2015

Farmer factors for targeting in the certified maize seed market of Western and Coastal Kenya

Githinji Pauline Bilha Wairimu
Strathmore Business School
Strathmore University

STRATHMORE BUSINESS SCHOOL

Farmer factors for targeting in the certified maize seed market of Western and Coastal Kenya

by

Githinji, Pauline Bilha Wairimu

Submitted in partial fulfillment of the requirements for the
Degree of Masters in Business Administration at Strathmore University

Strathmore Business School
Strathmore University
Nairobi, Kenya

May, 2015

Declaration of Authorship

I declare that this work has not been previously submitted and approved for the award of a degree by this or any other University. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

GITHINJI, PAULINE BILHA WAIRIMU

Signed:

Date:

STRATHMORE BUSINESS SCHOOL

Abstract

Masters of Business Administration

Farmer Factors for Targeting in Certified Maize Seed Market of Western and Coastal Kenya

by
Githinji, Pauline Bilha Wairimu

Agricultural input organizations at best group farmers on the basis of geography and yet customers, the farmers, are the final arbiters on the financial performance of the organization. This research investigated multiple characteristics of farmers for farmer factors that may be used for targeting in the certified maize seed market of Western and Coastal Kenya. The latent class finite mixture method of cluster analysis was used to model a survey sample data of 313 observations and, therefore, to first define the farmer groups or segments in the study market and then investigate for those farmer factors that would be influential in targeting social or extension initiatives and marketing strategies to the farmers. The study found that the study market is dominated by smallholder farmers at 98%, and that the farmers may be grouped into two distinct farmer groups, the empowered smallholder farmer and the challenged smallholder farmer, with a proportion of 85% and 15% respectively. The empowered farmer has the desired positive agronomic practices but is socioeconomically challenged, while the challenged farmer has poor or negative agronomic practices and is socioeconomically adverse. Socioeconomic status and gender continued to be significant factors in the smallholder dominated market but negative or resistant agronomic practices were found to have the strongest associations. Consequently, farmers may be addressed as belonging to one of either profiles, and the empowered farmer profile may be the focus for marketing strategies design while the challenged farmer profile may be the focus for social initiatives design. Additionally, seed recycling resistant behavior should be a key factor and in the context of other factors as well.

Acknowledgements

I would like to extend immense appreciation to Maxwell Muyale of Strathmore Business School, who was most supportive as an MBA program administrator at the school and as a friend. I also acknowledge Chege Macharia and his team of Brenda, Galgalo, Jesca and Mary for the field support. My appreciation also goes to Dr. Robert Mudida for the continued and patient coaching and direction even when I did not seem to see the path set. Lastly but not least is the team at Root Capital, Nairobi, for the support and encouragement.

Table of Content

List of Figures

List of Tables

I dedicate this to the local entrepreneurs in agriculture that are not only doing business but are also doing social good through their commitment to the farmers.

Chapter 1

Introduction

Most development economists share the consensus that agriculture must play an active and indispensable role in the process of economic development especially for the low income developing countries (Todaro & Smith, 2010).World Bank (2009) reports that every dollar of growth from agricultural products sold outside the local area in poor African countries leads to a second dollar of local rural growth from additional spending on services, local manufactures, construction materials and prepared foods.

Unfortunately, Sub-Sahara Africa (SSA) missed out on the Green Revolution driven by Norman Borlaug in the 1960s and 1970s, which advocates for the use of improved seed and farm inputs such as fertilizer to intensify production (Guenette, 2007; Todaro & Smith, 2010). For instance, SSA farmers on average apply 10Kg of nutrients per hectare compared to 140kg in Latin America and 73kg in East Asia.

In Kenya, the adoption of certified or improved seed by smallholder farmers remains dismal, at 10%, with 90% of farmers using informal or traditional seed, and old seed technology is still in use as seen with the case of a 1986 maize seed hybrid (FAO, 2010; Munyua, Jon, Nyikal & Mburu, 2010; Bernard, Hellin, Nyikal & Mburu, 2010; Livingston, Schonberger & Delaney, 2011; Olwande & Smale, 2012; Mathenge, Smale & Olwande, 2014). Additionally, maize, which is a staple in Kenya and accounts for 20% of total agricultural production and 25% of agricultural employment, is 70% produced by smallholder farmers (Schroeder et al., 2013). Kenya has 1.6M ha of land under maize cultivation area with limited scope for further expansion as most of the arable land is already under cultivation (Schroeder et al., 2013). Furthermore, while the formal maize seed sector in Kenya is one of the best functioning in SSA, the average maize yield in Kenya is 1.8 tonnes per hectare but a yield potential of 6tonnes per hectare is possible (Schroeder et al., 2013). This is because a large proportion of the farmers still use local varieties and prefer OPVs over hybrids (Schroeder et al., 2013).

SSA, and thus Kenya, is smallholder farmer dominated and yet for smallholder farmers to improve productivity and thus transform, they must invest, innovate and take risks on agricultural technologies (DFID, 2005; Todaro & Smith, 2010). This creates a conundrum since smallholder farmers by definition are resource poor and are constrained by their ability to manage the risk-return trade-offs in moving towards intensified agriculture (Pfitzer, Krishnaswamy & Genier, 2009; Todaro & Smith, 2010; Livingston et al., 2011). Moreover, in SSA, these risks vary significantly with the different agro-ecological zones and it is hard to generalize given the diversity of farming and marketing systems (Livingston et al., 2011).

Agriculture can achieve up to four times the impact of non-agricultural growth in improving livelihoods but transformation requires a long-term perspective on market development and smallholder farmer engagement opportunities for agricultural input companies (World Bank, 2008; Pfitzer et al., 2009; Todaro & Smith, 2010).

Agricultural input companies, therefore, are being both financially and socially strategic in their service delivery and marketing strategies so as to tap into this large smallholder farmer market and to transform it into a sustainable market (Pfitzer et al., 2009; Salami, Kamara, Brixiova & Bank, 2010; Smale et al., 2011).

The business models of agricultural input companies are designed to deliver technological innovations for productivity intensification to farmers through the use of social initiatives, extension services and information flow (Pfitzer et al., 2009; Smale et al., 2011). The goal is to transform smallholder farmers into commercially viable and sustainable units, thus creating long-term demand opportunities for the agricultural input company in a shared value approach (Pfitzer et al., 2009; Livingston et al., 2011).

Furthermore, there is growing interest in data driven segmentation of markets by organizations and the acknowledgement that not all customer types will enhance revenue growth, and it is therefore critical to understand which customers enhance and which diminish profitability (Wedel & Kamakura, 2000; SAS, 2012; Bisolutions.us, 2014).

1.1 Statement of the Research Problem

Smallholder farmers are required to improve farm productivity and to transform into commercially viable entities that make for better and more sustainable long term demand for agricultural input companies and more sustainable long term food supply for the nations. In order to effectively achieve this, therefore, the social initiatives, marketing strategies, extension services and information flow activities that will transform the smallholder farmers must be appropriately targeted to the right group of farmers.

However, the characteristics and nuances of farmer groups in the Kenyan certified maize seed market remain undefined. Agricultural input organizations at best group the farmers on the basis of a single dimension, such as geographic ecological zones or a product type they are offering, and yet customers, the farmers, are the final arbiters on the financial performance of the organization.

This research, therefore, investigates multiple characteristics or behaviour of farmers for farmer factors that influence targeting in the certified maize seed market of Western and Coastal Kenya.

1.2 Research Objectives

The overall objective of this research is to identify farmer factors that influence targeting in the certified maize seed market of Western and Coastal Kenya.

The sub-objectives of this research are to determine and profile meaningful farmer groups in the certified maize seed market of Western and Coastal Kenya, and to explore the influence of socioeconomic status relative to other influencing factors in targeting the farmer groups.

1.3 Research Questions

The associated research questions are as follows.

1. What number of distinct farmer groups exist in the certified maize seed market of Western and Coastal Kenyafor targeting?
2. What are the characteristics of the identified viable farmer groups in the certified maize seed market of Western and Coastal Kenyathat make each group distinct and identifiable?
3. What farmer characteristics or factors influence targeting in the certified maize seed market of Western and Coastal Kenya?

1.4 Justification of the Research

This research builds on prior academic works by contributing to agricultural debates on how best to deliver inputs to and to empower smallholder farmers towards intensification of farm productivity and subsequent commercialization of farm output. Additionally, this research supports the argument that the base of the pyramid markets also benefit from customer driven approaches where the market perspective on customer needs and preferences is incorporated. Furthermore, the research applies the latent class mixture model in cluster analysis for market segmentation. Mixture models provide powerful new tools in target marketing and are increasingly being recognized in marketing research.

This research also builds on prior studies on the application of poverty outreach measures, for customer targeting in emerging markets or pro-poor initiatives. Poverty outreach measures serve a dual purpose; a social reporting measure in double or triple bottom line as well as a market research variable for marketing mix strategies and customer or service targeting.

Furthermore, this research contributes to the formulation and implementation of agricultural policies and strategies by both public and private entities by providing a more informed understanding and profiling of the smallholder dominated agricultural market in of Western and Coastal Kenyafor certified maize seed. It enhances the effective delivery of marketing mix strategies, promotional and social initiatives, extension services and product development and innovation of improved maize seeds. Furthermore, by doing so, this research supports driving access to and adoption of the right maize seed variety in order to enhance agricultural productivity in Kenya.

This research may be used by agricultural businesses in the certified maize seed market to inform their marketing mix strategies, product development and innovation, and their targeted extension and social services to farmer consumers.

This research may also be used by extension, public and development organizations interested in smallholder agriculture for the purposes of programming initiatives, policies and strategies on agricultural intensification and commercialization.

1.5 Scope

This research looks at the maize crop in the Kenyan certified seed market, which is significantly larger and better developed than other crops. As a result, the research does not address market characteristics for other crop seeds or seek to draw relationships between maize and other crops in generalizing the resulting market characteristics.

Furthermore, this research covers the of Western and Coastal Kenyacertified maize seed markets, which were accessible logistically to the researcher.

In addition, this research does not explore the organizational end characteristics that are also an input in market targeting.

1.6 Overview

The rest of this work proceeds as follows. The next section, chapter two, evaluates prior works in market development for agricultural inputs with a particular focus on certified maize seed in of Western and Coastal Kenya. Chapter two also details the theoretical framework. Chapter three then details the methodology, data design and the latent class cluster analysis model used in the evaluation. Chapter four follows with an analysis of the empirical results and potential recommendations are detailed in Chapter five.

Chapter 2

Literature Review

In order to address the research objective of which farmer characteristics influence targeting in the certified maize seed market of Western and Coastal Kenya, we first answer the questions what viable sub-groups, or segments, exist in that market and what are their distinct characteristics. We then address the primary objective of the research by investigating which farmer characteristics are influential in how the farmers group and are, therefore, relevant for targeting.

This chapter, therefore, presents a theoretical framework for the study and a thematic review of works done in certified maize seed market segmentation. The emerging thematic issues for this literature review are, firstly, what are the observable characteristics of the market of interest? Secondly, which farmer characteristics or behaviour are relevant and measurable in defining the farmer groups? Particularly, what segmentation variables are used. Thirdly, which method would optimally partition or group the farmers using the farmer context-based data and provide practical results? Lastly, what do we learn from related prior works on the segmentation of a certified seed market in a smallholder dominated region.

2.1 Introducing Segmentation

Organizations employ marketing functions in order to deliver value to the customer. Customers will purchase when the benefits exceed the costs and the products or services offer superior value compared to alternatives, even if for a higher price point (Iacobucci & Churchill, 2009; Gilligan, 2009; Kotler, 2012). Consequently, organizational decisions stand to benefit more if they incorporate the voice or behavior of the customer.

Customers have different motivations and preferences in their consumption, which makes any market generally heterogenous (Iacobucci & Churchill, 2009; Gilligan, 2009; Kotler, 2012). However, the same customers will tend to exhibit similarities in certain dimensions that enable them to be grouped and treated as a single entity more effectively and with better resource allocation(Iacobucci & Churchill, 2009; Gilligan, 2009; Kotler, 2012). This is the premise and motivation for segmenting markets.

Market Segmentation is a fundamental strategic marketing concept, defined as a data driven statistical process of dividing a heterogeneous customer base on several relevant dimensions into subgroups that are similar or homogeneous (Wedel & Kamakura, 2000; Iacobucci & Churchill, 2009; Gilligan, 2009; Kotler, 2012). The better the segments chosen for targeting by an organization, the more successful the organization is in being competitive and in growing its long-term profitability (Ulwick, 2003; Iacobucci & Churchill, 2009; Gilligan, 2009; Kotler, 2012; SAS, 2012; Bisolutions.us, 2014). However, data driven approaches tend to be resource intensive and management will often opt to rely on their subjective knowledge and expertise in defining markets and customer groups (Wedel & Kamakura, 2000; Ulwick, 2003; Iacobucci & Churchill, 2009; Gilligan, 2009; Kotler, 2012; SAS, 2012; Bisolutions.us, 2014). Organizations may, therefore, benefit more from incorporating a data driven approach that provides reasonable statistical rigour at minimal effort, such as public reports on or analysis of markets, as opposed to relying solely on internal subjective measures.

Market segmentation has considerable potential benefits to creating a competitive advantage including price differentiation, niche markets and product innovation and development (Ulwick, 2003; Hunt & Arnett, 2004; Iacobucci & Churchill, 2009; Gilligan, 2009; Kotler, 2012). However, there are two other approaches to marketing strategies, besides segmenting, namely undifferentiated or mass marketing and concentrated or one-on-one marketing (Gilligan, 2009). All in all, segmentation still plays a role, to some degree, in these strategies as well and each approach is suited for a different context (Gilligan, 2009). Targeting prioritizes customer groups for different products, services or communication approaches providing a more customer need based approach in delivery, which is also better focused (Wedel & Kamakura, 2000; Ulwick, 2003; Iacobucci & Churchill, 2009; Gilligan, 2009; Kotler, 2012; SAS, 2012; Bisolutions.us, 2014). Furthermore, Ulwick (2003), adds that many managers believe they do a good job capturing their customers requirements, while in reality, they do not and their failure to do so is preventing them from managing innovation as a key business process. Therefore, agricultural input companies may use such a customer need based and data driven approach to be more effective in how they deliver social and extension services, formulate marketing mix strategies and develop products.

2.2 Characteristics of The Market

Karnani (2007) argues that large-scale industries are better placed to realize economic development and focus, therefore, efforts should be redirected from the poor or resource challenged households that are less likely to invest at scale. However, developing economies are dominated by these resource challenged households and agriculture is the main economic activity. Therefore, economic development in developing countries is pegged on rural and agricultural development, which requires transformation of the majority inhabitants into commercially viable entities (Todaro & Smith, 2010; Livingston et al., 2011; Juma, 2011). Particularly, agricultural intensification practices, a premise for improving farm productivity and subsequent commercialization of farm output, are low in SSA and only 10% of farmers are using improved seed for instance (Nyoro, Kirimi & Jayne, 2004; Munyua et al., 2010; Bernard et al., 2010; Pfitzer et al., 2009; Schroeder et al., 2013). Consequently, smallholder farmers, constituting 75% of these resource challenged households in Africa, remain an unexploited market for agricultural inputs and intensification technologies. Markets, institutional and social efforts should be made to drive access to and adoption of agricultural intensification inputs or technologies by smallholder farmers to empower and transform them into commercial producers, who are more likely to invest at scale.

By definition, agricultural inputs or technologies entail improved or certified seeds, plant materials, crop protection, fertilizers, animal feed, veterinary medicines and services, agricultural equipment and machinery, irrigation schemes, knowledge or education of agricultural best practices, ICT and financial services for agriculture (Livingston et al., 2011). Among agricultural inputs, certified seed is recognized to have the greatest ability of increasing on-farm productivity since seed determines the upper limit of crop yields and the productivity of all other agricultural inputs (Livingston et al., 2011; Munyua et al., 2010; Bernard et al., 2010; Olwande & Smale, 2012; Mathenge et al., 2014).

While commercialization is desired, markets generally fail to address social issues due to the single minded focus on the bottom line; the financial goals and shareholder wealth (yunus09). It may be argued, though, that the primary objective of business is to maximize shareholder wealth. However, doing good business makes for sustainable growth and profitability for businesses and, thus, is in line with maximizing shareholder wealth (Yunus & Weber, 2009). Good business entails a socially and or environmentally responsible approach that incorporates the society interests by minding the people and the planet that constitute the context in which organizations operate (). To achieve sustainability and growth, therefore, agricultural input business will do better to respond to their smallholder dominated context, which requires a social lens at the very least.

Consequently, the business models of agricultural input companies are being designed to deliver technological innovations for productivity intensification to farmers through the use of social initiatives, extension services and information flow (Pfitzer et al., 2009; Smale et al., 2011). The goal is to transform smallholder farmers into commercially viable and sustainable units, thus creating long-term demand opportunities for the agricultural input company in a shared value approach (Pfitzer et al., 2009; Salami et al., 2010; Todaro & Smith, 2010). However, subjective management decisions and singular treatment of smallholder farmers in the markets may not be as effective as a data-driven and a more focused approach in the delivery of the products and the social and extension services.

While there are varying definitions of who a smallholder farmer is based on land size, it is generally agreed that smallholder farmers own little household assets, practice subsistence farming and are resource and capital constrained (Pfitzer et al., 2009; Munyua et al., 2010; Bernard et al., 2010; Todaro & Smith, 2010; Schroeder et al., 2013). Principal livelihoods for smallholder farmers are based on agricultural production, for home consumption, market sale and off-farm employment in agricultural inputs and post-harvest processing and distribution (Todaro & Smith, 2010; Smale et al., 2011).

The ability of smallholder farmers in SSA to increase on-farm investment is constrained by their ability to manage the risk-return trade-offs in moving towards intensified agriculture (Todaro & Smith, 2010; Livingston et al., 2011). As a result, traditionally, economists have identified smallholder farmers as being economically irrational (Todaro & Smith, 2010). However, risk is both business related and social as the poor rural families have livelihood challenges, which is social risk, and this influences their economic decisions. Therefore, they are highly risk averse financially and thus less inclined than non-poor groups to move up the risk-return ladder towards potential higher incomes (Livingston et al., 2011). As a result, they are rational as they prioritize to minimize social risk and then balance in maximizing economic returns (Todaro & Smith, 2010; Livingston et al., 2011). Consequently, effort should be put to empower and transform smallholder farmers using both the financial and social lenses.

2.3 Theoretical Basis of Segmentation

Analytic approaches, such as cluster analysis, support decision making in market segmentation by identifying granular segments of customers, predicting membership to and behavior of a particular segment and by optimizing capabilities that maximize economic outcomes or some desired objective (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Jung & Wickrama, 2008; Iacobucci & Churchill, 2009; Gilligan, 2009; Everitt, Landau, Leese & Stahl, 2011; Kotler, 2012; SAS, 2012; Bisolutions.us, 2014). Cluster analysis identifies customer groups that minimize differences among members of same group (highly internally homogeneous groups) while maximizing differences between different groups (highly externally heterogeneous groups)(Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Jung & Wickrama, 2008; Everitt et al., 2011).

There is a plethora of cluster analysis methods available and the decision on which method to apply has been ad hoc in practice (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Jung & Wickrama, 2008; Gilligan, 2009; Everitt et al., 2011; SAS, 2012; Bisolutions.us, 2014). It is important that the selected method serve the purpose of the study, for instance descriptive and/or predictive, and the structure of the data as well, for instance nominal data resilience in addition to interval data analysis (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Jung & Wickrama, 2008; Gilligan, 2009; Bisolutions.us, 2014). Advancement in computation capabilities have seen to the recent growing interest in the more statistically robust mixtures latent class cluster analysis methods for market segmentation as opposed to the traditional hierarchical approaches(( )wedelk00, verm02, jung08, tuma13).

In the over 50 years of segmentation development, since its introduction in 1956 by Wendell Smith, a variety of approaches have emerged but the common thread has been the importance of understanding in detail the structure of a market (Gilligan, 2009). Furthermore, the task to a strategist continues to be in deciding which of the approaches, or combination of, best partitions the market in question and for the desired objectives (Gilligan, 2009; SAS, 2012; Bisolutions.us, 2014). Growing interests in market driven strategies and innovations, and improvements in technological and computational systems are making it possible to apply the more robust data driven segmentation approaches but the practice of it is still wanting (SAS, 2012; Bisolutions.us, 2014).

Additionally, recent trends and issues within segmentation approaches touch on the selection of segmentation variables or basis and selection of segmentation model or clustering technique (**etc**; Iacobucci & Churchill, 2009; Gilligan, 2009; Kotler, 2012; SAS, 2012; Bisolutions.us, 2014). There is also the question of choosing between a priori and post hoc methods. In an a priori approach, the segmentation factors of focus are decided upon in advance, while in a post hoc, the interesting factors are allowed to emerge from a data driven analysis (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Jung & Wickrama, 2008; Gilligan, 2009). Post hoc approaches are more robust and objective in exploring how customers group and behave in reality, while finite mixture methods are more statiscally robust (Gilligan, 2009; Everitt et al., 2011).

2.3.1 Latent Class Cluster Analysis Method

The connection between latent class method and cluster analysis was first made in 1970 by Wolfe, while the application of latent class cluster analysis in market segmentation was first suggested by Green, Carmone and Wachspres in 1976 (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Jung & Wickrama, 2008). However, the computational intensity of latent class methods made it impractical until recently, with the advancement in computational power and software engineering capabilities, enabling more developments in and applications of the computationally intensive fuzzy clustering methods and particularly the more statistically robust mixtures of latent class methods (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Jung & Wickrama, 2008; Tuma & Decker, 2013; Bisolutions.us, 2014).

Latent class analysis or mixture models is a class of methods that attempts to explain the observed associations between factors in data by introducing unobservable underlying (latent) classes, which are the hidden clusters and constitute the dependent variable in the analysis (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Jung & Wickrama, 2008; Everitt et al., 2011; Tuma & Decker, 2013). Latent class analysis is also referred to as latent discriminant analysis in marketing research and the hidden clusters are the discovered segments in the marketing research.

The Latent class cluster analysis (LCCA) is a post-hoc, predictive, fuzzy and non-hierarchical approaches to cluster analysis (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Magidson & Vermunt, 2002; Jung & Wickrama, 2008; Everitt et al., 2011). In addition, unlike traditional approaches, the model does not rely on traditional modeling assumptions of linearity, normal distribution and homogeneity (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Magidson & Vermunt, 2002; Jung & Wickrama, 2008). This makes it more applicable to real life situations with varying characteristics.

While hierarchical and other non-hierarchical approaches have their own merits, the LCCA approach compares in the following ways. Firstly, compared to a priori method, LCCA does not require prior knowledge of the type and number of groups or classes in the data in questions (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Magidson & Vermunt, 2002; Jung & Wickrama, 2008; Everitt et al., 2011; Tuma & Decker, 2013). Furthermore, unlike other works, which introduce Ward analysis or the Heckman correction to avoid selection bias, LCCA achieves correction for selection bias in the same step that it does the clustering (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Magidson & Vermunt, 2002; Jung & Wickrama, 2008; Everitt et al., 2011). This makes for a more objective exploration of the data for its underlying characteristics and, therefore, a more representative and truer discovery of the number of relevant or viable segments and of the partitioning variables that are statistically influential in defining the resulting segments.

Lastly, while, as a fuzzy approach, LCCA is computationally intensive compared to traditional hierarchical approaches and other non-hierarchical approaches, it is computationally faster than other fuzzy methods and achievable using existing technology (Jung & Wickrama, 2008; Everitt et al., 2011; Hennig & Liao, 2013; Bisolutions.us, 2014). This makes it a more practical state of the art approach in clustering.

On the other hand, LCCA has the challenge that it is likely to converge on a local solution as opposed to a global maximum of the entire data set under analysis, resulting in clusters that are not representative of the population (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Magidson & Vermunt, 2002; Jung & Wickrama, 2008; Everitt et al., 2011). LCCA deals with this issue by using random starting values to reduce the likelihood of converging on a local solution (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Magidson & Vermunt, 2002; Jung & Wickrama, 2008; Everitt et al., 2011). An additional approach is a hybrid approach that uses Bayesian prior information in conjunction with the randomized starting values, which eliminates the possibility of obtaining boundary or extreme solutions and reduces the chance of obtaining local solutions (Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Magidson & Vermunt, 2002; Dolnicar, 2003; Jung & Wickrama, 2008; Everitt et al., 2011).

2.3.2 Segmentation Variable Selection

A segmentation base is a set of variables or characteristics used to define the homogeneous subgroups in a customer base (Wedel & Kamakura, 2000; Dean & Raftery, 2010). There are no scientific procedures for selecting segmentation variables. However,four segmentation bases have emerged as the most popular in segmentation studies namely geographic segmentation, demographic segmentation, psychographic and behavioural segmentation (Kotler, 2012; Gilligan, 2009; SAS, 2012; Bisolutions.us, 2014). For instance, geographic bases include geographic region and climate, demographic bases include age and gender, psychographic bases include lifestyle activities and opinions and behavioural bases include benefits sought (Kotler, 2012; Gilligan, 2009).

A successful segmentation plan must produce market segments which meet the four basic criteria of sustainability, identifiability, accessibility and responsiveness (Kotler, 2012; Gilligan, 2009). Geographic and demographic approaches more practically applicable and, therefore, more popular. However, rarely can a single dimension be used to segment effectively (Gilligan, 2009; Tuma & Decker, 2013). Variable selection is a critical part in the segmentation process.

Recent trends and issues in segmentation touch on the use of bases other than geographic and demographic, and in particular behavioural and psychographic techniques that better capture motivation compared to preferences, which are dynamic (Kotler, 2012; Gilligan, 2009; SAS, 2012; Bisolutions.us, 2014). Behaviour explains why a decision is made and is a habit that is more stable in the immediate term. However, a holistic or multidimensional approach is recommended for greater depth and the strength of behavioural or psychographic variables does not negate the role of other types of variables (Gilligan, 2009; Kotler, 2012). Moreover, the selected variables should connect with the intended purpose (Gilligan, 2009). As already seen, smallholder farmer choices are driven by the need to minimize social risk first, which makes behavioural and psychographic analysis a better bases of analysis in addition to direct product preference and geo-demographic bases.

Consequently, this research considers the following segmentation variables.

2.3.2.1 Socioeconomic Status Variable

Socioeconomic status is derived from measures involving household income and education level and it is indicative of both the social and economic risks of a household in achieving a reasonable standard of living (Wedel & Kamakura, 2000; Sebstad & Cohen, 2000; Gonzalez, 2006). Smallholder farmers are considered to have unfavourable socioeconomic status, which presents a challenge in converting them into sustainable intensive commercial entities by affecting their purchasing power and their tendency to try out new agronomic best practices and technologies (Pfitzer et al., 2009; Munyua et al., 2010; Bernard et al., 2010; Livingston et al., 2011; Schroeder et al., 2013). Therefore, this study looks at the overall socioeconomic well being as opposed to a single economic or social risk component.

To measure for this socioeconomic variable, a multifaceted approach on income, assets and consumption is necessary. Poverty measurement approaches measure for socioeconomic status and different approaches exist with rapid assessment methods being a key issue as opposed to the traditional large household methodologies (Gonzalez, 2006; Coleman & Spellberg, 2007; The Iris Center, 2011). The progress out of poverty index (PPI) is a rapid assessment poverty measurement methodology, which derives from the underlying national household surveys on socioeconomic status (Zeller, 2004; Progress out of Poverty, 2010; The Iris Center, 2011). This provides a tested and practical method for the measure of socioeconomic status in a holistic way.

Furthermore, the PPI requires minimal expertise to use, takes about five minutes to complete and it is built for practical use over statistical rigour, without compromise to statistical relevance (Coleman & Spellberg, 2007; Progress out of Poverty, 2010; CGAP, Ford, EU & Taskforce, 2010; The Iris Center, 2011; SAS, 2012; Ele-Ojo, Eme & Fonta, 2013). These are useful characteristics for market research design in that measures can be readily observed and measured for effectively and with minimal resources. Furthermore, the same data that is used for the segmentation can be used by an organization for social impact reporting purposes as is fit.

2.3.2.2 Personal Characteristics Variables

Personal characteristics entail other demographic and geographic variables, in addition to the socioeconomic status. Such variables provide observable segmentation bases called covariants that are not necessarily predictors but serve to enhance the accessibility of segments discovered

(Wedel & Kamakura, 2000; Vermunt & Magidson, 2002; Magidson & Vermunt, 2002; Jung & Wickrama, 2008; Everitt et al., 2011; Tuma & Decker, 2013).

The underlying household questions of the PPI also capture demographic information in the same set of questions and incorporate them, in weighted form, in the computation of the final socioeconomic status value (Zeller, 2004; Progress out of Poverty, 2010; The Iris Center, 2011). Some of the captured variables of interest are occupation, education and household size. Therefore, this study only adds age and gender measures to the demographic dimension.

Furthermore, infrastructural elements, such as access points for agronomic information and for input and output markets, would be interesting as they create access, convenience and awareness in order for the consumer to act and thus purchase (Gloy & Akridge, 1999; Alexander, Wilson & Foley, 2005; Munyua et al., 2010; Bernard et al., 2010; Roucan-Kane, Alexander, Boehlje, Downey & Gray, 2011; Feeney, Berardi & Steiger, 2011; Feeney & Berardi, n.d.). However, in the next sub-section, we consider farmer practices that are indicative of awareness.

Additionally, while studies do recognize that land-size is a common measure in the definition of who is a smallholder farmer, there is not a single agreed upon range or measure of what that measure should and the range may vary significantly by geography (Pfitzer et al., 2009; Todaro & Smith, 2010; Olwande & Smale, 2012; Schroeder et al., 2013; Mathenge et al., 2014). However, farmers may also have larger tracks of land but the capacity to only cultivate a very small portion of it or the practice of subsistence farming (World Bank, 2009; Todaro & Smith, 2010; Bernard et al., 2010; Livingston et al., 2011; Schroeder et al., 2013). Consequently, this study captures size of land under actual cultivation for maize.

In the research design phase, therefore, the land size variable was designed to capture a higher level of granularity at the lower band of zero to one acres of land so as to further define smallholder farmers in that range.

2.3.2.3 Agronomic Practices Variables

Agricultural practices of farmers are indicative of behaviour that either supports or resists the adoption of new agricultural technologies or inputs. Seemingly resistant behaviour and activities to new technology, however, may be due to a lack of knowledge of the benefits of the new technology or a lack of awareness of when and/or where the different product types are best applicable (Pfitzer et al., 2009; Schroeder et al., 2013). The use of certified maize seed in Kenya is dismal despite Kenya having one of the most mature certified maize seed market in SSA (Ouma et al., 2002; Muhammad et al., 2003; Nyoro et al., 2004; Pfitzer et al., 2009; Munyua et al., 2010; Bernard et al., 2010; Erenstein, Kassie & Mwangi, 2011; Schroeder et al., 2013). However, some seed types, such as the open pollinated seeds (OPV), are designed for at least one re-use (Munyua et al., 2010; Bernard et al., 2010; Erenstein et al., 2011; Schroeder et al., 2013). In this study, we consider the recycling of certified maize seed as indicative of resistant behaviour or lack of knowledge. We differentiate between recycling of hybrid seeds and OPV, which are popular varieties in the Kenya market and that are meant for different agro-ecological zones but are not necessarily applied as intended.

Supportive behaviours will be observed in practices that complement the target practice and which further improve farm output and farm returns (Pfitzer et al., 2009; Schroeder et al., 2013). Furthermore, such behaviours also identify farmers that are highly likely to be innovative in their farm activities and to, therefore, invest in agricultural technologies and inputs (Pfitzer et al., 2009; Munyua et al., 2010; Bernard et al., 2010; Erenstein et al., 2011; Feeney et al., 2011; Livingston et al., 2011; Roucan-Kane et al., 2011; Schroeder et al., 2013). Such practices include the use of fertilizer alongside certified seed or the existence of commercial activities from the farm output as a consequence (Pfitzer et al., 2009; Munyua et al., 2010; Bernard et al., 2010; Erenstein et al., 2011; Feeney et al., 2011; Livingston et al., 2011; Roucan-Kane et al., 2011; Schroeder et al., 2013). This study, therefore, considers whether the farmer applies fertilizer on his farm.

2.3.2.4 Product Characteristics Variables

Benefits are the basic reasons for the heterogeneity in choice behavior and are, therefore, the most relevant bases for segmentation providing general understanding of markets and supporting decisions about positioning, new product concepts, advertising and distribution because of their actionability (Wedel & Kamakura, 2000; Gilligan, 2009). Certified seed are an agricultural technology for improving farm productivity and commercialization, with the expectation that yields are higher for instance, and that there is surplus over subsistence use for sale. (Nyoro et al., 2004; Pfitzer et al., 2009; Munyua et al., 2010; Bernard et al., 2010; Schroeder et al., 2013). Consequently, we include perceived harvest benefits from the certified seed and sale of harvest in our segmentation variables.

2.4 Empirical Review of Prior Works in Agricultural Inputs Market Segmentation

There are other works on the segmentation of farmer markets for agricultural inputs. These, however, have been done primarily in the United States of America and have looked at commercial producers (Gloy & Akridge, 1999; Alexander et al., 2005; Roucan-Kane et al., 2011; Borchers et al., 2012). Consequently, this study addresses a new geography, Kenya, in studies on agricultural inputs and for certified seed in particular. Furthermore, by addressing this new geography, this study also looks at a different class of farmers, the smallholder farmer, that dominates agriculture in developing countries, thus adding an emerging markets or base of the pyramid dimension for SSA in particular.

On the other hand, where farmer market studies have touched on the smallholder markets, they have had an access to credit lens as the segmentation objective (Christen & Anderson, 2013). This research, however, looks at the agricultural input sector, specifically certified seed input, for the purposes of targeting marketing activities and delivery of associated social and extension services.

Moreover, Bernard et al. (2010) have specifically looked at the significance of market transaction costs, as an economic component, on the tendency of smallholder farmers to adopt certified seed in Kenya. In their analysis, they did not define the certified maize seed market structure. They also find that market transaction costs were not a significant influence on the decision to purchase certified seed. Market transaction costs looked at the incremental amounts, on the cost of the product itself, that a farmer spent to learn about and access the product and or to deliver his produce to the markets for sale (Bernard et al., 2010). Conversely, in this research, we consider socioeconomic status as a holistic view of the resource capacity of the farmer and as a proxy of their tendency to prioritize minimizing social risk over maximizing economic returns and thus increasing on-farm investment. Moreover, we first define the way in which farmers in the certified maize seed market under study group, and can, therefore, be targeted for marketing mix strategies and other related social and extension services. This is not included in the study by Bernard et al. (2010).

Additionally, prior works in the agricultural markets have applied two step analysis such as using non-hierarchical k-means algorithms or Ward and Heckman correction on selection bias for descriptive analysis, and multinomial logit regression for overlapping predictive analysis (Gloy & Akridge, 1999; Alexander et al., 2005; Munyua et al., 2010; Bernard et al., 2010; Roucan-Kane et al., 2011; Feeney et al., 2011; Feeney & Berardi, n.d.). This research, on the other hand, applies the more recent and statistically robust finite mixture approach by using the LCCA method.Furthermore, the above works have been primarily descriptive of the existing markets without additional input on addressing future customers, with the exception of Feeney et al. (2011), who have added a predictive model dimension in their research of the Argentine farmer seed market.

2.5 The Gap

This study takes a data driven and customer needs based decision making approach to understanding an organization's market. It explores for those factors that would influence targeted marketing mix strategies and the delivery of farmer services as opposed to treating a smallholder farmer dominated market as one large group.

Furthermore, unlike the traditional single-dimension internal practice of looking at agricultural markets by geography, this study takes into consideration additional variables and the voice or needs of the customer in defining the agricultural market. In addition, that multi-dimensional approach in defining the agricultural market, uses behavioural or psychographic variables that may allude to motivation which is a more stable characteristic.

The study also employs the more statistically robust finite mixture model clustering techniques that have previously been impractical due to their computational intensity and the less advanced state of computational technologies of prior times.

2.6 Theoretical Framework

The theoretical framework used in this study is a cluster analysis statistical method to segmenting markets and identifying influential factors in the market. It explores for a statistically relevant number of viable farmer groups in the certified maize seed market of Western and Coastal Kenyaas well as the defining characteristics of those groups. The results are then applied in identifying those farmer characteristics or factors that may influence targeting in the certified maize seed market of Western and Coastal Kenya.

Consequently, the dependent variable is the farmer group or class, which is the market segment being explored for. This farmer group may be expressed as a function of the following independent variables.

$$
\begin{aligned}
\text{Farmer Group or Segement} = y = f(&Personalcharacteristics, Socioeconomicstatus,\\
&Agronomicpractices, Productcharacteristics)\\
&\textbf{Where:}\\
\text{Personal characteristics} &= \text{age, gender, land size,}\\
\text{Socioeconomic Status} &= \text{Probability below \$2.45 international povertly line}\\
\text{Agronomic practices} &= \text{Resistant behaviour: recycling of seed,}\\
&\quad\text{Supportive bahaviour: use of fertilizer}\\
\text{Product characteritics} &= \text{benefit expected from yields}
\end{aligned}
\tag{2.1}
$$

As a result, further analysis will be done on the resulting models that define the farmer groups so as to evaluate for those independent variables that are influential in the farmer grouping.

The resulting theoretic framework is as follows.

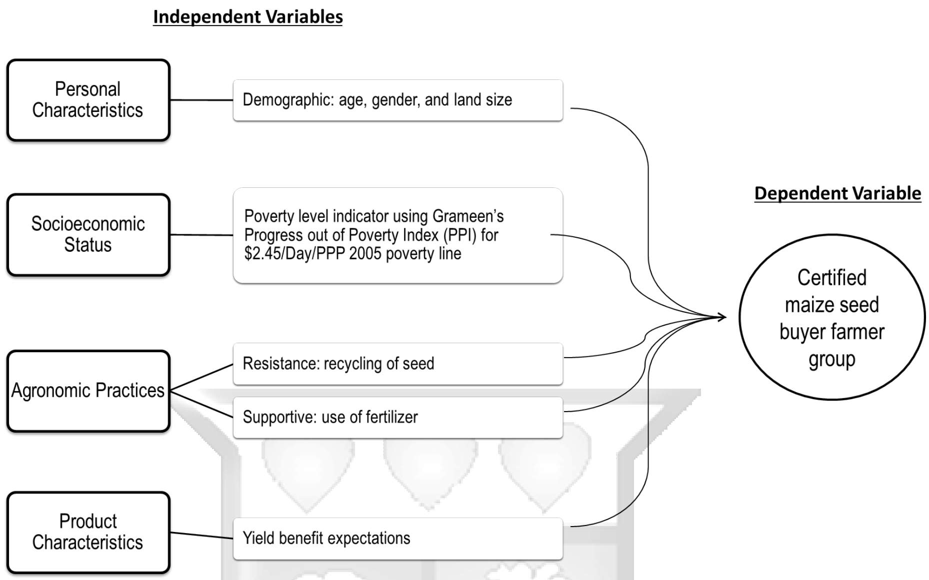


FIGURE 2.1: Theoretical framework

It is expected that some of the independent variables may have a significant influence and therefore act as predictor variables that are indicative of the factors that would influence targeting. Similarly, some of the independent variables may end up being just descriptive without any significant correlation to how the farmers group.

Firstly, it is hypothesized that there exist distinct farmer subgroups within the larger smallholder farmer dominated market seed market of Western and Coastal Kenya, deserving of different treatments. The expected number of clusters or farmer groups is greater than one.

Secondly, it is hypothesized that socioeconomic status has a significant influence on the grouping of farmers in the certified maize seed market of Western and Coastal Kenya.

Thirdly, it is hypothesized that resistant agronomic practices are more influential than supportive agronomic practices as a factor for targeting the farmers in the certified maize seed market of Western and Coastal Kenya.

Lastly, it is hypothesized that land size, while an identifying factor of farmers in a smallholder dominated market, does not have a significant influence in the way farmers group in a smallholder dominated market such as the certified maize seed market of Western and Coastal Kenya.

Chapter 3

Research Methodology

3.1 Research Design

The research is an exploratory study of households that consume certified seed in the Western and Coastal certified maize seed markets of Kenya for the factors that influence targeting within this group. The study first explored the characteristics of the farmers in the certified maize seed market of Western and Coastal Kenya, so as to identify what viable sub-groups or segments exist in that market, and what their distinct characteristics are. The study investigated which farmer characteristics were influential in how the farmers group and were, therefore, relevant for targeting.

The statistical unit of the survey, the participant, was an individual, who would also represent a household. The survey ran once during the planting season when participants buy certified seed.

3.2 Population and Sampling

Since there was not a single complete registry of all certified maize seed buyers and the general populace constitutes the universe of certified maize seed customers, it would have been prohibitively expensive to develop a sampling frame, and therefore, there was no defined sampling frame for the study unit.

To achieve a balance between rigour and practicability with regard to cost and effort, a 90% confidence level with a +/-5% margin of error is considered reasonable for market research, resulting in an expected sample size of 272 for a large populace of 100,000 or more (Iacobucci & Churchill, 2009; Burns & Bush, 2003). It was estimated that, for this study, a sample size of 312 would be reasonable for the 90% confidence level, while giving room for erroneous data of up to 40 records.

A two stage process was used in sampling for the participants. First, survey sites were selected based on retail distribution points or agro-dealer stores. These distribution points or survey sites were the points from which participants were then sampled and interviewed. Second, individuals were selected at the identified sites for participation.

3.2.1 Survey Site Selection

A registry of agro-dealer stores that distribute certified maize seed products was obtained from one of the local certified maize seed processors. There were a total of 125 agro-dealer stores registered with the processor for the Western and Coastal region of Kenya. This provided a sampling frame for survey site selection.

To select a survey site, first, the agro-dealer stores were grouped by region and six agro-dealer stores were randomly selected for each region. Random selection was achieved by first having the list of agro-dealer stores from the processor in an excel worksheet. Second, the excel random function was used to generate a random number for each record in the worksheet and all in one go with one random number generator seed. Third, the generated random number values were locked in to avoid updating each time there was activity in the worksheet since because that is how the excel random function operates. Fourth, the records were ordered by the random number column in ascending order and the top ”n” records were selected accordingly. In this case the top six survey sites were selected for each region.

Second, these selected sites were then reviewed for logistics feasibility and to ensure that active sites were represented as these would also give us greater access to consumers for better participant selection. As a result, some of the sites were dropped off to leave four sites per region. The four sites per region were not in the same town and were within four hours of travel distance from each other. In addition, a route map was taken into consideration to facilitate with movement that was efficient for the enumerator, as an enumerator would be responsible for a region and not just a single survey site.

Fourth, one of the four sites acted as a backup site should we have required to bolster collected records in an area within the budgeted time or should a target site have been unavailable for whatever reason. Therefore, enumerators were expected to survey at the first three of the sites and use the fourth site as a backup site.

The Western and Coastal regions of Kenya were selected due to accessibility by the researcher.

3.2.2 Respondent Selection

Since we did not have a sampling frame for the study unit, the certified seed consumer, we selected participants as follows. Firstly, an enumerator reported at a survey site and had an introduction session with the agro-dealer. This session informed the agro-dealer about the survey and that the enumerator would be setting camp outside the store to interview certified seed buyers. The agro-dealer was invited to facilitate by informing customers about the survey and requesting their cooperation. However, he had no knowledge of the survey content and objectives.

Secondly, the enumerator did the following for each potential respondent:

- On the first day, the enumerator interviewed the 3rd customer that walked out of the store after buying certified seed. He then interviewed every 2nd farmer after that.
- On the second day, the enumerator interviewed the 2nd customer that walked out of the store after buying certified seed. He then interviewed every 3rd farmer after that.
- The enumerator then repeated the above over the rest of the survey days.

3.2.3 Enumeration

A team of four enumerators were purposely selected to identify a homogeneous team with regard to literacy and occupation level. The data collection was conducted simultaneously across the country.

The sample size of 312 was equally shared among the resulting 12 sites to make a total of 26 records per survey site. One enumerator was assigned to every three sites to make a total of 78 records per enumerator.

The enumerators were University students, in agronomic studies, that had also interned with a certified maize seed processor and were therefore familiar with the domain area and had conducted survey interviews before, thus shortening the learning curve. We selected four enumerators from this group of students, who applied for the roles upon posting of a paid internship application request.

A list of the agro-dealers sites by region and the final list of selected sites is included in the appendix as item 6.

3.3 Data Collection Tools

There were two data collection tools. First, was the progress out of poverty indicator (PPI) survey for Kenya, which was used for measuring socioeconomic status. Second, was the certified maize seed market survey, which measured for farming behaviour and certified maize seed related characteristics. The PPI questions can be found in item 6 of the annex, while the certified maize seed market questions can be found in item 6 of appendix A.

In order to enforce data quality checks on completeness, response consistency, data types and user input, the questionnaires was digitized onto a mobile data collection tool, as opposed to using a paper-based questionnaire. In addition, use of mobile ensured accountability of enumerators with regard to location and targets. Item 1 of appendix B includes some screenshots of the mobile data collection tool.

Furthermore, the enumerators were trained on the methodology and the questions, and a buffer had been included in the sample size in order to further mitigate systematic errors from wrong use. The training also covered how to select a participant and how to use the mobile tool for data collection.

3.4 Data Analysis

The survey data was quantitative. It was expected that given the sample design and the sample size and allocation detailed above, the sample results would provide representative estimates for the target population and would have the desired small sampling error.

The Latent Class Cluster Analysis (LCCA) method was used for data analysis to partition or cluster the data and determine predictive group membership models.

The LCCA model operates the market segment or class, in this case the farmer group, as the dependent variable to be explored for and identified by the data. The model then categorizes the dependent variables into either 1. Demographic variables and other covariates, which add depth to the description of the identified segments, or 2. Predictors, which define the class membership probability model. Consequently, the farmer group, market segment or class may be expressed as a function of independent descriptive characteristics (or covariates) and select independent predictor variables.

Recalling the theoretical framework

$$
\begin{aligned}
\text{Farmer Group or Segement} &= y = f(Personalcharacteristics, Socioeconomicstatus,\\
&\quad Agronomicpractices, Productcharacteristics)\\
\textbf{Where:} &\\
\text{Personal characteristics} &= \text{age, gender, land size,}\\
\text{Socioeconomic Status} &= \text{Probability below \$2.45 international povertly line}\\
\text{Agronomic practices} &= f(\text{Resistant behaviour: recycling of seed,}\\
&\quad \text{Supportive bahaviour: use of fertilizer})\\
\text{Product characteritics} &= \text{benefit expected from yields}
\end{aligned}
\tag{3.1}
$$

Therefore, the resulting expected regression model was as follows

$$
\begin{aligned}
\text{Farmer Group or Segement} &= y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4\\
&\quad + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \epsilon\\
\textbf{Where:} &\\
x_1 &= \text{age}\\
x_2 &= \text{gender}\\
x_3 &= \text{land size}\\
x_4 &= \text{socioeconomic status as per}\\
&\quad \text{probability below \$2.45 international povertly line}\\
x_5 &= \text{recycling of seed}\\
x_6 &= \text{use of fertilizer}\\
x_7 &= \text{benefit expected from yields}\\
\epsilon &= \text{other unexplained factors or the residual error}\\
\text{And } \beta_1, \beta_2 \cdots \beta_7 &\text{ are the associated regression coefficients}
\end{aligned}
\tag{3.2}
$$

LCCA outputs a class membership probability model for each latent classes or segments found in the data. The regression coefficients, therefore, represent probability values of class membership for a given independent variable. In interpreting the model, therefore, influential independent variables will have regression coefficients greater than zero in magnitude with a positive or negative sign for type of relationship. Where a regression coefficient is zero, then that factor has no influence. On the other hand, positive regression coefficients of independent variables will indicate factors that support farmer choice or desired behaviour in consuming certified maize seed in of Western and Coastal Kenya, while negative regression coefficients will indicate farmer characteristics that are not supportive of consumption of certified maize seed in of Western and Coastal Kenya. Furthermore, the larger the regression coefficient the more influential the factor or independent variable is relative to the other independent variables. Additionally, p-value statistical tests for each of the regression coefficient will be computed to ascertain the significance of that independent variable on the overall model. Consequently, independent variables with positive regression coefficient and statistically significant p-values may be prioritized by an organization as the farmer characteristics for targeting of marketing mix strategies and service delivery. Conversely, independent variables with negative regression coefficients and statistically significant p-values may be treated as conditions the organization may mitigate for the farmer in order to drive consumption of certified maize seed in of Western and Coastal Kenya.

While the LCCA method does not rely on the modeling assumptions of linearity and normal distribution, chances are that it may converge on a local maximum solution as opposed to a global maximum, which is then not representative of the true nature of the data under investigation and, therefore, of the population estimate. In order to mitigate this, random starting values are used when initializing the model.

In preparation, the dummy table 3.1 on page 25 was designed to ensure that the data collected fitted with the variables required for analysis

Variable Class	Questions	Data class	Set of answers
Personal Characteristics			
Demographic	Gender	Nominal	Male, Female
	Age	Ordinal	Under 13, 13 – 17, 18 – 24, 25 – 30, 31 – 40, 41 – 50, 51 – 60, Above 60
	Land size	Ordinal	0.5 acre or less, 0.5 acre to 1 acre, 1 acre to 5 acre, 5 acre to 10 acre, Above 10 acres
Predictors			
Socioeconomic	Probability of falling below a given poverty line	Ordinal	None, Low, Moderate low, Moderate, Moderate high, high, extreme
Agronomic Practices	Resistant: Has recycled seed before	Nominal	Not recycled, Hybrid seed, OPV seed, Other seed
	Supportive: Uses fertilizer	Ordinal	No fertilizer, Most of the land, Half of the land, A small portion of the land
Product Benefit	Perceived certified seed yield benefit	Ordinal	Higher yields, Same yields, Lower yields

TABLE 3.1: Model variables and expected data from survey instrument

3.5 Research Quality

While the ultimate sampling unit was the individual participant, we intended to make the sample as representative as possible by conducting the pseudo-random selection of survey sites or agro-dealer sites. However, stratification was not necessary as there were no expected differences in the dynamics within the regions.

The surveys also ran until at least 312 fully completed records had been collected. This was meant to mitigate on non-response errors. Furthermore, to enhance and ensure data quality with regard to completeness, response consistency, data types and user input, the questionnaire was digitized and administered using mobile devices.

In addition, and given the quantitative nature of the survey, the questions were designed to be close ended and to require answers to be selected from a defined set of options. The defined set of options had been checked for completeness and representativeness.

The training of interviewers was intended to mitigate on systematic errors due to interviewer capabilities and to enhance the homogeneity of the interview team. Furthermore, as part of the training, the enumerators were expected to complete at least four test records and an error buffer was made of 40 records in the sample size of 312.

Moreover, the LCCA method is robust to outliers and irrelevant segmentation variables and applies a random selection of the cluster analysis initial seed to avoid selection bias.

3.6 Ethical Issues

Survey participants and interviewers were notified of the purpose of the research and their expected voluntary participation in the research. Furthermore, survey participant identities are not revealed in the data and research output; the data is coded.

Chapter 4

Data Analysis and Discussion

In this chapter, the results of the data analysis are presented. The data collected was in response to the research questions posed in chapter one and analyzed accordingly. The primary objective was to investigate the farmer characteristics or factors that would influence targeting. The associated sub-objectives were to define and profile the farmer segments or groups in the market to be targeted, and to investigate the significance of socioeconomic status as an influencing factor. The findings in the chapter demonstrate how these objectives were accomplished.

The R Project language and environment for statistical computing is used for the analysis of the data together with the Rmixmod package for finite mixture model cluster analysis using latent class variables (R Core Team, 2015).

This chapter first presents summary statistics on the independent variables as identified in the theoretical framework 2.1 and modeled in the data analysis equation 3.2. Next is an investigation of the latent class clusters to identify the number and characteristics of farmer groups in the target market as per research questions one and two. Finally, the relationships between the independent variables are further investigate for those independent variables that would influence targeting in the target market as per the research question three.

4.1 Summary Statistics

A description of the study variables and their associated data type is seen in Figure 4.1. The variables gender and resistant agricultural practice of recycling seed are nominal labels, while the other variables are ordered ordinal data types. For each variable, the length of 313 represents the number of complete observations used in the analysis.

```
---------- VARIABLE STRUCTURE ------------------

                      Length   Class   Mode
gender                   313    factor  numeric
age                      313    ordered numeric
land_size                313    ordered numeric
resistant_recyle         313    factor  numeric
supportive_fertilizer 313       ordered numeric
benefit_yields           313    ordered numeric
socioeconomnic_status 313       ordered numeric

-----------------------------------------------
```

FIGURE 4.1: Structure of the variables used

The data is qualitative; categorical nominal and ordinal scale data types are used. Qualitative data types are analyzed for frequencies and proportions as well as modal for central tendency and bar plot for distribution.

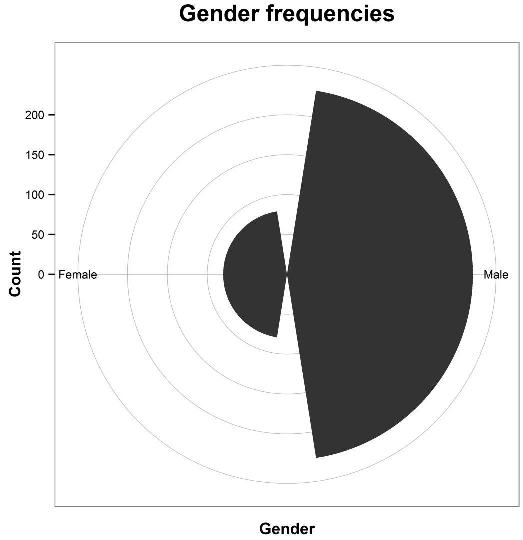
4.1.1 Personal Characteristics

Personal characteristics were used for accessibility of the segments but they may act as predictors in the final model for defining the segments. The personal characteristic variables were gender, age and land size.

4.1.1.1 Gender

Figure 4.2 tabulates and charts the frequencies and associated proportions for the gender variable. 74% of the respondents were male, while 26% were female. The data shows that the mode value is male.

This finding is consistent with other research, which has shown that more men than women own land and make decisions on farm production and finances at the household, and that, instead more women than men make up the farm labour, with women doing up to 70% to 80% of the agricultural work (Bernard et al., 2010; Livingston et al., 2011).



Gender Frequency Table

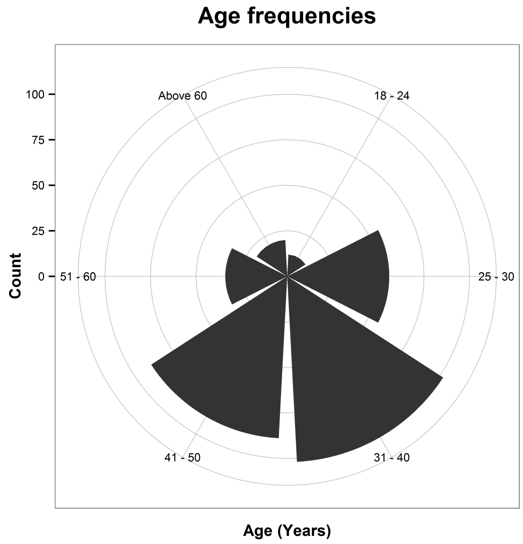
	Male	Female	Total
Count	233	80	313
% Proportion	74%	26%	100%

FIGURE 4.2: Frequencies and proportions for the variable gender

4.1.1.2 Age

From figure 4.3, 78% of the respondents were above 30 years old with the mode being ages 31 to 40 years at 33%. Mature adults between the ages of 31 and 50 made up 61% of the respondents, while the youth between ages 18 and 30 years constituted 22% of the respondents. The radial plot of the frequencies in 4.3 indicates that the distribution has one peak at ages 31 to 40. The observations increase from ages under 13, peak at ages 31 to 40 and then decrease towards ages above 60. The distribution is skewed towards the older age groups.

This is consistent with the growing concern that fewer and fewer youth are engaging in agriculture and they prefer to go to urban centers for employment, which is in turn stagnating rural development (Todaro & Smith, 2010).



Age Frequency Table

	Under 13	13 to 17	18 to 24	25 to 30	31 to 40	41 to 50	51 to 60	Above 60	Total
Count	0	0	12	56	102	89	34	20	313
% Proportion	0%	0%	4%	18%	33%	28%	11%	6%	100%

Age bands re-summarized

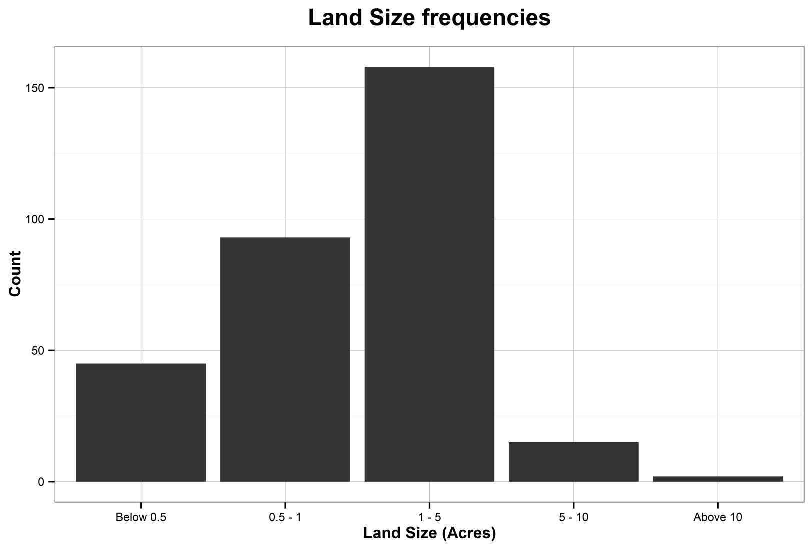
	Child	Young Adult	Mature Adult	Seniors	
	0 to 17 years	18 to 30 years	31 to 50 years	51 and Above	Total
Count	0	68	191	54	313
% Proportion	0%	22%	61%	17%	100%

FIGURE 4.3: Frequencies and proportions for the variable age

4.1.1.3 Land Size

Figure 4.4 shows that, in total, this band of zero to one acre of land constituted 44% of the responses, with the largest count being observed in the bucket 0.5 to one acre of land at 30%. Additionally, the mode for the entire sample is one to five acres of land at 50% of the respondents while only 6% of the respondents had more than five acres of land. From the bar plot of the frequencies in 4.4 the distribution has one peak at the one to five acres bin. However, a look at the The observations increase from below 0.5 acres, peak at one to five acres and then decrease towards above 10 acres. A breakdown of the one to five acre bin into sub-bins may exhibit a more gradual decrease in the distribution towards the above ten acre bin, indicating a more bell-shaped curve.

These results are consistent with the prior research in defining smallholder land ownership and the accompanying subsistence type of farming(World Bank, 2009; Todaro & Smith, 2010; Bernard et al., 2010; Livingston et al., 2011; Schroeder et al., 2013). Moreover, the proportion of farmers with large tracks of land of five acres and above is only 6%, which is further in line with other works in describing SSA, and Kenya in particular, as a smallholder farmer dominated market (World Bank, 2009; Todaro & Smith, 2010; Schroeder et al., 2013).



Land Size Frequency Table

Farmer cultivates maize on land of size

	Less than 0.5 acres	0.5 to 1 acre	More than 1 acre to 5 acres	More than 5 acres to 10 acres	More than 10 acres	Total
Count	45	93	158	15	2	313
% Proportion	14%	30%	50%	5%	1%	100%

Land size re-summarized

	Less than 1 acre output	More than 1 acre to 5 acres	More than 5 acresto 10 acres	More than 10 acres	Total
Count	138	158	15	2	313
% Proportion	44%	50%	5%	1%	100%

FIGURE 4.4: Frequencies and proportions for the variable land size

4.1.2 Socioeconomic Status

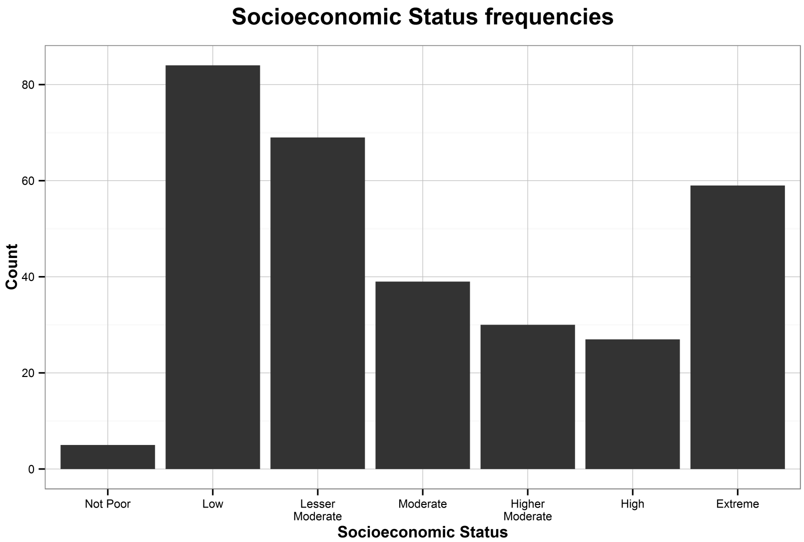
The study uses the progress out of poverty index (PPI) to measure for socioeconomic status. The PPI returns a probability value of falling below a particular poverty line. The study codifies this value into an ordinal scale variable as per figure 4.5. This study also uses poverty level or poverty likelihood to refer to this socioeconomic status.

SOCIOECONOMIC STATUS

$2.45/Day/PPP 2005 Probability value	Coded Value	Description
0	none	Not poor
1 - 15%	low	Low probability is poor
16 - 30%	low_mod	Lower moderately poor
31 - 45%	moderate	Moderately poor
46 - 65%	mod_hi	Higher moderately poor
66 - 80%	high	Very poor
81 - 100%	extreme	Extremely poor

FIGURE 4.5: Coding of socioeconomic status variable

The frequencies and proportions as seen in 4.6 indicate that 29% of the respondents have low or no (none) likelihood of being poor. On the other hand, 28% of the respondents are highly or extremely poor. A majority of the respondents, at 44%, have moderate levels of poverty. Only 2% of the respondents have no (none) or zero likelihood of being poor or socioeconomically at risk. The bar plot of the frequencies in 4.6 indicates that the distribution has two likely peaks at low socioeconomic status and extreme socioeconomic status. The observations decrease from the peak low poverty likelihood, deep at high poverty likelihood and then increase towards the peak extreme poverty likelihood.



Socioeconomic Status Frequency Table

Farmer probability below the international $2.45/Day/PPP 2005

	none	Low	Moderate low	Moderate	Moderate high	High	Extreme	TOTAL
Count	5	84	69	39	30	27	59	313
% Proportion	2%	27%	22%	12%	10%	9%	19%	100%

FIGURE 4.6: Frequencies and proportions for the variable socioeconomic status

Prior works describe smallholder farmers as socioeconomically at risk, which is consistent with the study findings (World Bank, 2009; Todaro & Smith, 2010). This research finds that 98% of the respondents do fall under the \$2.45/Day/PPP 2005 international poverty line. Interestingly though, this research finds that there are more respondents with lower levels of poverty likelihood than with high or extreme poverty likelihood; 49% have a 30% or less poverty likelihood compared to 28% who have a 66% or more poverty likelihood. This suggests that, the severity or intensity of the socioeconomic status risk is probably lower compared to the expectation of extreme poverty.

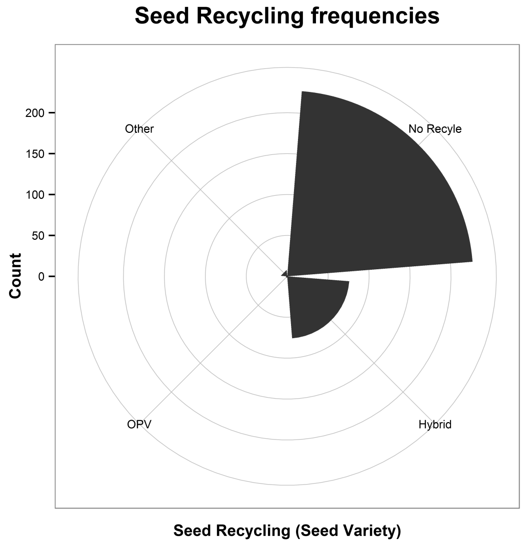
4.1.3 Agricultural Practices

Agricultural practices are measured for using the variables for seed recycling and fertilizer usage.

4.1.3.1 Seed Recycling

This study finds that a majority of the farmers using certified seed do not recycle it. Figure 4.7 shows that a majority of the respondents, at 73%, do not recycle seed. Hybrid seed type is the most recycled at 24% of the respondents compared to a total of 4% for OPV and other seed type. This is differs from other research that shows that, while certified maize seed market is most mature in Kenya, there are high rates of seed recycling (Bernard et al., 2010; Erenstein et al., 2011; Schroeder et al., 2013).

Additionally, there are higher proportions of recycled hybrid seed variety over OPV variety. This agrees with prior research, which indicates that while OPV varieties can withstand recycling but hybrid varieties should not be recycled, farmers do not distinguish that (Munyua et al., 2010; Bernard et al., 2010; Erenstein et al., 2011; Schroeder et al., 2013). Furthermore, hybrid seed may be inferred to be more popular with the farmers generally than OPV varieties resulting in the lower observations.



Seed Recycling Frequency Table

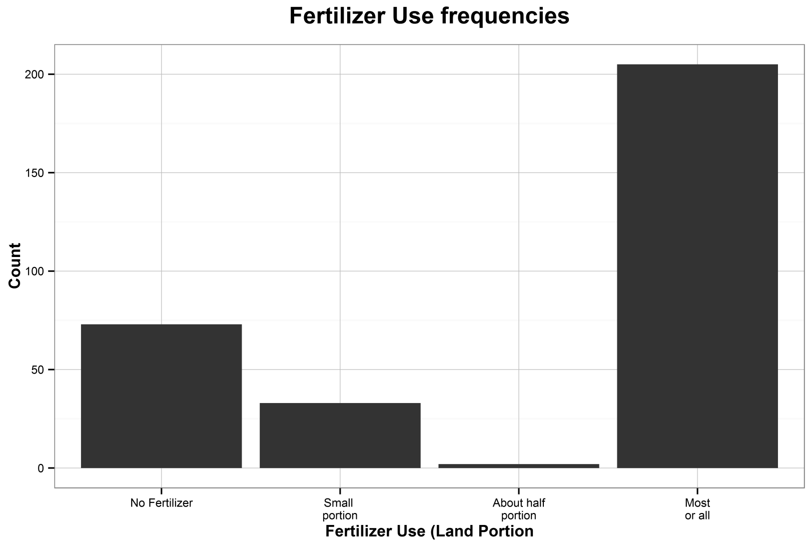
Farmer reused seed from previous crop of

	Not recyled	Hybrid seed	OPV seed	Other seed	Total
Count	227	76	2	8	313
% Proportion	73%	24%	1%	3%	100%

FIGURE 4.7: Frequencies and proportions for the resistant agricultural practice of recycling seed variable

4.1.3.2 Fertilizer Usage

Figure 4.8 shows that fertilizer usage seems to be high in the sample group with 65% of the respondents using fertilizer on most or all of their land and 23% not using fertilizer.



Fertilizer Use Frequency Table

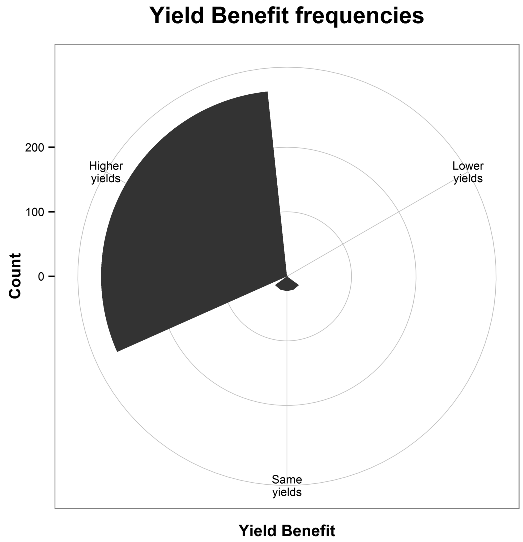
Farmer uses fertilizers amounts applicable to

	No fertilizer	A small portion of land	About half the land	Most or all the land	Total
Count	73	33	2	205	313
% Proportion	23%	11%	1%	65%	100%

FIGURE 4.8: Frequencies and proportions for the supportive agricultural practice of fertilizer usage variable

4.1.4 Product Characteristic

A large proportion of the respondents, at 92%, have experienced higher yields in farm output as a result of using certified maize seed compared to when they plant non-certified seed. Only 1% reported having experienced lower yields and 7% as having experienced no yield benefit. This is consistent with prior works on the yield value of using certified seed (Munyua et al., 2010; Bernard et al., 2010; Erenstein et al., 2011; Schroeder et al., 2013).



Yield Benefit Frequency Table

Farmers has experienced yield changes from use of certified seed

	Lower yields	Same yields	Higher yields	Total
Count	2	23	288	313
% Proprotion	1%	7%	92%	100%

FIGURE 4.9: Frequencies and proportions for the variable yield benefit

4.2 Cluster or Segmentation Results

A description of the clustering model paramaters used is shown in figure 4.10 below. Three criteria are used to select the appropriate model and number of clusters namely, the bayesian information criterion (BIC), integrated completed likelihood (ICL), which is a classification version of BIC, and the entropy criterion(NEC) method. The model then selects the criterion with the best fit. Additionally, since the data is qualitative, the parameter distribution is set to a multinomial distribution. Consequently, the models are estimated and tested using the maximum likelihood estimation approach using expectation maximization (EM) algorithms, which outputs a log-likelihood value. The best model identified is selected by comparing the log-likelihood values of each model in each iteration and for each criterion. Therefore, the model with the best fit, as per estimation maximization, has the largest log-likelihood value. The model also tests for the proportions. Figure 4.11 on page 38, shows the clustering model output. The best fit model is selected from the BIC criterion. There is theoretical and empirical support for the use of the

BIC model criterion, which has been found to be a consistent estimator and to perform better computationally and in handling categorical data (Gilligan, 2009).

```
**************************************
        MODEL SETUP
**************************************
* nbCluster =  2 3 4 5 6
* criterion =  BIC ICL NEC
**************************************
        MIXMOD Strategy:
**************************************
* algorithm            =  EM
* number of tries      =  1
* number of iterations =  200
* epsilon              =  0.001
*
*** Initialization strategy:
*
* algorithm            =  smallEM
* number of tries      =  50
* number of iterations =  5
* epsilon              =  0.001
* seed                 =  NULL
**************************************
```

FIGURE 4.10: The clustering model setup

```
****************************************
*** BEST MODEL OUTPUT:
*** According to the BIC criterion
****************************************
* nbCluster   =  2
* model name  =  Binary_pk_Ekjh
* criterion   =  BIC(4543.5673) ICL(4598.9378) NEC(0.4777)
* likelihood  =  -2119.5093
****************************************
* number of modalities =  2 8 5 4 4 3 7
*** Cluster 1
* proportion =  0.8549
* centre     =  1.0000 5.0000 3.0000 1.0000 4.0000 3.0000 2.0000
* scatter    = |     0.2069     0.2069 |
        |     0.0005     0.0005     0.0367     0.1677     0.6915     0.3064     0.1200     0.0598 |
        |     0.1285     0.2875     0.4766     0.0524     0.0082 |
        |     0.1556     0.1290     0.0009     0.0257 |
        |     0.1817     0.1056     0.0082     0.2955 |
        |     0.0087     0.0015     0.0102 |
        |     0.0168     0.7072     0.2396     0.1273     0.0872     0.0773     0.1590 |
*** Cluster 2
* proportion =  0.1451
* centre     =  2.0000 5.0000 3.0000 2.0000 1.0000 2.0000 7.0000
* scatter    = |     0.4523     0.4523 |
        |     0.0027     0.0027     0.0518     0.2413     0.5821     0.1497     0.0436     0.0904 |
        |     0.2343     0.3489     0.6159     0.0284     0.0043 |
        |     0.0157     0.0983     0.0484     0.0342 |
        |     0.4678     0.1104     0.0063     0.3512 |
        |     0.0072     0.4992     0.4919 |
        |     0.0168     0.1214     0.1062     0.1097     0.1480     0.1405     0.6427 |
****************************************
```

FIGURE 4.11: The clustering model output

4.2.1 Research Question 1: Number of Farmer Groups

The clustering model output in figure 4.11 on page 38 shows that two clusters are identified as having the best fit. The first cluster has a proportion value of 85% and the second cluster makes up 15%. The BIC model criterion has been found to be a consistent estimator of the correct number of latent classes in data (Gilligan, 2009).

Furthermore, Livingston et al. (2011) in their segmentation of smallholder farmers by value chain for credit access, they found two major groups as commercial and non-commercial and three relevant segments for their study as non-commercial, commercial in tight value chains and commercial in loose value chains, which was a further sub-grouping of the commercial farmers into two groups by type of value chain. This research, however, does not consider commercialization of farm production but may infer that there are at least two top-level groups of farmer.

The study does find that there is more than one viable group of farmers, which agrees with the hypothesis that there exist farmer groups within the larger group of smallholder farmer deserving of targeted treatment.

A plot of the two identified clusters is shown in figure 4.12. The red circles represent observations for cluster one, while the green triangles represent observations for cluster two.

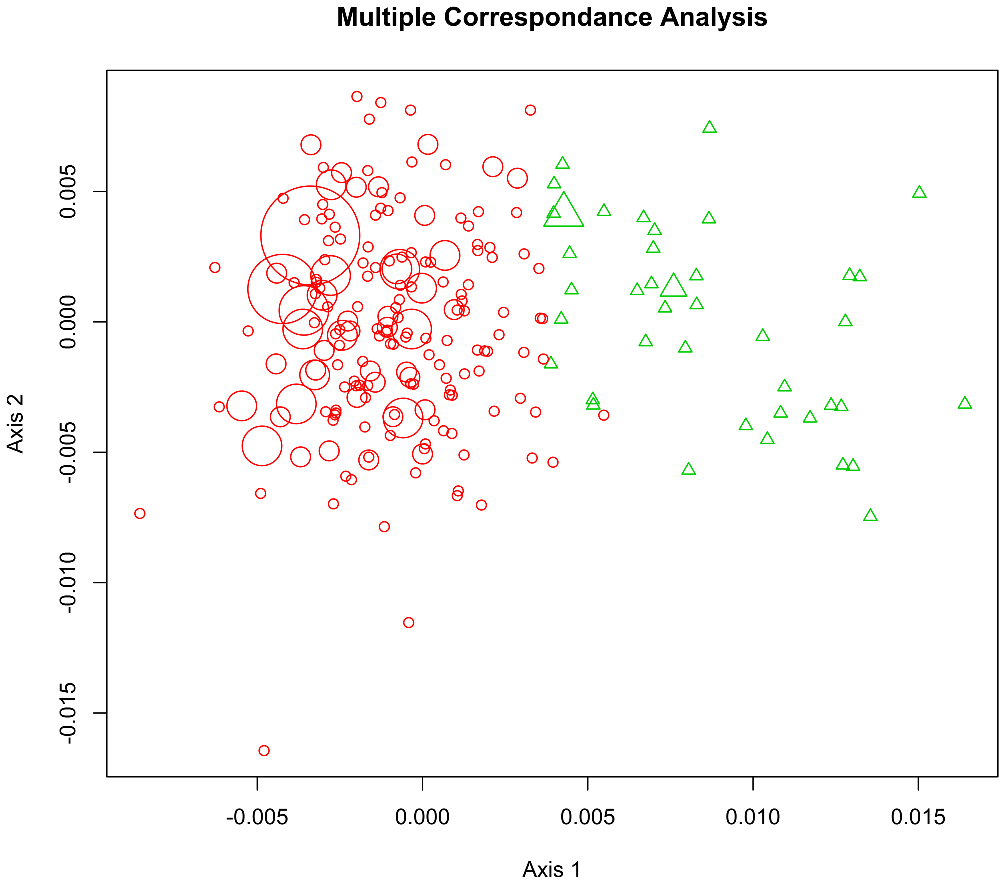


FIGURE 4.12: A plot of the clusters

4.2.2 Research Question 2: Characteristics of the Farmer Groups

Recall that the study theoretical framework is 2.1 with a resulting model of 3.2. Each cluster identified represents a farmer group in the study market, which is the dependent variable y. For each of the identified clusters, the model is then defined using the cluster center values, which are coded values for the qualitative variables modalities. These cluster centers are indicative of the defining characteristics in that cluster. Figure 4.13 shows what the coded values translate to for each of the variables. Additionally, figure 4.14 on page 42 plots the probabilities of the modalities for each of the clustering variables, therefore, showing the probability of a modality (variable response) belonging in a particular cluster. These probabilities are constitute the coefficients for each of the explanatory variables.

			Model Result center values	
Variable	**Modalities**	**Model code**	**Cluster 1**	**Cluster 2**
Gender	Male	1	1	2
	Female	2		
Age	Under 13 years	1	5	5
	13 to 17 years	2		
	18 to 24 years	3		
	25 to 30 years	4		
	31 to 40 years	5		
	41 to 50 years	6		
	51 to 60 years	7		
	Above 60	8		
Land Size	Less than 0.5 acres	2	3	3
	0.5 acre to 1 acre	3		
	1 acre to 5 acres	4		
	5 acres to 10 acres	5		
	More than 10 acres	6		
Seed recycling	Not recycled	1	1	2
	Hybrid seed	2		
	OPV seed	3		
	Other seed	4		
Fertilizer use	No fertilizer	1	4	1
	A small portion of the land	2		
	About half the land	3		
	Most of the land	4		
Yield benefit	Lower yields	1	3	2
	Same yields	2		
	Higher yields	3		
Socioeconomic status	None	1	2	7
	Low	2		
	Less moderate	3		
	Moderate	4		
	Higher moderate	5		
	High	6		
	Extreme	7		

FIGURE 4.13: Coding of the variables for clustering showing the center for each variable under each cluster identified

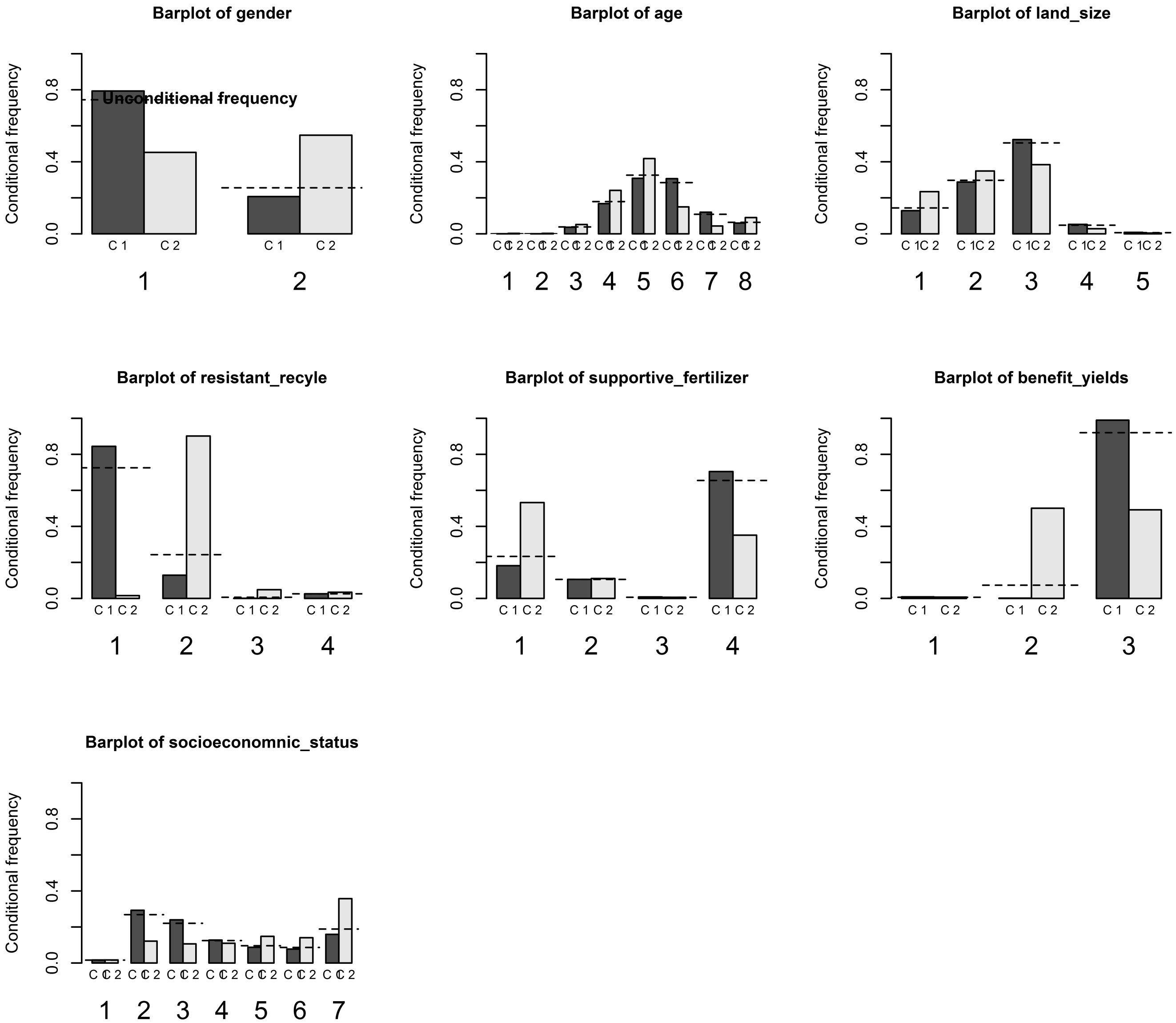


FIGURE 4.14: Barchart of the clustering variables

The model results, therefore, show that variable age and land size have the same center values in both cluster one and two, meaning that the two variables are not relevant in identifying the clusters. These two variables, however, may be used to access or enrich the description of the cluster. This is consistent with general segmentation theory that demographics add depth to the analysis but do not perform as well in partitioning the markets (Gilligan, 2009; Kotler, 2012). Furthermore, from the summary statistics, the demographic data variables suggest that the market is smallholder farmer dominated. Therefore, it is the non-demographic segmentation bases that would be more meaningful in further subdividing this market.

Consequently, the predictor variables in the clusters are seed recycling, fertilizer use, yield benefit and socioeconomic status as expected from the theoretical framework and resulting estimation model in models 2.1 and 3.2 respectively. Additionally, while gender is a demographic base, it is identified by the clustering model as an important factor in the partitioning of the data.

The identified clusters are further defined as follows, using the discovered predictor variables gender, seed recycling, fertilizer use, yield benefit and socioeconomic status.

4.2.2.1 Cluster One

The model results in figures 4.11 and 4.14, indicate that a farmer is likely to belong to cluster one group if he or she exhibits the following characteristics

1. The farmer is a male,
2. The farmer does not recycle seed,
3. The farmer uses fertilizer on most or all of the land,
4. The farmer observes high yield benefit from using certified seed
5. The farmer is likely to experience low levels of poverty or socioeconomic status. Using the coding table of socioeconomic status in figure 4.5, this translates to a poverty likelihood value of 1% to 15% under the $2.45/Day/PPP 2005 international poverty line. Additionally, from the bar plot, members in this group may also experience lesser moderate levels of poverty likelihood. As a result, from the coding table of socioeconomic status in figure 4.5, the farmer would have a poverty likelihood value of at most 30% under the $2.45/Day/PPP 2005 international poverty line.

Additionally, from the bar plot 4.14, members of this group are likely to be of ages 25 years to 40 years. Members are also likely to have land of size one to five acres.

4.2.2.2 Cluster Two

The model results in figures 4.11 and 4.14, indicate that a farmer is likely to belong to cluster two group if he or she exhibits the following characteristics

1. The farmer is a female,
2. The farmer recycles hybrid seed variety,
3. The farmer does not use fertilizer,
4. The farmer observes no change in yields from using certified seed
5. The farmer is likely to experience extreme levels of poverty or socioeconomic status. Using the coding table of socioeconomic status in figure 4.5, this translates to a poverty likelihood value of 81% to 100% under the $2.45/Day/PPP 2005 international poverty line. Additionally, from the bar plot, members in this group may also experience higher moderate levels or high levels of poverty likelihood. As a result, from the coding table of socioeconomic status in figure 4.5, the farmer would have a poverty likelihood value of 46% at the very least under the $2.45/Day/PPP 2005 international poverty line.

Additionally, from the bar plot 4.14, members of this group are likely to be of ages 41 years and above, and to have land of sizes one acre or less.

4.2.2.3 Summary on the Clusters

Cluster one signals an empowered group of smallholder farmers with the desired or positive agricultural practices for the use of certified maize seed for farm productivity. Socioeconomic risk is the main challenge for this group. This group is seemingly compelled and enabled to use certified seed.

Cluster two signals a challenged group of smallholder farmers that is significantly socioeconomically at risk and has undesired or negative agricultural practices for the use of certified maize seed for farm productivity. This group needs to be first compelled and empowered to properly use the certified seed.

4.3 Research Question 3: Factors that Influence Targeting

The cluster centers in figure 4.13 on page 41 and the associated conditional bar plots in figure 4.14 on 42 show that variables age and land size are not influential predictors in determining the farmer groups.

The clustering model identifies gender, socioeconomic status, seed recycling, fertilizer use and yield benefit as influential factors in belonging to a given farmer group. Consequently, in evaluating for those farmer factors that influence targeting in the certified maize seed of Western and Coastal Kenyamarket, the independent variables socioeconomic status, seed recycling, fertilizer use and yield benefit are further evaluated in comparison to each other and against gender.

Mosaic plots are used to visualize the area proportional of the variables contingency tables and therefore show the relationship between the categorical variables. In particular, an extended mosaic plot visualization from the vcd R package is used in this study, which has the added advantage of visualizing fit using residual based shading. The maximum residual shading in these mosaic plots, highlights cells where the associated residual exceeds critical values of the maximum tests, which are at 90% and 99% by default, and therefore, highlights variables with association. To interpret the mosaic plots, therefore, the plots with colored shading show variables with association. This is also supported by the smaller p-values in such mosaic plots.

The associated analysis of the relationships between the variables are stated as follows and a combined discussion is provided in the summary of the influential factors subsection.

4.3.1 Socioeconomic Status

Figure 4.15 on page 46 has four mosaic plots, one each for the relationship between socioeconomic status and the other four variables identified by the clustering model as influential. As per the residual-based maximum shading, socioeconomic status has an association with variable fertilizer use but no significant association with the other variables.

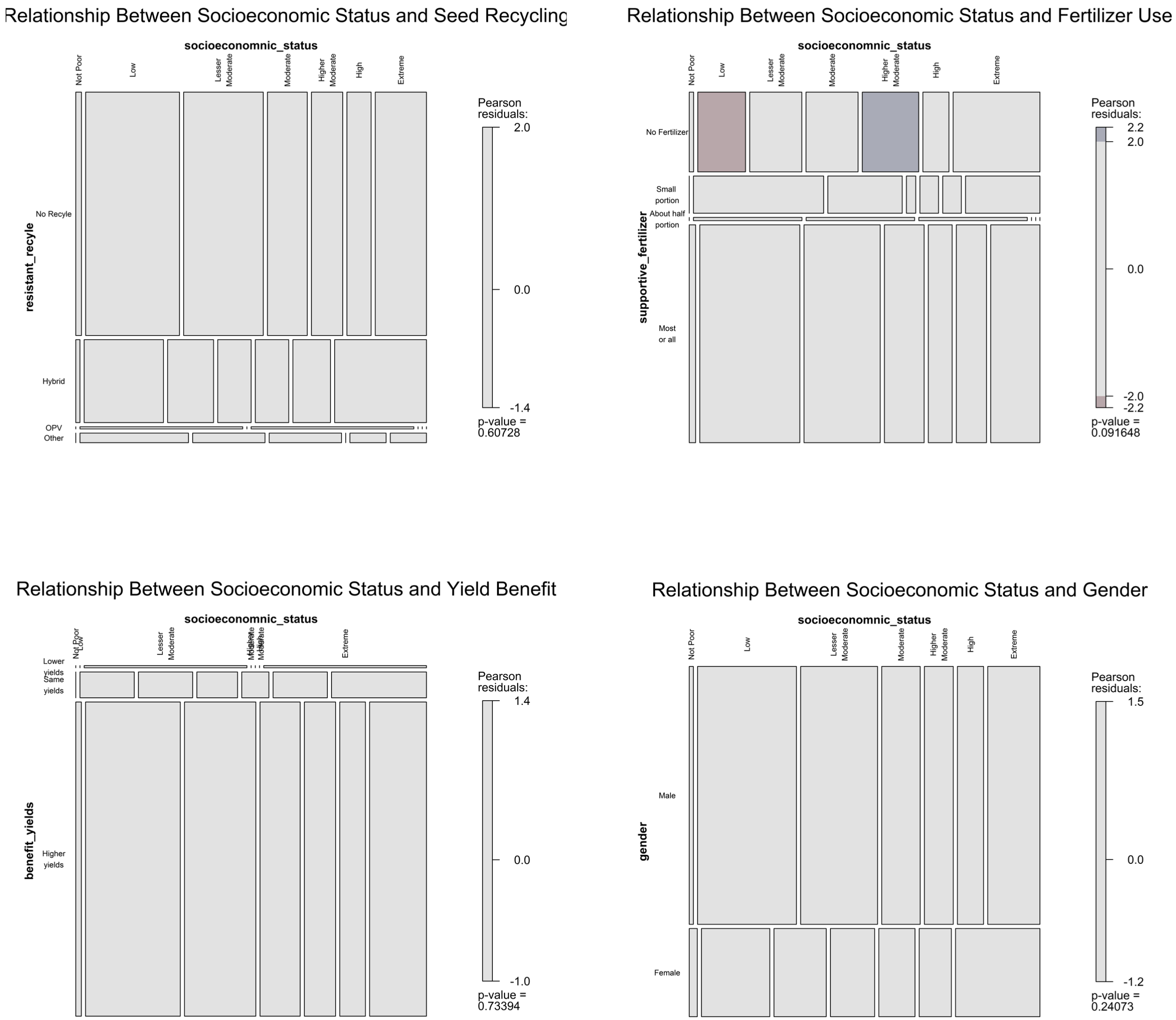


FIGURE 4.15: Grouping of socioeconomic status by other variables

4.3.2 Agricultural Practices

4.3.2.1 Fertilizer Use

In addition to the previous mosaic plot, figure 4.16 on page 48 shows the relationship between fertilizer use and the remaining variables seed recycling, yield benefit and gender. As per the residual-based maximum shading, fertilizer use has an association with variable seed recycling but not with yield benefit or gender.

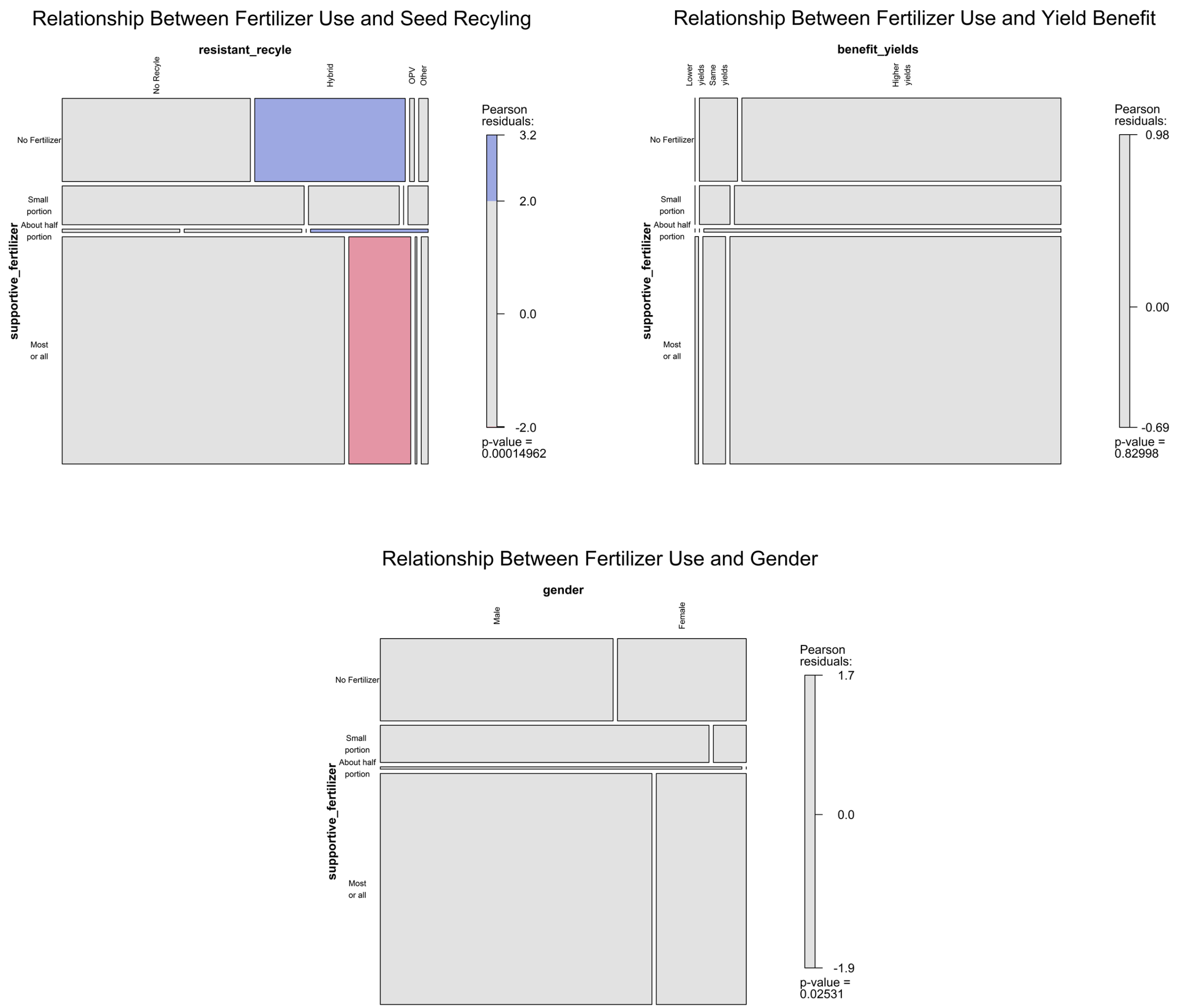


FIGURE 4.16: Grouping of fertilizer use by other variables

4.3.2.2 Seed Recycling

In addition to the previous mosaic plots, figure 4.17 on page 49 shows the relationship between seed recycling and the remaining variables yield benefit and gender. As per the residual-based maximum shading, seed recycling has an association with variable yield benefit.

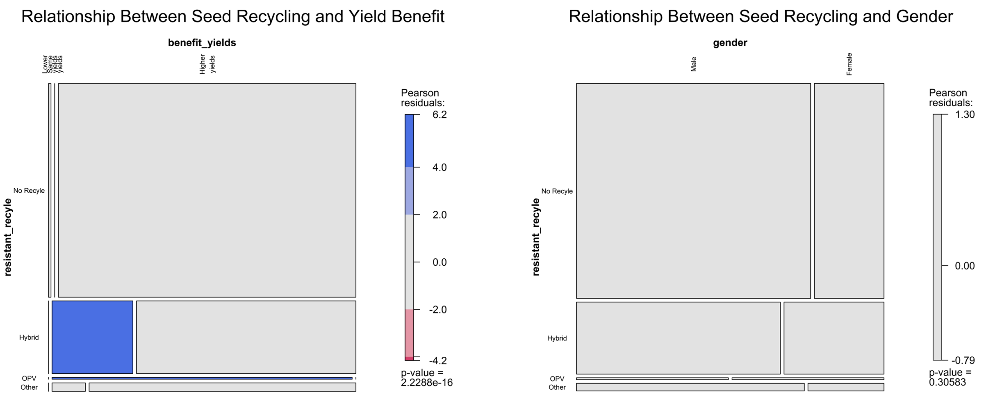


FIGURE 4.17: Grouping of fertilizer use by other variables

4.3.3 Product Benefit

In addition to the previous mosaic plot, figure 4.18 below shows the relationship between yield benefit and the remaining variable gender. As per the residual-based maximum shading, yield benefit has an association with variable gender.

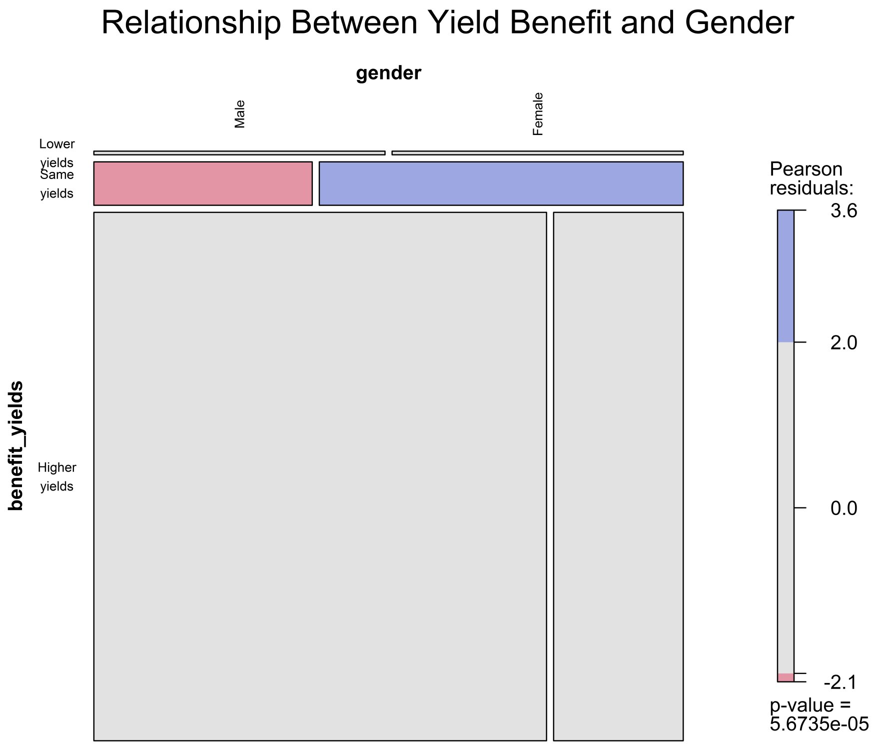


FIGURE 4.18: Yield benefit by gender and by age

4.4 Summary of Influential Factors

Form the figures, the color intensity of the shading in the mosaic plots indicates the statistical significance of the relationship between the variables, which is inline with the p-values. Consequently,

1. While there is a relationship between socioeconomic status and fertilizer use, the shading for this relationship has the least colour intensity; there is more gray than actual red or blue in the shading. This is, therefore, the weakest association found between all the variables.

2. The relationships between fertilizer use and seed recycling, and between yield benefit and gender have similar shading and residual values.

3. The richest colour intensity, and therefore the strongest association, is found in the relationship between seed recycling and yield benefit.

The study finds the following. Firstly, among the demographic variables only gender distinguishes the clusters found, which is inline with other findings that demographic or personal characteristics are generally not as good predictors as behavioural characteristics (Alexander et al., 2005). Furthermore, consistent with prior works, gender remains a factor in farm productivity particularly with respect to yield benefit of certified maize seed (Bernard et al., 2010; Mathenge et al., 2014). This is also consistent with the study hypothesis that land size, while and identifying factor of a smallholder farmer, it is not an influential factor in how farmers group in the market.

Secondly, the recycling of hybrid seed is also observed in this study as a significant characteristic in the segmentation. Seed recycling is an undesired characteristic given that seed determines the upper limit of yields, and particularly for the hybrid variety (Livingston et al., 2011; Munyua et al., 2010; Bernard et al., 2010; Olwande & Smale, 2012; Mathenge et al., 2014). However, OPV variety is designed for at least one re-use (Munyua et al., 2010; Bernard et al., 2010; Erenstein et al., 2011; Schroeder et al., 2013). Farmers may be recycling to save on costs or because they do not have knowledge on seed in general or the right knowledge on seed varieties and their application.However, as seen in previous section, a bigger proportion of the farmers do not recycle seed. Additionally, this study does not find a significant association between socioeconomic status and seed recycling to suggest lack of resources in acquiring certified seed that would lead to significant seed recycling. This is unlike expectations in other works that smallholder farmers are resource constrained and invest less in agriculture (Pfitzer et al., 2009; Munyua et al., 2010; Bernard et al., 2010; Todaro & Smith, 2010; Schroeder et al., 2013).

Thirdly, prior works identify fertilizer as a big component of agricultural inputs costs for smallholder farmers and farmers will tend to skimp on it or opt between it and other inputs (). In line with that, this study finds that there is a significant association between socioeconomic status and fertilizer use.

Fourthly, research indicates greater productivity benefits when certified seed is used with other inputs such as fertilizer (Pfitzer et al., 2009; Bernard et al., 2010; Erenstein et al., 2011; Feeney et al., 2011; Livingston et al., 2011; Roucan-Kane et al., 2011; Schroeder et al., 2013). Accordingly, this research finds a significant association between fertilizer use and recycling of seed.

Fifth, prior works find that price and transaction costs are a significant factor in segmenting, and targeting, in agricultural inputs markets or agricultural credit market, both in smallholder farmer markets and in the more developed farmer markets (Alexander et al., 2005; Feeney et al., 2011; Christen & Anderson, 2013; Bernard et al., 2010). Surprisingly, though, while this research does find that socioeconomic status is a predictor in the segmentation, socioeconomic status is not found to be a stronger influencing farmer factor in the study market than the other variables. This is in contrast with prior works expectations that socioeconomic risk is influential in smallholder farmer economic decisions (Todaro & Smith, 2010; Pfitzer et al., 2009). However, in the cluster profiles, the two ends of socioeconomic status levels, low and extreme poverty likelihood, are defining characteristics of the identified farmer groups. All in all, this finding supports the hypothesis that socioeconomic status is a significant factor in the farmer segments.

Lastly, research indicates that farmers are likely to adopt agricultural technologies if the benefits are demonstrated (Pfitzer et al., 2009; Alexander et al., 2005; Feeney et al., 2011; Christen & Anderson, 2013; Bernard et al., 2010). This study finds the relationship between yield benefit and gender as most influential, and together with the significant relationship of yield benefit and seed recycling as well, it could be inferred that gender is a hindrance in realizing yield benefits and particularly due to the practice of recycling seed.

Chapter 5

Conclusion and Recommendations

In chapter one of this study, the problem under investigation was presented and the research objectives articulated. A subsequent thematic literature review was presented in chapter two and a theoretical framework for the analysis posed. Chapter three then details the research methodology used for data design and collection. An analysis and discussion of the data and associated findings is then reviewed in chapter four. This chapter, therefore, entails the conclusions and recommendations that can be made from the research results in addressing the research objectives.

The study finds that the market is dominated by smallholder farmers at 98% of the respondents. Furthermore, the study finds that there are two distinct types of smallholder farmer groups; one group is seemingly more able and moving towards intensification of farm production while the other group seems to be struggling with socioeconomic risk and bad agricultural practices. The first group is referred to as the empowered smallholder farmer while the second group is referred to as the challenged smallholder farmer. Fortunately, the later group, the challenged smallholder farmer, is small at 15% proportion compared to the more able group, the empowered farmer. Looking at the farmer group proportion of 85% for the empowered farmer, it may apparently suggest a viable market that does not call for much consideration of social and extension services in addition to financial and market strategies. However, socioeconomic status remains a significant factor and the farmers continue to experience varying degrees of poverty. Furthermore, the challenged smallholder farmer, while making up a smaller proportion, presents a wider range for potential action and subsequent impact on socioeconomic status, intensification and farm productivity.

In addition, this study finds that behaviour and activities of farmers make for better segmentation variables than demographic characteristics. Furthermore, gender and socioeconomic status continue to be significant farmer factors in how farmers group in this research as well. Moreover, negative or undesired agricultural practices are found to be significant factors in how farmers group in the study market and more influential, therefore, in targeting than positive or supportive agronomic practices.

5.1 Conclusion

This study set out to identify those farmer factors that influence targeting in the target market certified maize seed of Western and Coastal Kenya. It also had the sub-objectives of defining the segments or farmer groups in the target market and the role of socioeconomic status as a factor influencing targeting. This study answered the research questions and met the research objectives that were set out at the beginning of the research.

The results of this study were similar to other works in most instances, both with theoretical and empirical research. The following conclusions, therefore, are based on the thematic issues in the literature review in chapter two and/or the data analysis results in chapter four in meeting the research objectives and the hypothesis.

First, as per segmentation theory, while personal or demographic characteristics are useful in accessing and enriching the descriptions of the segments, they are found not to be stronger partitioning variables for segment membership (Gilligan, 2009; Kotler, 2012). Instead, behaviour or characteristics alluding to motivation for use partition the market. The study finds age and land size to be weak or non-influential farmer factors in the segmentation.

Second, while the study results support that the study market is dominated by smallholder farmers, it also identifies two distinct groups of smallholder farmers in that study market. This addresses the first research question on the number of farmer groups exist in the target market and supports the hypothesis that there exist more than one group of smallholder farmers in the study market.

Third, in addressing the second research question regarding the description of the identified study groups, the study characterizes the two segments or farmer groups as empowered farmer and challenged farmer. An empowered farmer is identified as experiencing low levels of socioeconomic risk and undertaking positive agricultural practices of not recycling seed and of using fertilizer. A challenged farmer is identified as experiencing high levels of socioeconomic risk and undertaking negative agricultural practices of recycling seed and not using fertilizer.

Fourth, in order to target these farmer groups and address the primary objective and the third research question, the study finds the farmer factors gender, socioeconomic status, recycling of seed, use of fertilizer and yield benefits experiences as influential.

Fifth, as per the hypotheses, socioeconomic status is found to be a significant factor, with the empowered farmer group being on the lower end of the poverty likelihood spectrum and the challenged farmer being on the higher end of poverty likelihood.

Sixth, it is found that fertilizer use and seed recycling are associated and that the resistant agricultural practices of seed recycling and failure to use fertilizer are more influential in targeting than supportive or positive agricultural practices. This is in support of the hypothesis that resistant agronomic practices are more influential factors for targeting than supportive agronomic practices.

Seventh, as hypothesized, land size is not found to be influential in how farmers group in the market.

Lastly, in line with other empirical findings in smallholder agriculture, gender continues to be a significant farmer factor in this study as well. However, gender is only influential in yield benefits and related recycling of hybrid seed varieties, suggesting a knowledge or skill gap in the women on the proper application of certified seed for yield benefit returns.

5.2 Recommendations

This section recommends how smallholder markets should be targeted based on farmer factors or characteristics in the certified maize seed market of Western and Coastal Kenya

Firstly, smallholder farmers be treated as belonging to either of the two farmer groups; empowered smallholder farmer and challenged smallholder farmer. The first group of farmers exhibits agricultural practices and product benefit experiences that support continued used and investment in certified seed. For the second group however, there is higher levels of socioeconomic risk and negative agricultural practices that need to be addressed first in targeting approaches for continued use and investment in certified seed to be addressed.

Secondly, a greater proportion of the empowered smallholder farmer, however, should not imply less or discontinued social and extensions services to smallholder farmers. Instead, that proportion should be indicative of where intensity of social and extension services should be directed at; more intensive support for the challenged smallholder farmer and reinforcement and continued support for the empowered farmer.

Thirdly, initiatives for the empowered smallholder farmer should be to maintain gains made in their positive agricultural practices and to further move them out of socioeconomic risk. Marketing communication for that group should reinforce the positive behaviour they already exhibit and financial initiatives would be a primarily focus.

Lastly, initiatives for the challenged smallholder farmer should focus on building behaviour and agricultural practices that drive adoption and proper application of the agricultural inputs certified maize seed and fertilizer.

References

Alexander, C. E., Wilson, C. a. & Foley, D. H. (2005). Agricultural Input Market Segments : Who Is Buying What ? *Agricultural Economics*, *2*(Fall), 113–132.

Bernard, M., Hellin, J., Nyikal, R. & Mburu, J. (2010). Determinants for use of certified maize seed and the relative importance of transaction costs. *Agricultural Economics*.

Bisolutions.us. (2014, 25 October). Cluster analysis vs market segmentation. Retrieved October 25, 2014, from http://www.bisolutions.us/Cluster-Analysis-vs.-Market-Segmentation.php

Borchers, B., Roucan-Kane, M., Alexander, C. E., Boehlje, M., Downey, W. S. & Gray, A. W. (2012). How large commercial producers choose input suppliers: expendable products from seed to animal health. *International Food and Agribusiness Management Review, International Food and Agribusiness Management Association (IAMA)*, *15*, 2.

Burns, A. C. & Bush, R. F. (2003). *Determining the Size of a Sample* (4th ed.). Pearson Prentice Hall.

CGAP, Ford, EU & Taskforce, S. (2010). POVERTY TARGETING AND MEASUREMENT TOOLS IN Progress out of Poverty Index and the Poverty Assessment. (October).

Christen, R. P. & Anderson, J. (2013, April). Segmentation of Smallholder Households: Meeting the Range of Financial Needs in Agricultural Families. *CGAP FocusNote*, (85).

Coleman, T. & Spellberg, J. (2007). *2007 global poverty and micro-fiance incidence map*. Progress out of Poverty, Grameen Foundation.

Dean, N. & Raftery, A. E. (2010). Latent class analysis variable selection. *Annals of the Institute of Statistical Mathematics*, *62*(1), 11–35. doi:10.1007/s10463-009-0258-9

DFID. (2005). Growth and poverty reduction: the role of agriculture. Department for International Development (DFID). Retrieved from http://dfid-agriculture-consultation.nri.org/launchpapers/roleofagriculture.pdf

Dolnicar, S. (2003). Using cluster analysis for market segmentation - typical misconceptions , established methodological weaknesses and some recommendations for improvement. *11*(2), 5–12.

Ele-Ojo, A. J., Eme, I. H. & Fonta, W. M. (2013). Multidimensional poverty assessment: applying the capability approach. *International Journal of Social Economics*, *40*(4), 331–354. doi:10.1108/03068291311305017

Erenstein, O., Kassie, G. T. & Mwangi, W. (2011). Challenges and opportunities for maize seed sector development in eastern Africa. (pp. 1–15). Retrieved from http://addis2011.ifpri.info/files/2011/10/Paper_2B_Olaf-Ernestein.pdf

Everitt, B. S., Landau, S., Leese, M. & Stahl, D. (2011). *Cluster Analysis*. doi:10.1007/BF00154794

FAO. (2010). *How to feed the world in 2050*.

Feeney, R. & Berardi, V. (nodate). seed market segmentation: how do argentine farmers buy their expendable inputs?

Feeney, R., Berardi, V. & Steiger, C. (2011, June). Agricultural Input Market Segmentation in Argentina : How do Argentine farmers buy their expendable inputs ? The Case of the Seed Industry, 1–22.

Gilligan, C. (2009). *Market Segmentation , Targeting*. Amsterdam Boston London: Elsevier/Butterworth-Heinemann.

Gloy, B. a. & Akridge, J. T. (1999). Segmenting the commercial producer marketplace for agricultural inputs. *International Food and Agribusiness Management Review*, *2*(2), 145–163. doi:10.1016/S1096-7508(00)00023-9

Gonzalez, V. (2006). Is micro-finance reaching the poor? *An Overview of Poverty Targeting Methods*. Retrieved from http://www.globenet.org/archives/web/2006/www.globenet.org/horizon-local/ada/c18.html

Guenette, P. (2007). The importance of input supply to value chain performance. *In World Report Fall 2006: The Value Chain Approach; Strengthening Value Chains to Promote Economic Opportunities*, 6–7.

Hennig, C. & Liao, T. F. (2013). How to find an appropriate clustering for mixed type variables with application to socio-economic stratification. *Applied Statistics*, *62*(3), 1–25.

Hunt, S. D. & Arnett, D. B. (2004). Market segmentation strategy, competitive advantage, and public policy: Grounding segmentation strategy in resource-advantage theory. *Australasian Marketing Journal*, *12*(1), 7–25. doi:10.1016/S1441-3582(04)70083-X

Iacobucci, D. & Churchill, G. (2009). *Marketing research: methodological foundations* (10th ed.). Cengage Learning.

Juma, C. (2011). *The new harvest*.

Jung, T. & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*(1), 302–317. doi:10.1111/j.1751-9004.2007.00054.x

Karnani, B. A. (2007). *Microfinance misses its mark*. Stanford Social Innovation Review.

Kotler, P. (2012). *Principles of marketing*. Boston: Pearson Prentice Hall.

Livingston, G., Schonberger, S. & Delaney, S. (2011). Sub-Saharan Africa : The state of smallholders in agriculture.

Magidson, J. & Vermunt, J. K. (2002). A nontechnical introduction to latent class models. *15*.

Mathenge, M. K., Smale, M. & Olwande, J. (2014). The impacts of hybrid maize seed on the welfare of farming households in kenya. *Food Policy*, *44*, 262–271. doi:10.1016/j.foodpol.2013.09.013

Muhammad, L., Njoroge, K., Bett, C., Mwangi, W., Verkuijl, H. & De Groote, H. (2003, November). *The seed industry for dryland crops in eastern kenya*. CIMMYT. Retrieved from http://books.google.com/books?hl=enandlr=andid=dhAPlaQY46ICandoi=fndandpg=PR4anddq=%22security+in+Kenya.+However,+these+impacts+have+not+been+replicated+in+the+semi-arid+midlands+due%22+%22The+study+found+that+the+low+quantity+of+seed+traded,+high+cost+of+production,+and+high+seed%22+andots=6CAHIKW10xandsig=SXOPsssc_q_5NihhHo3P3gMC6WY

Munyua, B. G., Jon, H., Nyikal, R. & Mburu, J. (2010). Determinants for use of certified maize seed and the relataive importance of transaction costs. Cape Town, South Africa.

Nyoro, J., Kirimi, L. & Jayne, T. S. (2004). *Competitiveness of kenyan and ugandan maize production: challenges for the future* (Working Papers KE-TEGEMEO-WP-10 No. 10). Nairobi: International Development Collaborative Working Papers KE-TEGEMEO-WP-10, Department of Agricultural Economics, Michigan State University. Retrieved from http://fsg.afre.msu.edu/kenya/wp9.pdf

Olwande, J. & Smale, M. (2012, August). Is Older Better? Maize Hybrid Change on Household Farms in Kenya. *International Association of Agricultural Economists*, 18–24. Retrieved from http://ageconsearch.umn.edu/bitstream/126669/2/Olwande.pdf

Ouma, J. O., Murithi, F. M., Mwangi, W., Macharia, G., Verkuijl, H. & De Groote, H. (2002, October). *Adoption of Maize Seed and Fertilizer Technologies in Embu District , Kenya Adoption of Maize Seed and Fertilizer Technologies in Embu District , Kenya*. CIMMYT.

Pfitzer, M., Krishnaswamy, R. & Genier, C. (2009, May). *Market Development Investments by Agricultural Input Companies and their Foundations: Transforming Smallholder Agriculture*. Syngenta Foundation for Sustainable Agriculture.

Progress out of Poverty. (2010). *Poverty targeting and measurement tools in microfinance: progress out of poverty index and the poverty assessment tool*. Ford Foundation, CGAP and Social Performance Task Force.

R Core Team. (2015). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from http://www.R-project.org/

Roucan-Kane, M., Alexander, C., Boehlje, M. D., Downey, S. W. & Gray, A. W. (2011). Large commercial producer market segments for agricultural capital equipment. *International Food and Agribusiness Management Review, International Food and Agribusiness Management Association (IAMA)*, *14*(4), 127–144.

Salami, A., Kamara, A. B., Brixiova, Z. & Bank, A. D. (2010, April). Smallholder Agriculture in East Africa: Trends, Constraints and Opportunities. *Working Paper No.105*.

SAS. (2012). *A markets's guide to analytics - using analytics to make smarter marketing decisions and maximize results*. SAS Institute for Advanced Analytics.

Schroeder, C., Onyango, T. K., Ranabhat, N., Jick, N. A., Parzies, H. K. & Gemenet, D. C. (2013). Potentials of hybrid maize varieties for smallholder farmers in kenya: a review based on swot analysis. *African Journal of Food, Agriculture, Nutrition and Development*, *13*, 2. Retrieved from http://www.ajol.info/index.php/ajfand/article/view/87480

Sebstad, J. & Cohen, M. (2000). *Microfinance, risk management, and poverty*. Assessing the Impact of Microenterprise Services (AIMS).

Smale, M., Nagarajan, L., Diakite, L., Audi, P., Grum, M., Jones, R. & Weltzien, E. (2011). Seed and fertilizer markets. *62*.

The Iris Center. (2011). *Manual for the implementation of usaid poverty assessment tools*. The IRIS Center.

Todaro, M. P. & Smith, S. C. (2010). *Economic development* (11th ed.). Pearson.

Tuma, M. & Decker, R. (2013). Finite mixture models in market segmentation: A review and suggestions for best practices. *Electronic Journal of Business Research Methods*, *11*(1), 2–15.

Ulwick, A. W. (2003). The Strategic Role of Customer Requirements in Innovation. *Strategyn inc*, *13*(January), 12. Retrieved from http://www.marketing4entrepreneurs.org/images/TheStrategicRoleofRequirementinInnovation.pdf

Vermunt, J. K. & Magidson, J. (2002). *Latent class cluster analysis*.

Wedel, M. & Kamakura, W. (2000). *Market segmentation*. Boston, MA: Springer US.

World Bank. (2008). *World development indicators 2008*. World Bank Publications. doi:10.1596/978-0-8213-7386-6

World Bank. (2009). *Ida at work : agriculture - an engine for growth and poverty reduction*. Washington, DC: World Bank Publications. Retrieved from http://documents.worldbank.org/curated/en/2009/09/11457996/ida-work-agriculture-engine-growth-poverty-reduction

Yunus, M. & Weber, K. (2009). *Creating a world without poverty: social business and the future of capitalism*. PublicAffairs.

Zeller, M. (2004). Poverty assessment tools. *Social Performance Map*, *48*, 180–185.

Appendix A - Study Questionnaire

Certified Maize Seed Market Segmentation Questionnaire

A. Instructions

- Italicized paragraphs are instructions for the enumerators, which are included in the digitized questionnaire for the enumerator to reference while interviewing respondents.

B. Context Questions

These questions identify the context of the interview and are answered by the enumerator before starting the respondent interview.

1. Enumerator initials

Other data collected in background by the mobile data collection tool for context and accountability purposes include

- Timestamp (start of interview, end of interview, submission of record)
- GPS coordinates of interview location

C. Personal Characteristics Questions

Before starting the survey interview, introduce yourself and the survey:

"Hello. I am [your name] and work with {a research company} and I would like to take a few minutes of your time to ask you some questions which will help to see how our products and services are helping you. Don't worry; your answers will be treated with confidentiality. Would you like to participate?

Thank you. I will ask you a few questions about yourself and your home and then some questions about seeds for planting…"

1. What is your full name?

2. What is your gender?

   Male Female

3. What is your age?
   a. Under 13
   b. 13 to 17 years
   c. 18 to 24 years
   d. 25 to 30 years
   e. 31 to 40 years
   f. 40 to 50 years
   g. 51 to 60 years
   h. Above 60

4. What is the size of your land?
   a. 0.5 acre or less
   b. 0.5 acre to 1 acre
   c. 1 acre to 5 acre
   d. 5 acre to 10 acre
   a. Above 10 acres

D. Socioeconomic Status Questions

5. How many members does the household have?
   a. Nine or more
   b. Seven or eight
   c. Six
   d. Five
   e. Four
   f. Three
   g. One or two

6. What is the highest school grade that the female head/spouse has completed?
   a. None, or pre-school
   b. Primary standards 1 to 6
   c. Primary standard 7
   d. Primary standard 8, or secondary forms 1 to 3

e. No female head/spouse
f. Secondary form 4 or higher

7. What kind of business (type of industry) is the main occupation of the male head/spouse connected with?
    a. Does not work
    b. No male head/spouse
    c. Agriculture, hunting, forestry, fishing, mining, or quarrying
    d. Any other

8. How many habitable rooms does this household occupy in its main dwelling (do not count bathrooms, toilets, storerooms, or garage)?
    a. One
    b. Two
    c. Three
    d. Four or more

9. The floor of the main dwelling is predominantly made of what material?
    a. Wood, earth, or other
    b. Cement, or tiles

10. What is the main source of lighting fuel for the household?
    a. Collected firewood, purchased firewood, grass, or dry cell (torch)
    b. Paraffin, candles, biogas, or other
    c. Electricity, solar, or gas

11. Does your household own any irons (charcoal or electric)?
    No Yes

12. How many mosquito nets does your household own?
    a. None
    b. One
    c. Two or more

13. How many towels does your household own?
    a. None
    b. One

c. Two or more

14. How many frying pans does your household own?
    a. None
    b. One
    c. Two or more

E. Agronomic Practices Questions

15. Which type of seed have you recycled (i.e. replanted seed from harvested crop)?
    a. Not recycled seed
    b. Hybrid maize seed
    c. OPV maize seed
    d. Other

16. How much of your land do you apply fertilizer?
    b. No fertilizer
    c. A small portion of the land
    d. Half of the land
    e. Most of the land

F. Product Characteristics Questions

17. What have been the results of using certified seed on your farm output?
    a. Harvested more than when not using certified seed
    b. Harvested the same volumes as when not using certified seed
    c. Harvested less than when not using certified seed

Appendix B - Additional Information

List of Agro-dealer Retail Stores

CUSTOMER DATA BASE

SL	BUSINESS NAME	TOWN	REGION
1	NGOSE STORES	CHUMUINI-TAVETA	COAST
2	IBSA AGROCHEMIST	GARSEN	COAST
3	MARERENI PHARMACY	GONGONI	COAST
4	MKULIMA AGROVET	GONGONI	COAST
5	TIMES AND SEASONS	KIBWEZI	COAST
6	MWENZENU AGROVET	KWALE	COAST
7	KAMCO AGROVET CHEMICALS	LOITOKITOK	COAST
8	KIMANA FARMERS	LOITOKITOK	COAST
9	KILIMO BIASHARA AGRODEALERS	LOITOKTOK	COAST
10	LOITOKTOK FARMERS CENTRE	LOITOKTOK	COAST
11	ROMBO FARM SUPPLIES	LOITOKTOK-ROMBO	COAST
12	KASARANI HARD WARE AND AGROVET	MAKINDU	COAST
13	NEW COAST FARMERS	MALINDI	COAST
14	RAFIKI CHEMIST	MALINDI	COAST
15	BADAR PHARMACY	MOMBASA	COAST
16	BIG A AGROVET	MOMBASA	COAST
17	COAST FARMCARE AGROVET	MOMBASA	COAST
18	FARM + PLUS	MOMBASA	COAST
19	LAMU FASHIONS	MOMBASA	COAST
20	MAKUPA CHEMIST	MOMBASA	COAST
21	MOMBASA AGRICULTURAL CENTRE	MOMBASA	COAST
22	PALMLAND PHARMACEUTICALS LTD	MOMBASA	COAST
23	SHIFA CHEM LTD	MOMBASA	COAST
24	MPEKETONI AGROVET	MPEKETONI	COAST
25	JIKAWA AGROVET	TAVETA	COAST
26	MPAKANI PHARMACY LTD	TAVETA	COAST
27	MWANZILISHI AGROVET	TAVETA	COAST
28	VOI SISAL ESTATE	VOI	COAST
29	LOMASTAR AGROVET $ SUPPLIES	WUNDANYI	COAST
30	MILIMANI HARDWARE	WUNDANYI	COAST
31	PAMTECH GENERAL MERCHANTS	WUNDANYI	COAST
32	PETMA AGROVET	AHERO	SOUTH RIFT
33	ISENYA FARMERS OUTPUT STORE	BOMET	SOUTH RIFT
34	KFA - BOMET	BOMET	SOUTH RIFT
35	SOT PHARMACY	BOMET	SOUTH RIFT
36	K.F.A BOMET	BOMET	SOUTH RIFT
37	RADS AGROVET	BONDO	SOUTH RIFT
38	ANIMAL DRAFT AND FAM-HOMABAY	HOMABAY	SOUTH RIFT
39	ANIMAL DRAFT AND FAM-OYUGIS	HOMABAY	SOUTH RIFT
40	AWENDO FARMERS STORES	HOMABAY	SOUTH RIFT
41	CARE- HOMABAY	HOMABAY	SOUTH RIFT
42	KERICHO MASHAMBANI	KERICHO	SOUTH RIFT
43	KERICHO WHOLESALERS	KERICHO	SOUTH RIFT
44	KIPSIGIS FARMERS STORES	KERICHO	SOUTH RIFT
45	PAKSONS ENTERPRISES	KERICHO	SOUTH RIFT
46	WALDAI AGROVET	KERICHO	SOUTH RIFT
47	BEST NINE SUPERMKT	KISII	SOUTH RIFT

48	ENOCHEM AGROVET	KISII	SOUTH RIFT
49	JOSEMO DISTRIBUTORS (K) LTD	KISII	SOUTH RIFT
50	KFA OYUGIS	KISII	SOUTH RIFT
51	R.M GUDKA	KISII	SOUTH RIFT
52	K.F.A MOLO	MOLO	SOUTH RIFT
53	FARMERS WORLD LTD	NAKURU	SOUTH RIFT
54	H.DOWNING LTD	NAKURU	SOUTH RIFT
55	MAMBO WHOLESALERS	NAKURU	SOUTH RIFT
56	NJORO CANNING FACTORY (K) LTD	NAKURU	SOUTH RIFT
57	PLAINSVIEW FARMERS LTD	NAKURU	SOUTH RIFT
58	SHAMBA INPUTS SUPPLIES	NAKURU	SOUTH RIFT
59	UKWALA SUPERMARKET (NAKURU) LTD	NAKURU	SOUTH RIFT
60	ENANYORAI AGROVET	NAROK	SOUTH RIFT
61	ESERIAN AGRICHEM	NAROK	SOUTH RIFT
62	EWASO NYIRO AGROVET	NAROK	SOUTH RIFT
63	MAU CHEMIST	NAROK	SOUTH RIFT
64	AMASAI ENTERPRISES	NAROK	SOUTH RIFT
65	DAO-GUCHA	OGEMBO	SOUTH RIFT
66	FARM ANIMAL FEEDS	OGEMBO	SOUTH RIFT
67	ELMART AGROVET	OYUGIS	SOUTH RIFT
68	SCODP KORINDA	SEGA	SOUTH RIFT
69	SIAYA FARMERS CENTRE	SIAYA	SOUTH RIFT
70	KFA - SOTIK	SOTIK	SOUTH RIFT
71	ROSELYNE GENERAL DISRIBUTORS	SOTIK	SOUTH RIFT
72	SALGAA AGENCIES	SOTIK	SOUTH RIFT
73	SOTIK AGROVET	SOTIK	SOUTH RIFT
74	KOIYET FARMERS STORES	SOTIK	SOUTH RIFT
75	OMBEKA ENTERPRISES	SUNA MIGORI	SOUTH RIFT
76	BUNGOMA CHEMIST	BUNGOMA	WESTERN
77	HENROSE FARMERS CENTRE	BUNGOMA	WESTERN
78	K.F.A BUNGOMA	BUNGOMA	WESTERN
79	KFA BUNGOMA	BUNGOMA	WESTERN
80	KHETIA DRAPERS LTD	BUNGOMA	WESTERN
81	NEW ADATIA WHOLESALERS	BUNGOMA	WESTERN
82	NOMADIC VET`SERVICES	BUNGOMA	WESTERN
83	RONAK PHARMACY	BUNGOMA	WESTERN
84	UPENDO AGROVET	BUNGOMA	WESTERN
85	WAKULIMA AGROVET	BUNGOMA	WESTERN
86	BUSIA AGROVET	BUSIA	WESTERN
87	GIATHI FARM INPUTS	BUSIA	WESTERN
88	KEPHIS BUSIA	BUSIA	WESTERN
89	ELGON FARMERS STORES	CHWELE	WESTERN
90	JAWA AGROVET	DAGO	WESTERN
91	KAIMOSI FTC	KAIMOSI	WESTERN
92	BAYA FARMCARE	KAKAMEGA	WESTERN
93	ELIZKIM AGROVET SUPPLIES	KAKAMEGA	WESTERN
94	HILL PHARMACY LTD	KAKAMEGA	WESTERN
95	K.F.A KAKAMEGA	KAKAMEGA	WESTERN
96	KFA KAKAMEGA	KAKAMEGA	WESTERN
97	KHAYEGA SEEDS	KAKAMEGA	WESTERN

98	PRIYANKA ENTERPRISES	KAKAMEGA	WESTERN
99	FARMERS DELIGHT	KEUMBU	WESTERN
100	KIMILILI FARMERS CENTRE	KIMILILI	WESTERN
101	AGMARK	KISUMU	WESTERN
102	DOMINION GROUP OF COMPANY	KISUMU	WESTERN
103	FARMCHOICE INPUTS &TECH	KISUMU	WESTERN
104	FOAMAT SUPERMARKET	KISUMU	WESTERN
105	HEALTHCARE FARM AND VET	KISUMU	WESTERN
106	K.F.A KISUMU	KISUMU	WESTERN
107	MAGOS FARM ENTERPRISES	KISUMU	WESTERN
108	MWANGA AGROVET	KISUMU	WESTERN
109	OSHWAL H/WARE & GENERAL TRADERS	KISUMU	WESTERN
110	PARRACK WHOLESALERS	KISUMU	WESTERN
111	WEVA SUPPLIES	KISUMU	WESTERN
112	ELMART AGROVET LUANDA	LUANDA	WESTERN
113	FARMERS PRIDE AGROVET	LUANDA	WESTERN
114	JUMBO AGROVET	LUANDA	WESTERN
115	LUANDA AGROVET	LUANDA	WESTERN
116	KEPHIS MALABA	MALABA	WESTERN
117	NAWAL AGROVET	MUMIAS	WESTERN
118	NEEMA FARMERS SHOP	NYAKOE	WESTERN
119	NYARKADERA ENTERPRISES	SARE AWENDO	WESTERN
120	MARTIN SHILABULA SHOP	SHINYALU	WESTERN
121	FARMERS STORE SONDU	SONDU	WESTERN
122	HODARI AGROVET	UGUNJA	WESTERN
123	K.F.A WEBUYE	WEBUYE	WESTERN
124	MWANZO STORES	WEBUYE	WESTERN
125	PRAMUKH PHARMACY	WEBUYE	WESTERN

Mobile Data Collection Tool

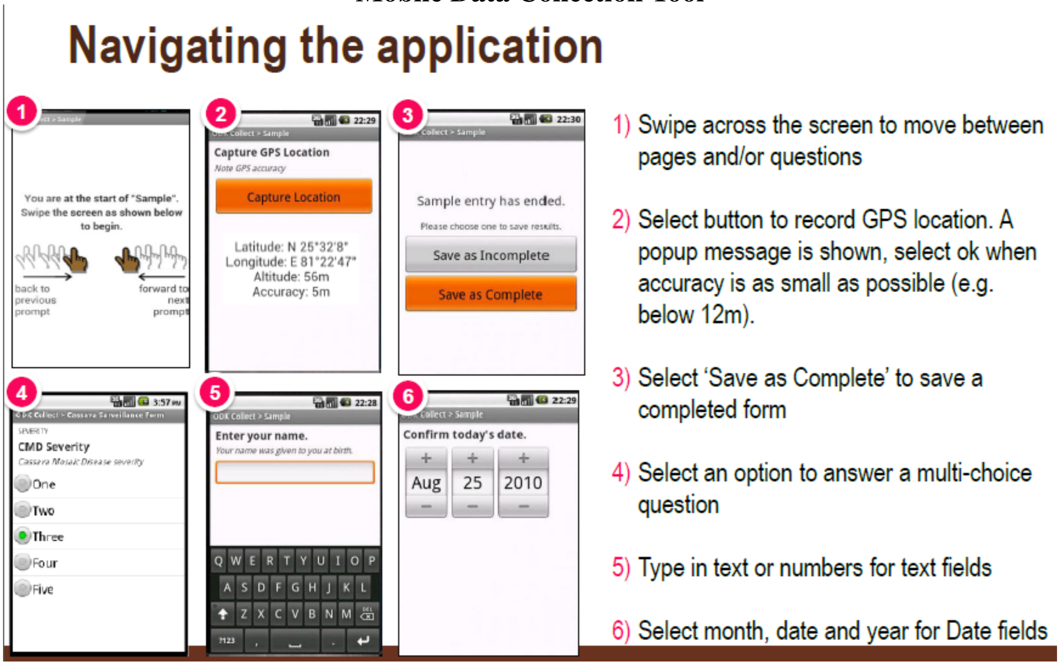


FIGURE 1: Screenshots showing how the mobile data collection tool works

Annex - Additional Questions

Progress out of Poverty Index™ for Kenya

Entity	Name	ID	Date (DD/MM/YY)	
Member:	____________	____________	Joined:	____________
Field agent:	____________	____________	Today:	____________
Service point:	____________	____________	Household size:	____________

Indicator	Value	Points	Score
1. How many members does the household have?	A. Nine or more	0	
	B. Seven or eight	5	
	C. Six	8	
	D. Five	12	
	E. Four	18	
	F. Three	22	
	G. One or two	32	
2. What is the highest school grade that the female head/spouse has completed?	A. None, or pre-school	0	
	B. Primary standards 1 to 6	1	
	C. Primary standard 7	2	
	D. Primary standard 8, or secondary	6	
	E. No female head/spouse	6	
	F. Secondary form 4 or higher	11	
3. What kind of business (type of industry) is the main occupation of the male head/spouse connected with?	A. Does not work	0	
	B. No male head/spouse	3	
	C. Agriculture, hunting, forestry, fishing, mining, or quarrying	7	
	D. Any other	9	
4. How many habitable rooms does this household occupy in its main dwelling (do not count bathrooms, toilets, storerooms, or garage)?	A. One	0	
	B. Two	2	
	C. Three	5	
	D. Four or more	8	
5. The floor of the main dwelling is predominantly made of what	A. Wood, earth, or other	0	
	B. Cement, or tiles	3	
6. What is the main source of lighting fuel for the household?	A. Collected firewood, purchased firewood, grass, or dry cell (torch)	0	
	B. Paraffin, candles, biogas, or other	6	
	C. Electricity, solar, or gas	12	
7. Does your household own any irons (charcoal or electric)?	A. No	0	
	B. Yes	4	
8. How many mosquito nets does your household own?	A. None	0	
	B. One	2	
	C. Two or more	4	
9. How many towels does your household own?	A. None	0	
	B. One	6	
	C. Two or more	10	
10. How many frying pans does your household own?	A. None	0	
	B. One	3	
	C. Two or more	7	

Microfinance Risk Management, L.L.C., Total score

This PPI was updated in March 2011. For up-to-date PPIs and other information on the Progress out of Poverty Index™ for Kenya and other countries go to www.progressoutofpoverty.org

Category Likelihoods according to Kenya PPI™ Score

PPI Score	National Poverty Line: Total Below the National Poverty Line	National Poverty Line: Total Above the National Poverty Line	National Food Poverty Line: Total Below the National Food Poverty Line	National Food Poverty Line: Total Above the National Food Poverty Line	150% of the National Poverty Line: Total Below the 150% of the National Poverty Line	150% of the National Poverty Line: Total Above the 150% of the National Poverty Line
0-4	95.4%	4.6%	95.4%	4.6%	100.0%	0.0%
5-9	95.0%	5.0%	72.6%	27.4%	100.0%	0.0%
10-14	85.8%	14.2%	57.1%	42.9%	96.5%	3.5%
15-19	82.5%	17.5%	47.4%	52.6%	95.7%	4.3%
20-24	77.3%	22.7%	37.8%	62.2%	93.2%	6.8%
25-29	67.9%	32.1%	32.8%	67.2%	89.1%	10.9%
30-34	63.7%	36.3%	23.5%	76.5%	83.3%	16.7%
35-39	46.4%	53.6%	12.7%	87.3%	75.7%	24.3%
40-44	36.9%	63.1%	9.9%	90.1%	64.8%	35.2%
45-49	30.0%	70.0%	4.7%	95.3%	64.3%	35.7%
50-54	17.8%	82.2%	1.9%	98.1%	49.4%	50.6%
55-59	13.9%	86.1%	0.9%	99.1%	41.8%	58.2%
60-64	6.1%	93.9%	0.5%	99.5%	32.3%	67.7%
65-69	4.6%	95.4%	0.9%	99.1%	20.4%	79.6%
70-74	3.8%	96.2%	0.2%	99.8%	11.1%	88.9%
75-79	0.0%	100.0%	0.0%	100.0%	4.1%	95.9%
80-84	0.4%	99.6%	0.4%	99.6%	6.7%	93.3%
85-89	0.0%	100.0%	0.0%	100.0%	4.1%	95.9%
90-94	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%
95-100	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%

Source: Microfinance Risk Management, L.L.C. based on the 2005/6 KNBS

This PPI was updated in March 2011. For up-to-date PPIs and other information on the Progress out of Poverty Index™ for Kenya and other countries go to www.progressoutofpoverty.org

Category Likelihoods according to Kenya PPI™ Score

PPI Score	USAID "Extreme" Poverty Line: Total Below the USAID "Extreme" Poverty Line	USAID "Extreme" Poverty Line: Total Above the USAID "Extreme" Poverty Line	\$1.25/Day/2005 PPP Poverty Line: Total Below the \$1.25/Day/2005 PPP Line	\$1.25/Day/2005 PPP Poverty Line: Total Above the \$1.25/Day/2005 PPP Line	\$2.50/Day/2005 PPP Poverty Line: Total Below the \$2.50/Day/2005 PPP Line	\$2.50/Day/2005 PPP Poverty Line: Total Above the \$2.50/Day/2005 PPP Line
0-4	91.5%	8.5%	100.0%	0.0%	100.0%	0.0%
5-9	73.9%	26.1%	97.2%	2.8%	100.0%	0.0%
10-14	57.9%	42.1%	83.7%	16.3%	99.3%	0.7%
15-19	46.9%	53.1%	87.6%	12.4%	99.1%	0.9%
20-24	46.3%	53.7%	81.1%	18.9%	99.2%	0.8%
25-29	36.5%	63.5%	70.7%	29.3%	96.2%	3.8%
30-34	27.6%	72.4%	63.1%	36.9%	95.4%	4.6%
35-39	16.8%	83.2%	48.4%	51.6%	91.0%	9.0%
40-44	15.4%	84.6%	35.1%	64.9%	82.7%	17.3%
45-49	7.4%	92.6%	25.4%	74.6%	75.5%	24.5%
50-54	2.5%	97.5%	8.7%	91.3%	61.1%	38.9%
55-59	2.3%	97.7%	7.8%	92.2%	44.0%	56.0%
60-64	0.3%	99.7%	1.0%	99.0%	29.0%	71.0%
65-69	1.2%	98.8%	1.1%	98.9%	20.0%	80.0%
70-74	0.2%	99.8%	0.2%	99.8%	9.4%	90.6%
75-79	0.0%	100.0%	0.0%	100.0%	6.0%	94.0%
80-84	0.4%	99.6%	0.4%	99.6%	2.2%	97.8%
85-89	0.0%	100.0%	0.0%	100.0%	4.1%	95.9%
90-94	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%
95-100	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%

Source: Microfinance Risk Management, L.L.C. based on the 2005/6 KNBS