

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          xx |       114    9.325529    .6252101   7.832655   10.80978
          xy |       114    8.149157    1.028865   5.388144   9.848806
          yx |        87    8.059088    1.360614   4.599279   11.08818
          yy |        87    7.664153    1.249962   4.061326   10.43195

.
. ttest xx=xy

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      xx |     114    9.325529    .0585563    .6252101    9.209519    9.44154
      xy |     114    8.149157     .096362    1.028865    7.958246    8.340067
---------+--------------------------------------------------------------------
    diff |     114    1.176373    .0574528    .6134278    1.062548    1.290197
------------------------------------------------------------------------------
     mean(diff) = mean(xx - xy)                                t =  20.4755
 Ho: mean(diff) = 0                         degrees of freedom =      113

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

```
.
. ttest yx=yy

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
      yx |      87    8.059088    .1458732    1.360614    7.769101    8.349074
      yy |      87    7.664153     .13401    1.249962     7.39775    7.930556
---------+--------------------------------------------------------------------
    diff |      87    .394935    .0940547    .8772842    .2079604    .5819097
------------------------------------------------------------------------------
    mean(diff) = mean(yx - yy)                                t =   4.1990
 Ho: mean(diff) = 0                         degrees of freedom =       86

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0001          Pr(T > t) = 0.0000



. use "C:\Users\user\Desktop\SHIRO_DESCRIPTIVE\Objective one\merged 22.dta"

. reg DID_YOU_PARTICIPATE_IN_THE_SOYBE SEX_HHHEAD AGE_HHHEAD YEARS_SCHOOL HH_SIZE log_otherincome
ACREAGE_SIZE_IN_ACRES
> EXPERIENCE_IN_SOYBEAN_PRODUCTION ARE_YOU_A_MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS
DISTANCE_TO_NEAREST_MARKET

      Source |       SS           df       MS      Number of obs   =       201
-------------+----------------------------------   F(10, 190)      =     21.46
       Model |  26.1707464        10  2.61707464   Prob > F        =    0.0000
    Residual |  23.1725372       190  .121960722   R-squared       =    0.5304
-------------+----------------------------------   Adj R-squared   =    0.5057
       Total |  49.3432836       200  .246716418   Root MSE        =    .34923

------------------------------------------------------------------------------------------------
DID_YOU_PARTICIPATE_IN_THE_SOYBE |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------------------+--------------------------------------------------------------
                     SEX_HHHEAD |   -.020484   .0565381    -0.36   0.718     -.132007     .091039
                    AGE_HHHEAD |  -.0088242    .002274    -3.88   0.000    -.0133097   -.0043388
                  YEARS_SCHOOL |   .0264677   .0076515     3.46   0.001     .0113749    .0415606
                       HH_SIZE |   -.055493   .0096414    -5.76   0.000    -.0745109   -.0364751
                log_otherincome |  -.0327265   .0205117    -1.60   0.112    -.0731864    .0077334
           ACREAGE_SIZE_IN_ACRES |   .0215064   .0177036     1.21   0.226    -.0134144    .0564272
EXPERIENCE_IN_SOYBEAN_PRODUCTION |   .0024326   .0042127     0.58   0.564    -.0058771    .0107423
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.0292717   .0549111    -0.53   0.595    -.1375853     .079042
              EXTENSION_VISITS |   .0226457   .0055951     4.05   0.000     .0116092    .0336821
    DISTANCE_TO_NEAREST_MARKET |  -.0086139   .0015255    -5.65   0.000    -.0116229   -.0056049
                         _cons |   1.572156   .2872086     5.47   0.000     1.005629    2.138683
------------------------------------------------------------------------------------------------

. vif

    Variable |       VIF       1/VIF
-------------+----------------------
YEARS_SCHOOL |      1.37    0.729308
EXTENSION_~S |      1.28    0.784161
DISTANCE_T~T |      1.26    0.796252
  AGE_HHHEAD |      1.24    0.808258
  SEX_HHHEAD |      1.20    0.836304
log_otheri~e |      1.19    0.843178
ARE_YOU_A_~T |      1.17    0.851498
     HH_SIZE |      1.13    0.884450
ACREAGE_SI~S |      1.12    0.895465
EXPERIENCE~N |      1.07    0.936971
-------------+----------------------
    Mean VIF |      1.20

. estat imtest

Cameron & Trivedi's decomposition of IM-test

---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
```

```
  Heteroskedasticity |      92.49     63    0.0091
            Skewness |      22.18     10    0.0142
            Kurtosis |       3.36      1    0.0670
---------------------+-----------------------------
               Total |     118.03     74    0.0009
-------------------------------------------------------

.
. estat hottest


Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of DID_YOU_PARTICIPATE_IN_THE_SOYBE

         chi2(1)      =     3.78
         Prob > chi2  =   0.1519

use "C:\Users\user\Desktop\SHIRO_DESCRIPTIVE\Objective three final\merged 22.dta"

. pwcorr SEX_HHHEAD AGE_HHHEAD YEARS_SCHOOL HH_SIZE log_otherincome ACREAGE_SIZE_IN_ACRES
EXPERIENCE_IN_SOYBEAN_PRODUCTION ARE_YOU_A_
> MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET

             | SEX_HH~D AGE_HH~D YEARS_~L  HH_SIZE log_ot~e ACREAG~S EXPERI~N
-------------+------------------------------------------------------------------
  SEX_HHHEAD |   1.0000
  AGE_HHHEAD |   0.0955   1.0000
YEARS_SCHOOL |   0.1717  -0.2995   1.0000
     HH_SIZE |   0.1208   0.1698  -0.1897   1.0000
log_otheri~e |   0.0292  -0.2767   0.1997  -0.0995   1.0000
ACREAGE_SI~S |   0.2320   0.1097   0.1390   0.0756   0.0261   1.0000
EXPERIENCE~N |   0.0039   0.0042   0.1221  -0.1610   0.1141   0.1089   1.0000
ARE_YOU_A_~T |   0.0040   0.0481   0.0874   0.0530   0.0650   0.0988   0.0210
EXTENSION_~S |  -0.2110  -0.0437   0.1645  -0.0487   0.1400  -0.0373   0.0570
DISTANCE_T~T |   0.0400   0.1678  -0.3353   0.2192  -0.2850   0.0202  -0.1384

             | ARE_YO~T EXTENS~S DISTAN~T
-------------+---------------------------
ARE_YOU_A_~T |   1.0000
EXTENSION_~S |   0.3521   1.0000
DISTANCE_T~T |  -0.1022  -0.1832   1.0000

. reg LNTHINCOME_PER_CAPITA SEX_HHHEAD AGE_HHHEAD YEARS_SCHOOL HH_SIZE log_otherincome ACREAGE_SIZE_IN_ACRES
EXPERIENCE_
> IN_SOYBEAN_PRODUCTION ARE_YOU_A_MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET

      Source |       SS           df       MS      Number of obs   =       201
-------------+----------------------------------   F(10, 190)      =     49.62
       Model |  246.271197        10  24.6271197   Prob > F        =    0.0000
    Residual |  94.291114        190  .496269021   R-squared       =    0.7231
-------------+----------------------------------   Adj R-squared   =    0.7086
       Total |  340.562311       200  1.70281156   Root MSE        =    .70446

------------------------------------------------------------------------------------------------
          LNTHINCOME_PER_CAPITA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------------------------+---------------------------------------------------------------
                  SEX_HHHEAD |   .0135735   .1140486     0.12   0.905    -.2113905    .2385375
                  AGE_HHHEAD |  -.0104178    .004587    -2.27   0.024    -.0194659   -.0013698
                YEARS_SCHOOL |   .0096653   .0154346     0.63   0.532    -.0207799    .0401105
                     HH_SIZE |  -.1985062   .0194486   -10.21   0.000    -.2368691   -.1601433
            log_otherincome |   .5912393   .0413761    14.29   0.000     .5096237    .6728549
         ACREAGE_SIZE_IN_ACRES |   .0708675   .0357116     1.98   0.049     .0004254    .1413096
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.0014731   .0084979    -0.17   0.863    -.0182354    .0152892
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |   .1349142   .1107665     1.22   0.225    -.0835759    .3534043
            EXTENSION_VISITS |  -.0081344   .0112864    -0.72   0.472    -.0303971    .0141283
    DISTANCE_TO_NEAREST_MARKET |  -.0087032   .0030772    -2.83   0.005    -.0147731   -.0026334
                       _cons |   4.476342   .5793568     7.73   0.000     3.333544    5.619139
------------------------------------------------------------------------------------------------

. test
last test not found
r(302);
```

```
. ptest
command ptest is unrecognized
r(199);

. reg LNTHINCOME_PER_CAPITA SEX_HHHEAD AGE_HHHEAD YEARS_SCHOOL HH_SIZE log_otherincome ACREAGE_SIZE_IN_ACRES
EXPERIENCE_
> IN_SOYBEAN_PRODUCTION ARE_YOU_A_MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET if
DID_YOU_PARTICIP
> ATE_IN_THE_SOYBE==0

      Source |       SS           df       MS      Number of obs   =        87
-------------+----------------------------------   F(10, 76)       =     45.81
       Model |   148.68511        10   14.868511   Prob > F        =    0.0000
    Residual |  24.6679068        76  .324577722   R-squared       =    0.8577
-------------+----------------------------------   Adj R-squared   =    0.8390
       Total |  173.353017        86  2.01573275   Root MSE        =    .56972

------------------------------------------------------------------------------------------------
         LNTHINCOME_PER_CAPITA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------------------+--------------------------------------------------------------
                  SEX_HHHEAD |    .081982   .1440228     0.57   0.571    -.2048643    .3688282
                 AGE_HHHEAD |  -.0069757   .0066346    -1.05   0.296    -.0201897    .0062382
               YEARS_SCHOOL |   .0070217   .0183666     0.38   0.703    -.0295584    .0436019
                    HH_SIZE |  -.1504183   .0205759    -7.31   0.000    -.1913987   -.1094379
            log_otherincome |   .8312304   .0502521    16.54   0.000     .7311447    .9313162
          ACREAGE_SIZE_IN_ACRES |   .0257193   .0495665     0.52   0.605    -.0730009    .1244394
EXPERIENCE_IN_SOYBEAN_PRODUCTION |   .0032861   .0141398     0.23   0.817    -.0248758     .031448
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |  -.1504576   .1419983    -1.06   0.293    -.4332718    .1323565
             EXTENSION_VISITS |  -.0056755   .0249797    -0.23   0.821    -.0554268    .0440758
     DISTANCE_TO_NEAREST_MARKET |  -.0030601    .003376    -0.91   0.368     -.009784    .0036637
                       _cons |   1.410956   .7672838     1.84   0.070    -.1172223    2.939134
------------------------------------------------------------------------------------------------

. reg LNTHINCOME_PER_CAPITA SEX_HHHEAD AGE_HHHEAD YEARS_SCHOOL HH_SIZE log_otherincome ACREAGE_SIZE_IN_ACRES
EXPERIENCE_
> IN_SOYBEAN_PRODUCTION ARE_YOU_A_MEMBER_OF_ANY_AGRICULT EXTENSION_VISITS DISTANCE_TO_NEAREST_MARKET if
DID_YOU_PARTICIP
> ATE_IN_THE_SOYBE==1

      Source |       SS           df       MS      Number of obs   =       114
-------------+----------------------------------   F(10, 103)      =     10.20
       Model |  44.2065083        10  4.42065083   Prob > F        =    0.0000
    Residual |  44.6453638       103  .433450134   R-squared       =    0.4975
-------------+----------------------------------   Adj R-squared   =    0.4487
       Total |  88.8518721       113  .786299753   Root MSE        =    .65837

------------------------------------------------------------------------------------------------
         LNTHINCOME_PER_CAPITA |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
---------------------------------+--------------------------------------------------------------
                  SEX_HHHEAD |  -.0547139   .1461825    -0.37   0.709    -.3446324    .2352046
                 AGE_HHHEAD |  -.0052469   .0057122    -0.92   0.360    -.0165757    .0060819
               YEARS_SCHOOL |  -.0042584   .0219572    -0.19   0.847    -.0478053    .0392885
                    HH_SIZE |  -.2156203   .0363544    -5.93   0.000    -.2877206    -.14352
            log_otherincome |   .3923518   .0563877     6.96   0.000      .28052    .5041835
          ACREAGE_SIZE_IN_ACRES |   .0582788   .0456568     1.28   0.205    -.0322707    .1488282
EXPERIENCE_IN_SOYBEAN_PRODUCTION |  -.0102414   .0096203    -1.06   0.290    -.0293211    .0088382
ARE_YOU_A_MEMBER_OF_ANY_AGRICULT |   .2856866   .1458071     1.96   0.053    -.0034874    .5748606
             EXTENSION_VISITS |  -.0107177   .0127234    -0.84   0.402    -.0359515     .014516
     DISTANCE_TO_NEAREST_MARKET |  -.0017069   .0059293    -0.29   0.774    -.0134662    .0100525
                       _cons |   6.517072   .7855414     8.30   0.000     4.959136    8.075009
------------------------------------------------------------------------------------------------


. use "C:\Users\user\Desktop\SHIRO_DESCRIPTIVE\Objective one\merged 22.dta"

. probit DID_YOU_PARTICIPATE_IN_THE_SOYBE DISTANCE_TO_NEAREST_MARKET

Iteration 0:   log likelihood = -137.50366
Iteration 1:   log likelihood = -107.77631
Iteration 2:   log likelihood = -107.69057
Iteration 3:   log likelihood = -107.69057

Probit regression                               Number of obs     =        201
                                                LR chi2(1)        =      59.63
```

```
                                        Prob > chi2       =     0.0000
Log likelihood = -107.69057                     Pseudo R2         =     0.2168

------------------------------------------------------------------------------
DID_YOU_PARTICIPATE_IN_THE_SOYBE |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------------------+--------------------------------------------------------------------
     DISTANCE_TO_NEAREST_MARKET |  -.0466123   .0068542    -6.80   0.000    -.0600464   -.0331782
                          _cons |   1.286657   .1862198     6.91   0.000     .9216726    1.651641
------------------------------------------------------------------------------

. predict
" found where varname expected
r(7);

. mspredict
something required
r(100);

. residual
command residual is unrecognized
r(199);

. predict y
(option pr assumed; Pr(DID_YOU_PARTICIPATE_IN_THE_SOYBE))

. reg y DISTANCE_TO_NEAREST_MARKET

      Source |       SS           df       MS      Number of obs   =       201
-------------+----------------------------------   F(1, 199)       =   2950.24
       Model |   12.261878         1   12.261878   Prob > F        =    0.0000
    Residual |  .827088576       199  .004156224   R-squared       =    0.9368
-------------+----------------------------------   Adj R-squared   =    0.9365
       Total |  13.0889666       200  .065444833   Root MSE        =    .06447

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
DISTANCE_TO_NEAREST_MARKET |  -.0136489   .0002513   -54.32   0.000    -.0141444   -.0131534
                     _cons |   .9018083   .0076141   118.44   0.000     .8867935     .916823
------------------------------------------------------------------------------
```

Appendix VI: Pictorial

Soybean Sample

Soybean crop

Appendix VII: Research Permit

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 527311 **Date of Issue: 10/March/2020**

RESEARCH LICENSE

This is to Certify that Miss: Martha Wanjiru Gachuhi of Egerton University, has been licensed to conduct research in Kakamega on the topic: Effect of soybean commercialization on income of smallholder farmers in Butere Sub-county, Kenya for the period ending : 10/March/2021.

License No: NACOSTI/P/20/4178 Amended

527311

Applicant Identification Number

Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

Appendix VIII: Paper Abstract

RAAE / xxxx, 2020: 23 (1) xx-xx, doi: 10.15414/raae.2020.23.01.xx-xx

REGULAR ARTICLE

DETERMINANTS OF INTENSITY OF SOYBEAN COMMERCIALIZATION AMONG SMALLHOLDER FARMERS IN BUTERE, KENYA

Martha Wanjiru GACHUHI[*1], *George OWUOR*[1], *Edith GATHUNGU*[1]

Address:
[1] Egerton University, Faculty of Agriculture, Department of Agricultural Economics and Agribusiness Management, P.O Box 536-20115, Egerton Njoro, Kenya.
* Corresponding author: mcirug@gmail.com, phone: +254 728 957 684

ABSTRACT

Research background: Soybean commercialization plays a vital role in enhancing the livelihoods, and income of small-scale farmers. Despite the government efforts to boost agricultural commercialization in Kenya, the intensity of soybean commercialization in the Butere Sub-County has remained low for unknown reasons.
Purpose of the article: This study investigates factors influencing the intensity of soybean commercialization in Butere Sub-county with an aim of recommending policies for improving the effectiveness and efficiency of the soybean commercialization process to improve rural livelihoods as well as realize major economic goals.
Methods: A sample of 201 smallholder soybean farmers was selected using a multistage sampling procedure. Face to face interviews using a pretested semi-structured questionnaire was used to collect the data. Data analysis was done using descriptive statistics and a double hurdle regression model.
Findings & Value added: The results revealed a relatively low soybean commercialization level (56.72 %) among soybean-producing households in the study area with schooling years, the number of extension contacts, and total land size under soybean production positively and significantly influencing soybean commercialization decisions. Similarly, schooling years, the number of extension contacts, and total land size under soybean production positively and significantly determined the intensity of soybean commercialization. The study, therefore, recommends equitable access to agricultural resources by all gender, the creation of exclusive land ownership rights, and the structuring and strengthening of the extension system.

Keywords: Soybean; Intensity; Commercialization; Double hurdle Model
JEL Codes: C01; C13; C31; Q12