HOUSEHOLD DEMAND FOR COMMON BEANS IN LILONGWE DISTRICT OF MALAWI

MSc. (AGRICULTURAL AND APPLIED ECONOMICS) THESIS

WUPE MSUKWA

LILONGWE UNIVERSITY OF AGRICULTURAL AND NATURAL SOURCES

BUNDA CAMPUS

SEPTEMBER, 2017

HOUSEHOLD DEMAND FOR COMMON BEANS IN LILONGWE DISTRICT OF MALAWI

WUPE MSUKWA

BSc. (Agricultural and Applied Economics), Malawi

A THESIS SUBMITTED TO THE FACULTY OF DEVELOPMENT STUDIES IN PARTIAL FULFILMENT OF THE REQUIRMENTS FOR AWARD OF THE DEGREE OF MASTER OF SCIENCE IN AGRICULTURAL AND APPLIED ECONOMICS

LILONGWE UNIVERSITY OF AGRICULTURE AND NATURAL RESOURCES

BUNDA CAMPUS

SEPTEMBER, 2017

CERTIFICATE OF APPROVAL

We, the undersigned, certify that this thesis is as a result of the authors' own work, and the best of our knowledge, it has never been submitted for any other academic qualification within the Lilongwe University of Agriculture and Natural Resources or elsewhere. The thesis is acceptable in form and content, and the satisfactory knowledge of the field was demonstrated by the candidate through an oral examination held on__________/______2017.

Major supervisor: Prof. J.H. Mangisoni

Signature: ____

Date: ___________/________________/___________

Supervisor: Associate Prof. L.D Mapemba

Signature: ____

Date: _____________/__________________/_________

Supervisor: Dr. T. Chilongo

Signature: ____

Date: ____________/__________________/___________

DECLARATION

I, Wupe Msukwa, declare that the work in this thesis is a result of my own effort and work and that to the best of my knowledge, the findings have not been presented in the University of Malawi or elsewhere for the award of any academic qualification. Where assistance was sought, it has been accordingly acknowledged.

Wupe Msukwa

Signature: ______________________

Date: ______________________

DEDICATION

This work goes out to my dear loving parents, Mr. and Mrs. Msukwa. Thank you for your moral, spiritual, financial and physical support. I also dedicate this work to my sister (Elluness) and cousins (Alice and Patricia). Thanks for your support.

ACKNOWLEDGEMENTS

I would like to thank the African Economic Research Consortium (AERC) – CMAAE and Bean Value Chain Research Network for the financial support provided for this study. Thanks to my supervisory team, Prof. J.H. Mangisoni, Associate Prof, L.D. Mapemba and Dr. T. Chilongo, for the valuable comments and guidance during the writing of both the thesis proposal and the final report.

Special thanks should also go to my classmates for their support during course work and research work. I thank my Mum and Dad for their unceasingly moral support. Thanks should also go to all people who helped me in one way or the other. Finally, but certainly not least I extend my gratitude to Prof. J.H. Mangisoni for the personal advice, moral support and unceasing mentorship. May the good LORD, God Almighty bless him abundantly above his expectation.

Above all, glory and honour be unto the Almighty God for His love and everlasting mercy to me.

ABSTRACT

Lack of consumer demand information can derail the development of the common bean sector in Malawi. Duly, the main objective of this study was to analyse the responsiveness of quantity demanded of different varieties of common beans to price and income changes among households of different locations in Lilongwe District of Malawi. The study used secondary data containing demand information on commonly traded varieties of common beans. In the analysis, three key innovations were made namely (1) consideration of zero consumption; (2) estimation of QuAIDS and its competitor LA/AIDS and (3) calculation of income elasticities from the estimated expenditure elasticities.

Based on performance, this study chose the QuAIDS over the LA/AIDS as the best representation of demand for common beans among households in Lilongwe District of Malawi. Households in Lilongwe average 4.15 adults. Majority consume common beans more than once/week and prefer the *Phalombe* variety. The *Khaki* and *Napilira* varieties are the most expensive in the high and medium, and low-density areas respectively. The computed demand elasticities showed that all common bean varieties are necessity goods with price-elastic demand, but not all are normal and substitutes. Furthermore, demand for common beans is more sensitive to price than income effects. The study therefore recommends that: (1) revenue maximisation from common beans traded by retailers can be maximised through reduction of prices and (2) if meaningful changes are to occur in common bean consumption, policy makers should pay more attention to price rather than income-related policy instruments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

AIDS	Almost Ideal Demand System
ASWAP	Agriculture Sector-Wide approach
CGIAR	Consultative Group on International Agricultural Research
CIAT	Centro Internacional de Agricultural Tropical
FAO	Food and Agriculture Organization
FAOSTAT	Food and Agricultural Organization Statistics
FCDS	Food Characteristic Demand System
FIML	Full Information Maximum Likelihood
FtF	Feed the Future
GoM	Government of Malawi
IAIDS	Inverse Almost Ideal Demand System
ICRISAT	International Crops Research Institute for the Semi-Arid Tropics
IICARDA	International Centre for Agricultural Research in the Dry Areas
IITA	International Institute of Tropical Agriculture
LA-AIDS	Linear Approximation of the Almost Ideal Demand System
LES	Linear Expenditure System
MDHM	Multivariate Double Hurdle Model
NARS	National Agriculture Research Systems
NASFARM	National Association of Smallholder Farmers
NNPSP	National Nutritional Policy and Strategic Plan
OLS	Ordinary Least Squares Estimator
PABRA	Pan African Bean Research Alliance
PIGLOG	Price Independent Generalized Logarithmic

QES	Quadratic Expenditure System
QuAIDS	Quadratic Almost Ideal Demand System
USAID	United States Agency for International Development

CHAPTER ONE

INTRODUCTION

1.1 Background Information

1.1.1 Global grain legume trends

Grain legumes (also called pulses) belong to the family of leguminous plants (also known as Fabaceae) and are grown worldwide primarily for their edible seeds (Centro Internacional de Agricultural Tropical (CIAT), 2013). Grown on 12-15% of the earth's arable land, the seeds of grain legumes are harvested mature and marketed dry for use as food or feed or processed into various products (Rusike *et al*., 2013). The plants have the ability to fix atmospheric nitrogen into the soil hence enriching the soil and reducing the cost of fertilizer inputs in crop farming. With this property and their versatile benefits in terms of farm incomes, food security, nutrition and health; the plants have qualified to be important food, cash, fodder and rotational crops in most parts of the world.

The worldwide yield of major grain legumes has increased modestly (<1%/year) over the past decades with an average of 0.90 tons/ha for the past decade (FAO, 2015; Akibode and Maredia, 2011). Bareja (2016) reveals that the major producers of grain legumes worldwide include Brazil (common bean), China (broad beans, horse beans and groundnuts), India (chick peas, pigeon peas), Nigeria (dry cow peas), USA (soybean), Canada (dry peas, lentils), Chile (lupins) and Ethiopia (Vetches). Worldwide consumption of grain legumes has trended upward over the recent 30 years, with an estimated average of 60 million metric tonnes per year globally and per capita consumption of 10kg/capita/year for the past decade (CIAT, 2013; Akibode and Maredia, 2011). According to CIAT (2012), the demand for grain legumes is growing more strongly than for other crop commodities in the developing world, yet local production is not able to keep up with this demand.

1.1.2 The current grain legume situation in Malawi

The Government of Malawi has committed to develop agricultural value chains that contribute to household income and nutrition while sustaining the biophysical environment. So far, grain legumes have shown the greatest potential to achieve this triple objective. Generally, grain legumes are grown because they are nutritious, have high gross margin and replenish soil fertility (Rusike *et al*., 2013). Furthermore, farmers grow legumes to spread risk for livelihood resilience. Katungi *et al.* (2009) reported that, in recent times, the production of grain legumes has increased especially with the support of stakeholders such as United States Agency for International Development (USAID), Feed the Future (FTF) Malawi, Pan African Bean Research Alliance (PABRA), National Agriculture Research Systems (NARS), Centro Internacional de Agricultural Tropical (CIAT), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), International Centre for Agricultural Research in the Dry Areas (IICARDA), International Institute of Tropical Agriculture (IITA) and Bunda college of agriculture. These stakeholders have played key roles in the breeding, production, marketing and consumption level strategies of grain legumes.

The major grain legumes grown in Malawi include common bean, groundnut, cowpea, pigeon peas and soybean. According to Rusike *et al.* (2013), the aforementioned grain legumes are very important for cash, food security, nutrition, gender, and sustainable natural resource management. Groundnut is the most commonly grown legume throughout Malawi (Katungi *et al*., 2013). Common bean ranks second to groundnut in terms of area planted and quantity produced (CIAT, 2012). Figure 1-1 shows an upward trend of quantity produced for groundnuts, dry beans (synonymous to common beans) and soybean for the past 15 years. These upwards trends are attributed to area expansion (Figure 1-2) and increase in interplay of stakeholders in the breeding, production, marketing and consumption level strategies. Figure 1-1 shows a downward trend in the production of dry cowpeas over the past 15 years. The trend in production of cowpeas directly correlates with the hectarage trend as shown in Figure 1-2.

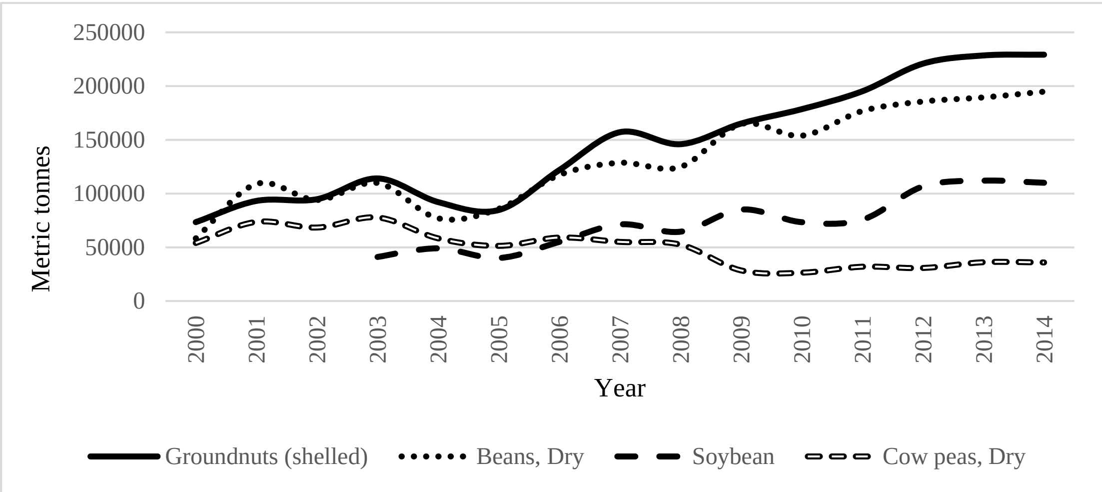

Figure 1.1 Production of major grain legumes in Malawi

Source of data: FAO (2015)

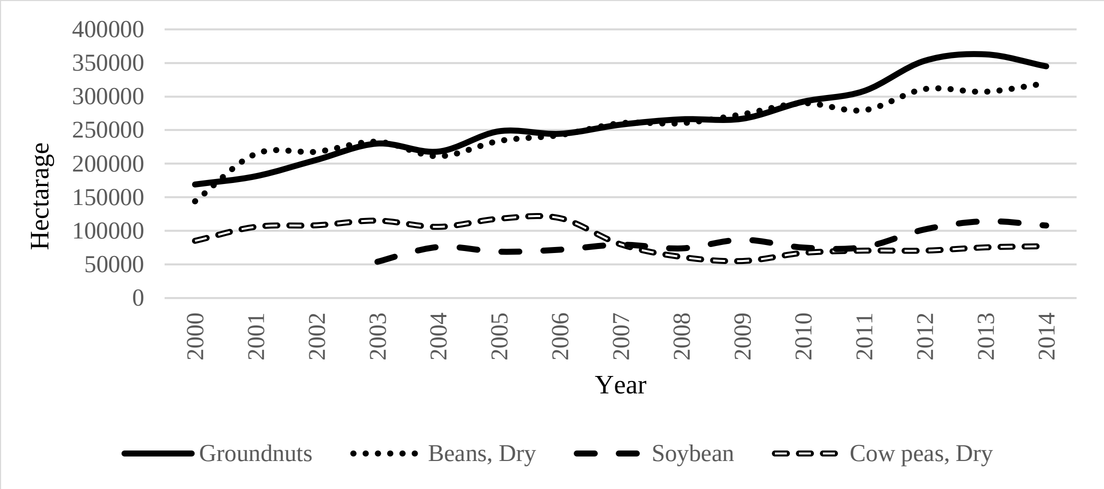

Figure 1.2 Hectarage of major grain legumes in Malawi

Source: FAO (2015)

1.1.3 Common beans in Malawi

Among the major grain legumes grown in Malawi, common bean (*Phaseolus vulgaris*) has proven to be the most commercially viable crop with greatest potential to sustainably reduce poverty and improve nutrition among farmers (CIAT, 2013; USAID, 2011). Common bean is an important food crop as it contains high amounts of proteins, vitamins, carbohydrates, and other valuable micronutrients like iron, zinc, potassium, and magnesium (Larochelle *et al*., 2015). Additionally, it matures early and offers relatively higher gross margins compared to priority crops like maize (Rusike *et al*, 2013). Common beans can also be sold at various stages of maturity like green leaves, fresh pods, and immature/dry grains (Birachi, 2012). Common bean sometimes supplements fishmeal or in livestock feed (Sichilima *et al.*, 2015). Ultimately, production of common bean has several benefits to the environment, farmers and the nation at large.

Common beans are produced mostly in higher altitude (1000-1700 meters above sea level) and high rainfall (800-1500mm/annum) areas (Muthoni *et al.*, 2008). More than 70% of common beans are produced in Dedza, Ntcheu, Lilongwe, Chitipa and Thyolo districts (Rusike *et al*., 2013). The beans are primarily produced under mixed/intercrop cropping system with crops like maize and cassava by 70% of the farmers (Birachi, 2012; CIAT, 2012). Although rare, some farmers use pure stands especially during the winter cropping season (Munthali, 2013). Smallholder farmers (mostly women) dominate production of common beans (Mwale et al., 2009).

The most popular common bean varieties marketed in Malawi include *Nanyati* (cream mottled), *Khaki/Mulanje* (cream), *Napilira/Kachiyata* (red mottled), *Phalombe/ Chimbamba/Thyolo* (red kidney), *Kaulesi* (purple/greyish), and mixed beans (various colours) (Chirwa and Phiri, 2007). Common beans are sold as dried grain to households, institutions, local/private traders, cottage industries, processors and exporters (Mwale et al., 2009; Muthoni *et al.*, 2008). The dried grain is used for direct human consumption or further processed into human food.

Although the production of common beans has trended upwards over the past years (Figure 1-1), there is still unmet market demand (GoM, 2013; Katungi *et al*., 2009). For instance, between 2009 and 2014 (when production peaked), about 177 metric tons of common beans were produced annually against a total market demand of 200 metric tons (FAO, 2015). Imports from Mozambique partly supplemented the supply shortfall. Birachi (2012) and Muthoni *et al.* (2008) expected the gap to increase given high population growth, weather turbulence, and stagnant-to-declining yields. Additionally, productivity of common beans is still low, estimated at 616kg/ha against a potential yield of 2000kg/ha (FAO, 2015). The supply shortfall of common beans reinforces the current annual per capita consumption which is estimated at 11.6kg/year, which is low compared to the African average of 17kg/year and the recommended 23.4kg/year (FAO, 2015; Larochelle *et al*., 2015; Katungi *et al*., 2009).

The current supply gap for common beans in Malawi is attributed to several biotic (e.g. pests and diseases), abiotic (e.g. low soil fertility and low rainfall) and socio-economic (e.g. poor access to improved bean varieties, lack of farm inputs, lack of demand information, inappropriate pricing strategies, poor storage facilities and poor access to markets) constraints (CIAT, 2013; Munthali, 2013; Rusike *et al.,* 2013; USAID, 2011). Of particular interest to this research is the dearth of demand information and inappropriate pricing strategies in the common bean value chain. Munthali (2013) and USAID (2011) explain that such constraints have often discouraged farmers from scaling up production of common beans. In addition, as explained by USAID (2011), they have exacerbated lack of good, established and organized market in the common bean sector. Currently most bean farmers are producing for subsistence purposes, with less than 20% producing surpluses for sale (CIAT, 2013).

1.2 Problem statement

The government of Malawi is absolutely committed in achieving food self-sufficiency at household level (GoM, 2013). In this regard, it has put in place a series of reforms that position the country for sustainable growth. There are several national leading documents[1] that contain specific measures for growth put in place by the government.

[1] See Agricultural Sector-Wide Approach (ASWAP) , National Nutritional Policy and Strategic Plan (NNPSP) and the Malawi Growth and Development Strategy

One of the key reforms put in place by the government is the development of agricultural value chains that contribute to household income and nutrition while sustaining the biophysical environment. This corresponds with the initiative by USAID to sustainably invest in high potential value chains to develop markets and improve nutritional options.

Despite the said efforts, there is slow development for some potential and promising value chains. For instance, common beans qualify as a crop that can help in achieving the triple objective of enhancing incomes, nutrition and soil fertility, yet its supply gap is very high and less than 20% of the farmers producing surpluses for sale (CIAT 2013). The government and other stakeholders in the legume industry (see CIAT 2013, GoM 2013; USAID, 2011) have officially acknowledged that the development of the bean value chain is slow citing low agricultural productivity and underdeveloped markets[2] as the main culprits. Munthali (2013) and USAID (2011) further explain that underdeveloped markets in the bean value chain are exacerbated among others by lack of information especially on consumer demand[3]. This concurs with Theil (1975) who explained that to achieve a good market system, one of the key ingredients is consumer demand information[4] because it helps in decision making of economic agents, projecting future market conditions, and bringing about needed changes.

[2] A good market system signals the relative scarcity of goods and services, guides the decision of economic agents, and ensures mobility of commodities over time and across space (Ravallion, 1986). If development is to be achieved, essentially any value chain needs a good market system

[3] This includes extent to which consumers adjust the consumption as a result of price and income changes vis-à-vis price and income elasticities of demand

[4] This is also reflected in the neo classical micro theory of consumer demand

Not many studies have been conducted on demand for common beans in Malawi. The few studies that have been conducted[5] have analysed beans at an aggregate level yet there are several varieties that differ by attributes in the markets. In addition, other studies that have disaggregated beans by varieties, have only used specific bean attributes to explain market demand, ignoring prices and income which ultimately inform the nature of demand[6]. Just like all trade, bean markets are demand driven, thus should the information gap persist, the common bean value chain will continue having underdeveloped markets and enhancing incomes, nutrition and soil fertility through beans would prove futile. In addition to widening the bean supply gap, the information gap has a potential of reinforcing the status quo where producers and sellers deal more in poorly demanded varieties hence leading to impediment of the common bean value chain. Thus, it is essential to gain insights on demand for beans and predict changes in household expenditures caused by price and income change from a proper and purposeful empirical econometric analysis.

1.3 Justification of the study

Information from this study is relevant to farmers and intermediate traders in making informed common bean production and trading decisions by revealing to them strongly demanded varieties of common beans and how to increase revenue through appropriate pricing strategies. This knowledge will help the two actors to invest and to maximize profits from common beans production/trading. Furthermore, the study will help marketing agents make informed targeting decisions, design and participate in effective

[5] Chirwa and Phiri (2007), Muthoni *et al*. (2008), Maganga et al. (2014), and Kankwamba *et al*. (2012).
[6] Whether the crop is normal or inferior, necessity or luxury, and the price elasticity of demand.

channels as well as formulate demand responsive marketing strategies. Consequently, this will reduce risks associated with various marketing functions.

Demand information from this research will also assist policy makers in devising price, trade, market development, and revenue creation policies applicable in the common bean value chain. The relative sizes of expenditure and price elasticities from this research will help the policy makers in the choice of policy instruments (i.e. whether price or cash related) relevant for improving demand and/or consumption. The demand elasticities can also be used to calculate welfare indicators such as change in consumer and producer surpluses, change in government revenue, net economic loss in consumption, net economic loss in production, and net effect of trade barriers and/or trade policies. The above indicators can assist in the formulation of appropriate macro policies.

Results from this study will also be relevant to other stakeholders interested in reducing poverty and improving nutrition of farmers through common bean's breeding, production, market and consumption level strategies. Such stakeholders include FTF Malawi, PABRA, NARS, CIAT, ICRISAT, IICARDA, IITA and Bunda College of Agriculture.

Finally, the insights on demand for common beans will form an important ingredient of some complete models such as multimarket and CGE models. These models aim at explaining production and consumption, price formation, trade flows, welfare impacts of policies, income levels, and government fiscal revenues.

1.4 Objectives

1.4.1 Main objective

The main objective of this study was to analyse the responsiveness of demand for common beans to price and income changes with a particular focus on households in Lilongwe district of Malawi. This objective was achieved through estimation of a household demand model that featured six commonly traded varieties of common bean.

1.4.2 Specific objectives

The specific objectives are:

i. To assess the sensitivity of budget shares for common beans to price and income effects.

ii. To assess the responsiveness of quantity demanded to price and income changes.

1.5 Research hypotheses

Based on objective one, this research made the following hypothesis;

i. Prices and income affect household budget share of common beans.

Based on objective 2, this research made the following hypotheses;

i. A unit percentage increase in own-price of a common bean variety decreases the quantity demanded by a less than unit percentage.

ii. A unit percentage increase in price of a common bean variety increases the quantity demanded of other common bean varieties.

iii. A unit percentage increase in income of the household increases the quantity demanded of a common bean variety by less than a unit percentage.

1.6 Research question

This research was guided by the following question:

> To which extent do households adjust their consumption of common beans as a result of price and income changes?

1.7 Thesis Organization

Chapter one has been discussed. Chapter two defines the key terms used in demand analysis, approaches used, estimation issues and their associated remedies, and presents some of the empirical studies on demand for beans across the world. The chapter further defines the research gap and gives a panorama view of how the thesis has addressed the gap. Chapter three describes the survey and data sources, presents the conceptual and theoretical framework. Thereafter the chapter presents the empirical model and discusses various econometric issues that have been addressed and the econometric tests that have been performed. Chapter four and chapter five present and discuss the results of descriptive and statistical analysis, respectively, while chapter six concludes with summary, conclusions, and policy implications.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

This chapter defines the key aspects of the thesis and presents related literature across the world. The first section defines the key terms used in demand analysis and the commodity analysed. The subsequent section presents the common approaches used in demand analysis, their estimation issues and known remedies. Thereafter, the next section discusses the related empirical studies that have been done across the world. Lastly, the chapter presents the research gap based on the literature reviewed and how this study intended to bridge the gap.

2.2 Definition of key terms

2.2.1 Consumer

A consumer is a purchaser of good/service in retail (Wetzstein, 2004). Consumers can be classified into industrial consumers (purchase for industrial use) and final (purchase for final consumption) (Moscati, 2007). A consumer can refer to an individual or group of individuals analysed as a unit i.e. household. This study will focus on the final consumer and most specifically households in Lilongwe of Malawi.

2.2.2 Quantity demanded

Demand is one of the most important decision making variables in an economy and conceptually it implies consumer readiness to satisfy desire by paying for goods and services (Wetzstein, 2004). Demand for a particular commodity is often gauged by the quantity demanded and has unique determinants, which are often summed in a demand profile plotted as slopes on a graph of quantity versus own-price. On such a plot, the vertical axis is the price while the horizontal axis is the quantity demanded. The neoclassical micro theory of consumer demand assumes the demand profile slopes downwards from left to right as price decreases. The miniature details of a demand profile are discussed in the subsequent sections.

2.2.3 Common bean (Phaseolus vulgaris)

Common bean is a herbaceous annual leguminous plant that is grown in most parts of the world. The crop originated from Mexico in the years 5500-7000BC and has more than 30 varieties across the world (Muthoni et al., 2008). With trifoliate large leaves, common bean is largely self-pollinated, has non-endospermic seeds and the varieties vary greatly in size and colour (Birachi et al., 2012; Katungi et al., 2009). The crop shows variation in growth habits from determinate, bush to indeterminate and extreme climbing types. The bushy bean is the most predominant type grown in Africa (Katungi et al., 2009). Common bean is largely produced in Brazil and Mexico, with Africa ranking second among the continents (CIAT, 2013). Major African producers include Kenya, Burundi, Congo, Uganda, and Rwanda (Rusike et al., 2013). This study will focus on six commonly traded varieties in Malawi namely *Kabalabala*, *Nanyati, Napilira, Khaki*, *Phalombe*, and mixed beans.

2.2.4 Elasticities

Many economic insights about nature of demand for a commodity often translate into the effect of income and prices on the quantity demanded. The response of quantity demanded from changes in prices and income can be postulated by various methods of which the famous ones are graphical and mathematical methods. Analysts examine how consumers respond to changes in prices and income by looking at derivatives or elasticities of the demand function. According to Nicholson and Snyder (2008), the use of derivatives has one major disadvantage for empirical work; derivatives are affected by unit of measurement. Hence most analysts often use elasticities which are metric independent to examine the nature of demand for a commodity.

The elasticities of interest in demand analysis often include own-price, cross price and income elasticities. Own-price demand elasticity refers to the percentage change in quantity demanded of a commodity due to a 1% change in its own price. Cross-price elasticity refers to the percentage change in quantity demanded of a commodity due to a 1% change in the price of another related commodity. The income elasticity refers to the percentage change in quantity demanded of a commodity due to a 1% change in the consumer income. More on demand elasticities including the mathematical derivations have been explained in chapter 3.

2.2.5 Forms of demand

There are different forms of demand with respect to demand elasticities. For instance, with respect to price elasticities we have elastic demand, unitary demand, and inelastic demand. A commodity is said to have inelastic demand if the own-price elasticity is less than one in absolute terms. If the own-price elasticity is greater than one in absolute terms, then the commodity has elastic demand. A commodity with unitary demand has own-price elasticity that is equal to one in absolute terms. Perfect elastic demand exists when the own-price elasticity of demand is equal to infinity and perfect inelastic demand exists when the own-price elasticity of demand is equal to zero. The cross-price elasticities show the relationship of a commodity with other commodities. Two normal goods are complements if the cross-price elasticity is negative and substitutes if the cross-price elasticity is positive.

2.2.6 Types of goods

In the realm of demand analysis, various types of a good exists depending on elasticities. A normal good is a good with positive income elasticity while an inferior good is a good with negative income elasticity (Binger and Hoffman., 1998). A Giffen good is a good with a positive own-price elasticity (Nicholson and Snyder, 2008). Necessity goods are goods with positive and less than one income elasticity, and luxury goods are goods with positive and greater than one income elasticity (Varian, 1987).

2.3 Demand analysis

2.3.1 Common approaches to demand analysis

Two approaches can be used to estimate the parameters of demand equations. One consists of specifying estimable single equation demand function without using economic theory as a guideline. An example of this is demand estimation from time series data (Sadoulet and Janvry, 1995). This approach is arbitrary, the functional forms assume constant elasticities, and the estimated parameters in general do not satisfy the requirements of demand theory particularly budget constraints (Taniguchi and Chern, 2000). The second approach involves using the theory of demand as a guideline for the choice of functional form and variables to be included in the model (Sadoulet and Janvry, 1995). The two most popular demand models in this category include the Linear Expenditure System (LES) and the Almost Ideal Demand System (AIDS).

Demand models are classified based on specification, estimation procedure, and number of equations (Taniguchi and Chern, 2000). The models can be linear or non-linear functions depending on fit and adequacy of data. Additionally, they may be modelled using direct or indirect utility concept. Based on the source of data, they can also be modelled using cross-sectional data, time series data, and pooled or panel data. Furthermore, demand models may be single equation or a simultaneous equation system models or complete demand system models. Complete demand systems comprise additive models (including LES and addilog demand models), *a priori* models (include constant elasticity model and quadratic expenditure systems), and non-additive models (including translog systems and AIDS) (Liu, 2006; Leserer, 2010; Taniguchi and Chern, 2000).

Much empirical work on demand analysis has focused on estimation of complete demand systems more specifically non-additive models. The non-additive complete demand systems are consistent with demand theory, are flexible, allow estimation of cross price elasticities with a group of substitutes or complements, do not assume any type of additivity, and allow for positive or negative, increasing or decreasing income elasticities (Liu, 2006; Sadoulet and Janvry, 1995; Sivaramane, 2012. One of the famous and commonly estimated non-additive complete demand system is the AIDS model. The AIDS model is a consistent and more flexible demand system that has the ability to impose and test the properties of consumer demand, exactly satisfies the axioms of choice, and exactly aggregates over consumers without invoking parallel linear Engel curves (Deaton and Muellbauer 1980). Due to estimation issues over the past years, different extensions of the AIDS model have emerged. The commonly applied extensions of the AIDS model in empirical analysis include the Linear

Approximation of the AIDS (LA/AIDS), the Quadratic Almost Ideal System (QuAIDS), and the Inverse of the AIDS model (IAIDS).

2.3.2 Common estimation issues in demand analysis

A number of demand estimation issues have been reported in empirical work. Examples include missing prices, presence of zero consumption, endogeneity in prices and expenditure, and measurement errors in expenditures. The subsequent sections discuss these issues.

2.3.2.1 Presence of zero consumption

Use of household data for demand analysis is characterized by censored response problem in which households report zero consumption of one or more of the items analyzed. The main reasons for this outcome include misreporting or mis-measurement and infrequency of purchase due to short survey period, consumer preferences, availability, inability of consumers to purchase the good at the current prices and income levels (Alviola, 2010; Yen and Lin, 2004). The censored response problem makes the standard procedure of estimating a demand system inefficient and biased (Tobin 1958). To remedy this problem, a number of parametric, semi-parametric and nonparametric approaches have emerged over the past decades. The most notable approach of dealing with the censored response problem is the parametric two-step estimation of demand models. Good examples of this approach include Tobin (1958), Amemiya (1974) and Heckman (1979) estimators. The Heckman two-step approach is the most widely applied approach in empirical demand analysis (Kenkel and Signorino, 2012). Two variants of the Heckman two-step approach namely Hein and Wessels

(1990) and Shonkwiler and Yen (1999) estimators have been widely employed in most empirical demand analysis.

2.3.2.2 Missing prices

The problem of missing prices is a common phenomenon in empirical demand analysis. In the cases of reported prices, missing prices can result from infrequency of purchase and misreporting or mis-measurement. In the case of derived prices, missing prices often result from zero consumption of commodities. Households with zero consumption provide no information on quantities consumed and/or prices. Missing prices pose a serious estimation issue as they may affect convergence of demand models especially in the case of large presence of zero consumptions. In order to remedy this problem, most empirical work on demand analysis (e.g. Lazaro, 2014; Maganga, 2014; Leserer, 2010) replaced missing prices with the mean price according to location.

2.3.2.3 Endogenous prices and expenditures

Expenditure endogeneity

The neoclassical micro theory of demand assumes expenditure is exogenous to income, but in empirical specification, in order for the budget shares to sum to one, income is computed as the sum of total expenditures across commodities analysed. As such, total expenditure may be determined jointly with the expenditure shares of the individual commodities being analysed, making it endogenous in the expenditure share equations (Sola, 2013). In addition, expenditure endogeneity may exist when the household expenditure allocation process is affected by unobserved factors not included among the explanatory variables in the demand equation hence bundled in the error term.

Disregarding expenditure endogeneity and treating expenditure as exogenous in demand models often results in inconsistent and biased estimates (Agostini, 2014).

The common practice for dealing with expenditure endogeneity depends mainly on the type of data used. In cross sectional data, the instrumental variable estimation procedure is the mostly applied method in mitigating expenditure endogeneity. The common instrument used in most empirical work is household income (see Sola, 2012; Bopape, 2006; Kebede, 2003). In panel data linear transformations of the original model, such as through fixed effects and first differencing are used to remove the unobserved heterogeneity component of the error term (Bopape, 2006). Other methods that have been used to mitigate expenditure endogeneity especially error correlations across equations include estimating the system of equations using the seemingly unrelated regression framework (SUR) (see Taylor, 2014). The SUR uses the correlations in the errors of other equations to improve the parameter estimates (Taylor, 2014)

Price endogeneity

The prices of commodities analyzed may correlate with the error terms in the system of equations. This arises from unobserved heterogeneity driving changes in the prices (Huber, 2012). This gives rise to endogenous prices within the system of equations. Prices may also be endogenous in demand models because consumers sometimes change their demand in relation to the price. Akbulutgiller (2008) explains that price endogeneity is a serious problem when price is determined by the interplay of supply and demand, and if the supply rather than the prices is assumed to be predetermined, the case with IAIDS estimation. The endogeneity in prices is resolved in two ways: by instrument variable estimation method and by explicit specification of price equations which reflect strategic firm behavior and supply cost (Huber, 2012: Akbulutgiller, 2008). The instrumental variable estimation is the major way of mitigating price endogeneity. Instruments are variables that have no direct association with outcome of interest but highly correlated with the regressor of interest (Greene, 2003). Two sets of instruments have often been applied to mitigate price endogeneity. The first set includes demand and supply shifters within city/region and second set includes one-period lagged endogenous prices (for time series data) or price of neighboring cities (see Akbulutgiller, 2008).

2.3.2.4 Measurement errors

Respondent and interviewer errors form a key characteristic of most survey data. Lewbel (1996) recognized the problem of measurement error in the literature of Engel curves estimation. Household expenditures are often contaminated with measurement errors and according to Hikaru and Kozumi (2001), the observed mean household expenditure often over-estimates the mean of the true total expenditure. The problem of measurement errors poses significant econometric and economic implications. When variables of interest are contaminated with measurement errors, one of the common methods applied is the use of instrument variables (See Agostini, 2014; Bopape, 2006). Thus, the instrument variables estimation not only helps solve endogeneity problems but also resolve measurement errors that may exist in the data.

2.4 Empirical studies on demand for beans

Consumers are the most important players in any agricultural value chain (Kinsey, 2001). Essentially, flow of information about consumer demand for a given commodity to the rest of actors drives the entire value chain. Consumer demand information allows for informed decision making amongst actors in the entire chain. Therefore, an understanding of consumer demand provides one of the greatest approaches towards improvement of value chains in an economy. It is for this background that literature on consumer demand for different commodities traded within economies is rapidly growing. The subsequent paragraphs discuss some of the reviewed literature on demand for legumes and common beans.

Musyoka et al. (2007) studied food consumption patterns in Kenya using the Quadratic Almost Ideal Demand System (QuAIDS) through the Shonkwiler and Yen (1999) approach. The Shonkwiler and Yen (1999) approach was employed to remedy the censored response problem arising from the zero consumption of some food commodities during the survey period. The uncompensated own-price elasticities indicated legumes (including beans) are non-Giffen goods and have elastic and inelastic demand for rural and urban Kenya, respectively. The compensated own-price elasticities indicated that legumes are non-Giffen goods and have inelastic demand regardless of location. The cross-price elasticities provided a mix of both complementary and substitution relationships between legumes and other household food groups. The income elasticities indicated that legumes are normal and necessity goods for the rural areas and are normal and luxury goods for the urban areas. Musyoka et al. (2007) also noted that household size, age and gender of household head affect demand for legumes.

Maganga et al. (2014) also estimated a QuAIDS model for food in rural Malawi using the Shonkwiler and Yen (1999) approach. Their results showed that legumes (including beans) have an expenditure share of 6% in household food expenditure and they exhibit the properties of a normal, non-Giffen, and necessity good. The results further revealed that legumes have inelastic demand and that they are a substitute for meat. Leserer (2010) estimated a LA/AIDS for food in both rural and urban areas of Indonesia. His results showed that legumes are a non-Giffen good and have inelastic demand for both urban and rural areas of Indonesia. The relationship between legumes and other food groups was both complementary and competitive. The results from the expenditure elasticities indicated that legumes belong to the group of normal and necessities irrespective of the income group, survey periods, and survey areas.

Kumar et al. (2011) estimated a food demand system for India using two alternative models namely multistage QuAIDS and a Food Characteristic Demand System (FCDS). The results revealed that budget shares of pulses (including beans) were independent of income groups and were almost similar to budgets shares of other protein sources such as meat, fish and eggs. Both models revealed that pulses are a normal, non-Giffen and necessity good and have inelastic demand.

Akerele et al. (2013) examined household food demand in Nigeria using a Multivariate Double-Hurdle model (MDHM). The objective was to address a censored response problem. Considering four food groups, the own price elasticities of beans indicated that beans are a non-Giffen good and have inelastic demand regardless of income group. The cross-price elasticities showed that beans have complementary relationship with maize for all income groups, and a mix of substitution and complementary relationships with tubers and snacks in the three income groups considered. The income elasticities revealed that beans are a normal and necessity good. The results further indicated that the factors that determine demand for beans across different income quintiles include age of household head, location factors, and ratio of children to adults in the household.

Through a QuAIDS model, Gonzalez and Wieck (2014) found that beans are a non-Giffen good and have inelastic demand in Mexico. Their expenditure elasticities indicated that beans are normal and necessity goods. The aforementioned results concur with Kankwamba et al. (2012) who estimated a LA/AIDS for legumes in Malawi through the Heien and Wessels (1990) approach to address the censored response problem. However, for Temitrope and Haruna (2013), their results from the compensated and uncompensated own-price elasticities estimated by the Almost Ideal Demand System (AIDS) indicated that beans have both unitary and inelastic demand. In addition, the results revealed that age, wealth status, occupation and gender affect demand for beans.

Mishili et al. (2009) studied consumer preferences as drivers of common bean trade in Tanzania from a marketing perspective. Using hedonic pricing which is based on a Lancaster model, the results indicated that consumer's market demand for beans is largely influenced by grain size, grain colour, grain damage, and whether or not the variety is natural. Besides the aforementioned attributes, other studies (Mangisoni and Bokosi, 2004; Mkanda 2007; Chirwa and Phiri 2007; Katungi et al. 2009) have recognised the significant influence of cooking time, taste/flavour, familiarity, availability, damage level, and flatulence on market demand for common beans in different parts of the world.

2.5 Summary

Beans fall in the category of normal and non-Giffen good, and can have inelastic, elastic or unitary demand. In addition, they can feature as a necessity or luxury good. The factors that affect the demand for beans across the world include consumer characteristics (such as age, education, gender, occupation, income, race, household size and location) and bean attributes (such as cooking time, grain size, grain colour, grain damage, flavour, taste, familiarity, and flatulence). The gap that has been identified from the reviewed literature is information about the nature of demand for different varieties of common beans. For Malawi's case, very few studies have focused on the demand side of the crop and examples include that of Chirwa and Phiri (2007), Muthoni *et al*. (2008), Maganga et al. (2014), and Kankwamba *et al*. (2012). Among the aforementioned, no study has assessed the responsiveness of demand for different varieties of common bean to changes in price and income. Studies by Chirwa and Phiri (2007) and Muthoni *et al*. (2008) focused on measuring specific bean attributes (such as cooking time, taste, colour, size, damage level, and flavour) to explain the market demand of common beans. Whilst these attributes are important in determining demand, they are not sufficient in revealing the true nature of demand since by consumer demand theory, demand is mainly a function of utility which ultimately depends on the income of individuals and the prices they face. Studies by Maganga *et al*. (2014) and Kankwamba *et al*. (2012) estimated censored food and legume demand systems in Malawi respectively. While the two studies featured beans in the estimation; the analysis did not disaggregate beans by varieties. There are several common bean varieties and they vary greatly in terms of grain size, taste and colour hence it is imperative for any study analysing the demand for beans to disaggregate beans by varieties and invest in the extent to which the quantity demanded fluctuates given changes in prices and income of consumers. Henceforth, it is against this background that this study assessed the nature of demand for six commonly traded varieties of common beans in Lilongwe district of Malawi. The next chapter expands the methodology used.

CHAPTER THREE

THEORETICAL AND EMPIRICAL APPROACHES

3.1 Introduction

This chapter outlines the methodology that was used to achieve the objectives and to test the hypotheses of the study. It gives a description of the data used and the analytical methods employed. The first section presents the conceptual framework that provided a roadmap for this study. The second section presents the theoretical framework, which provided the building blocks of the analytical framework. Thereafter, the analytical framework is presented to explain methods employed, variables used and a *priori* expectations. The last section describes the data that was used in the study, location and sampling.

3.2 Conceptual framework

The neoclassical micro theory of consumer demand explains demand as a process that is based on utility maximization framework (Moscati, 2007; Theil, 1975). Empirically, utility is not only a function of physical consumption of commodities, but is also an interactive process of commodity attributes, consumer characteristics, and the general environment which may include socio-cultural, biological, geographical and other relevant factors (Walker and Alkiva, 2002; Parraga, 1990; Lancaster, 1966). Thus, demand studies are supposed to take into account all the utility forming variables. However, due to data difficulties researchers working on demand narrow down their focus to few quantifiable options.

In consonance with the random utility theory, this study assumed the decision to purchase a common bean variety is affected by consumer and commodity characteristics. The important characteristics of a common bean variety include price, colour, cooking time, grain size, and taste (Muthoni et al., 2008; Chirwa and Phiri, 2007; Mkanda et al., 2007; Mangisoni and Bokosi, 2004). The important consumer characteristics include age, gender, marital status, household income, household size, and education (Temitope and Haruna, 2013; Katungi et al, 2009) Musyoka et al., 2007). Since many economic insights stem from the recognition that utility ultimately depends on the income of individuals and on the prices they face; this study focused on prices and income in explaining the nature of demand for six commonly traded common bean varieties in Lilongwe district.

Following the Heckman two-step analysis, the demand for a common bean variety was viewed as a two-step decision were in the first step the consumer decides whether to buy and in the second step allocates a budget if he/she decides to buy. The second stage was the main focus of this study. This step is largely affected by prices and income and reveals the true nature of demand for a particular variety. Figure 3.1 illustrates the conceptual framework for this study, which shows the two-step demand process for common beans as understood by the author through literature review.

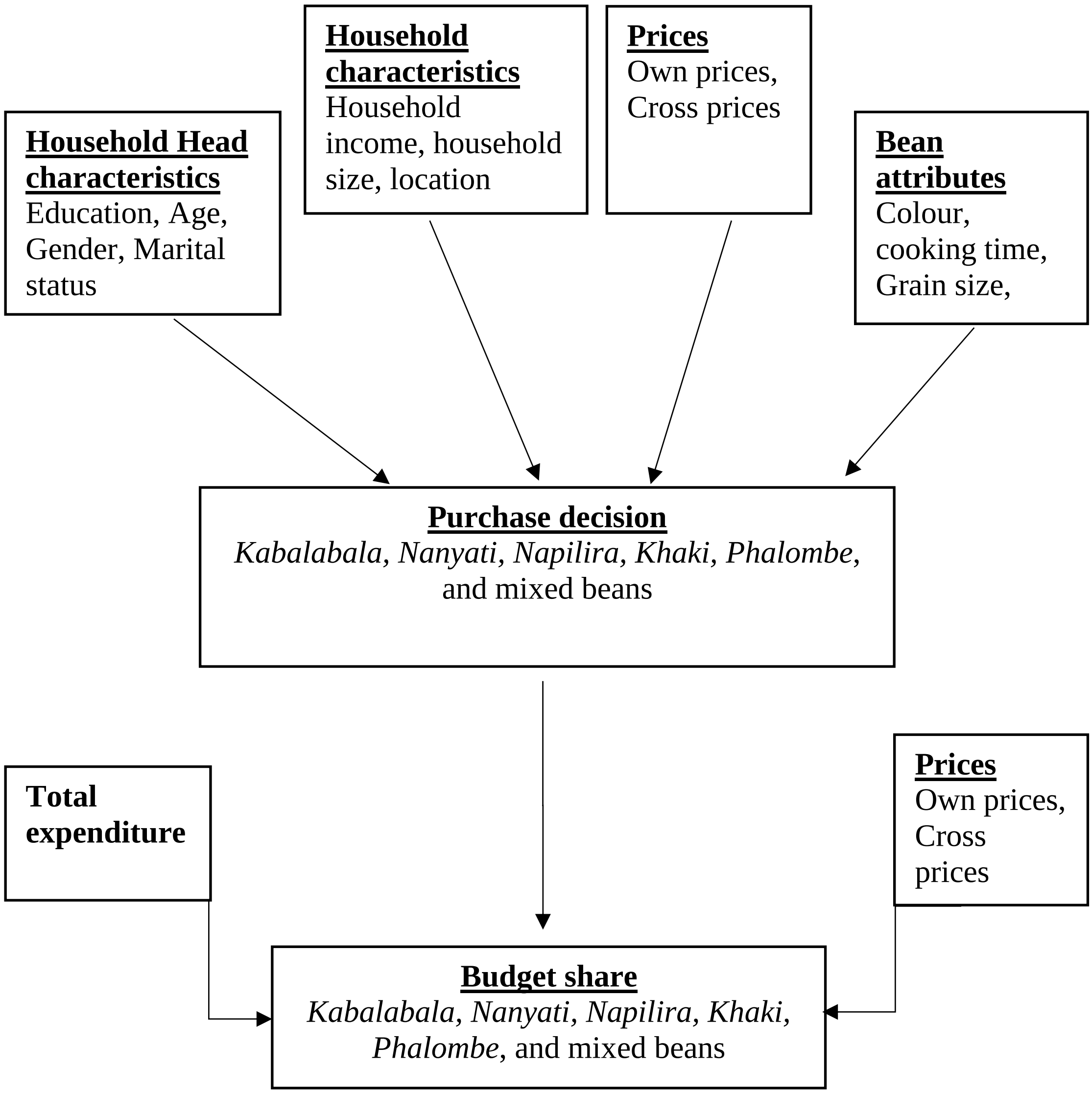

Figure 3.1 Two-step demand process for common beans

3.3 Theoretical framework

3.3.1 The neoclassical micro theory of consumer demand

The central concern in neoclassical consumer economics is the choice decision made by consumers and the associated motives during the budgetary allocation activities (Moscati, 2007). The choices refer to consumption baskets that provide a mix of all goods purchased by consumers at given prices and the consumer purchasing power. The aforementioned micro foundation provides the concept of the theory of consumer demand. The theory of consumer demand is built by concepts of utility, commodity set, and the consumer preference axioms (Raunikar and Huang, 1987).

In the theory of consumer demand, the commodity set must consist of commodities that are non-negative, divisible and unbounded. The preference axioms include those that show the rationality of consumer preference (such as completeness[7], asymmetry, and transitivity[8]) and those that ensure that the continuous utility function is well behaving[9] (such as continuity, monotonicity[10], convexity[11] and differentiability). According to Leserer (2010), as long as the commodity set properties and the preference axioms hold, the utility function would also be in possession of nice properties such as quasi-concavity[12], monotonicity, order-preserving, real-value and continuity. Deaton (1986) explains two extra utility axioms which facilitate the best choice that maximises consumer's utility namely non-satiation[13] and convexity.

Micro theory of consumer demand is based on utility maximization framework and the standard tactic of demand analysis is the estimation of the following general equation;

$$q_{mt} = f_m(y_t, p_{mt}, z_{mt}, u_{mt}) \qquad m=1,2,\ \ldots,\ h \qquad (1)$$

[7] Consumer is able to rank the bundles and choose between them
[8] If an individual report that "A is preferred to B" and "B is preferred to C," then he or she must also report that "A is preferred to C."
[9] This enables a consumer to get quantitative information that can be used to explain or forecast consumer behavior
[10] Monotonic transformations, by definition, preserve the order of the relationship between the arguments of a function and the value of that function
[11] Average is preferred to extremes
[12] Quasi-concave functions have the property that any two points in the set can be joined by a line contained completely within the set.
[13] More is preferred to less

where q_{mt} is per capita consumption of commodity m in time t, y_t is per capita income, p_{mt} are prices at time t, z_{mt} refers to consumer characteristics at time t, and u_{mt} is error term. The *m* equations can be estimated by a single equation or by a system of equations. In this study, a replica of equation (1) was estimated in a budget share form through the AIDS model (refer to section 3.4.2.1).

Demand functions are derived from a constrained optimization process (Nicholson and Snyder, 2008). A budget constrained optimization results in Marshallian (uncompensated) demand functions. In this optimization, utility is maximized subject to a given level of income. Mathematically, the objective of the consumer is

$$\max u(q) \text{ } subject \text{ } to \text{ } \sum_{m=1}^{n} p_m q_m = y \tag{2}$$

where *u* refers to utility, p_m is price of commodity *m*, q_m is quantity of commodity *m*, *n* is the total number of commodities, and *y* is income. Using mathematical optimization techniques like the Lagrangean multiplier technique, the above optimization results in Marshallian (uncompensated) demand function in the following form

$$q^* = q(p, y) \tag{3}$$

The Marshallian demand function shows the relationship between prices and the quantity demanded of a commodity while holding income constant (Nicholson and Snyder, 2008; Varian, 1987). According to Nicholson and Snyder (2008), the properties regarding Marshallian demand functions include homogeneity[14], adding up[15], negativity[16] and symmetry[17].

The reformulation of the constrained optimization in equation (4) is where the total expenditure is minimized subject to a certain utility level and this problem is described as a dual to the former. This process is mathematically expressed as

$$\min \sum_{m=1}^{n} p_m q_m \ \ subject\ to\ U(q) = \bar{U} \quad (4)$$

where $\sum_{m=1}^{n} p_m q_m$ is the cost/expenditure function and $\bar{U}$ is a given utility level. Using mathematical optimization techniques such as the Lagrangean multiplier technique, the above optimization results in Hicksian (compensated) demand functions which show the relationship between quantity demanded, prices and utility. The Hicksian demand function is expressed as follows

$$q^* = h(p, U) \quad (5)$$

It has to be noted though that much as the Marshallian[18] and Hicksian[19] demand functions seem different in their estimation; their solutions are identical thus present a

[14] The Marshallian demands are supposed to be homogenous of degree zero in prices and income thus the quantity demanded should remain unchanged if both prices and income are multiplied by a constant *k*.

[15] The budget allocation of the Marshallian demands must exhaust the total available budget.

[16] The *n* by *n* matrix formed by $\partial q_m / \partial p_m$ must be negative semi definite thus all leading diagonals must be positive. The negativity property also stems from the quasi-concavity of the utility function.

[17] If one set of cross price elasticities and budget shares are known along with income and own price elasticities, another set of cross price elasticities could be calculated.

[18] Marshallian demand function is observable but not predictive

[19] Hicksian demand function is predictive but not observable

duality principle. This duality[20] method provides a link between Marshallian and Hicksian demand functions and allows the possibility of estimating one given the other[21] (Wetzstein, 2004).

The Hicksian demand functions are homogenous of degree zero in prices[22] (Varian, 1987). After the derivation of Hicksian demand functions from the Lagrangean multiplier technique, the Hicksian demand functions can be substituted into the cost/expenditure function. This yields the minimized expenditure function, which becomes a function of prices and utility as follows

$$\sum_{m=1}^{n} p_m q_m = C^*(p, U) \tag{6}$$

The minimized cost function shown in equation (7) is very important in demand analysis. This applies in terms of its properties, which are useful in understanding demand restrictions. The function is homogenous of degree one[23] in prices, increasing function of prices and utility, and continuous and concave[24] in prices (Nicholson and Snyder, 2008). In addition, the derivative of the minimized cost function with respect to prices yields the Hicksian demand functions (Shepard, 1953). This is the Shepard lemma principle and is mathematically presented as

$$\frac{\partial C}{\partial p_m} = q^* = h(p, U) \tag{7}$$

[20] It's good to estimate both Marshallian and Hicksian to get better insights into the nature of demand. However, conclusions are made on the Marshallian estimates because these provide the ordinary case
[21] The Marshallian demand functions could be derived by substituting the inverse of expenditure function into the Hicksian demand function
[22] If prices are multiplied by a constant *k*, the quantity demanded remain unchanged
[23] Multiplying the prices by *k* increases the cost by the same factor
[24] Concave function resemble inverted t-cup shape and they always lie below any plane that is tangent to them

The Shepard lemma principle is very important in empirical demand models. It allows for estimation of demand functions from cost functions. Expositions of how the neoclassical theory of consumer demand works have applied the Shepard lemma principle to derive famous demand models. For instance, Deaton and Muellbauer (1980) used the Shepard lemma principle and a few derivatives to derive the AIDS model from a cost function that belongs to a preference class known as the Price Independent Generalized Logarithmic (PIGLOG) class (refer section 3.4.2.1).

3.4 Analytical framework

3.4.1 Choice of demand model

In demand system estimation, there is no clear-cut answer to the best model to use in cross-sectional data analysis. The answer lies with the comparison of different models according to their assumptions, their popularity, strengths and weaknesses. Additionally, based on the data and before and after estimation, different pre-estimation and post-estimation diagnostic tests including specification tests can be performed to help choose a best-fitting demand model. Again, all this has to fall in the suite of finding the proper balance between realism and manageability as this is the essence of good modelling.

A number of demand models have been developed over the past decades. This study employed the AIDS model proposed by Deaton and Muellbauer (1980) due to its multiple advantages as shown in section 2.3.1. Due to estimation issues, different authors have developed extensions of the AIDS model. The commonly applied extensions of the AIDS model include the QuAIDS, LA/AIDS and IAIDS.

The QuAIDS model due to Banks, Blundell and Lewbel (1997) is a comprehensive and popular demand system in most empirical work on demand analysis. The model retains the attractive features of the AIDS model and allows for flexibility in Engel curves by including a quadratic term in logarithm of income thus allowing for adequate curvature in the Engel curves. However, besides its theoretical power, the model has been criticised for requiring large amount of data, having econometric and computational difficulties, having dimensionality problems as number of option increases, and having collinear prices for narrowly defined products (Taniguchi and Yen, 2000; Liu, 2006).

The IAIDS due to Eales and Unnevehr (1991) is an extension of the AIDS model that assumes supply on the market is predetermined and prices adjust so that the available quantity must be consumed, thus prices are not predetermined. This assumption is ideal for modelling perishables, which must be consumed shortly after harvest. For non-perishables, the aforesaid assumption is not viable since prices are mostly assumed to be predetermined and it is the supply that adjusts according to prices.

The LA/AIDS is an extension of the AIDS model that only differs with the original AIDS model in the price index used; which is a linear approximate of the translog price index in the original AIDS model (see section 3.4.2.1). The original AIDS model has been reported to bring a lot of estimation difficulties in most empirical applications (Alviola, 2010). This makes the LA/AIDS retain the attractive features of the AIDS model while having less econometric and computational difficulties. However, since the LA/AIDS assumes linear Engel curves, its theoretical power is compromised and weak (Khaliukova, 2013; Moschini, 1995).

Since this study assumes supply is not predetermined, the QuAIDS and the LA-AIDS are the two most competing variants of the AIDS model for this research. From the pre-estimation literature-based comparison of the two models, each model variant has merits and demerits; thus, this mode of choice would offer unreliable results. Bopape (2006, p41-44) describes one simpler pre-estimation specification test that can be done to determine whether the AIDS demand model should be specified in QuAIDS or LA/AIDS form. However, this study used both the pre-estimation and post-estimation methods to choose the best fitting model. The post-estimation methods required estimating both the QuAIDS and the LA/AIDS model[25]

3.4.2 Achieving the first specific objective

The first objective was achieved by estimating a complete demand system that featured the budget shares of the six common bean varieties analysed as dependent variables and prices and expenditure as explanatory variables. The estimation procedure was in reference to the known estimation issues of demand analyses presented in section 2.3.2. Two demand systems were estimated namely LA/AIDS and QuAIDS, one of which was selected to represent consumer demand in this research. The subsequent sections explain in detail how the first objective of this research was achieved.

3.4.2.1 Addressing the problem of zero consumption

As shown in section 2.3.2, two variants of the Heckman two-step approach namely Hein and Wessels (1990) and Shonkwiler and Yen (1999) estimators have been widely employed in most empirical work on demand analysis to address the problem of zero

[25] Results on the choice between QuAIDS and LAIDS are presented in chapter five

consumption of some commodities analysed during the survey period. This study utilized the latter to account for censored response problem due to its asymptotic efficiency and consistency in estimation over the former as reported by Barslund (2011), Musyoka et al. (2007) and Shonkwiler and Yen (1999). In the first step, a selection model was estimated from which the probability density function (pdf) and the cumulative density function (cdf) were obtained. The pdf and cdf were then used as instruments in the second step (estimation of complete demand system) to correct for zero consumption.

Selection model: Multivariate probit

The multivariate probit model was used as a selection model for the first step of the Shonkwiler and Yen (1999) approach. In the multivariate probit, demand was modelled as binary decision of purchase for each of the six varieties of common bean analyzed. The multivariate probit was used because its likelihood function is well behaved as it gives consistent Maximum Likelihood Estimate (MLE) of the coefficients (β) and standard errors of the estimates (Maddala, 1983). Furthermore, the multivariate probit was preferred in estimation or over separate probit models by each variety because the latter can result in endogeneity within the data (error terms from different probits might be correlated) since information about consumption of different varieties of common bean is corrected from one subject at a point in time. Following Cappellari and Jenkins (2003), the *m*-equation multivariate probit was specified as follows;

$$y_{im}^{*} = \beta_m{}' X_{im} + \epsilon_{im} \qquad m = 1, \dots, 6 \ , \ i = 1, \dots, n \tag{8}$$

$$y_{im}^{*} = 1 \ if \ y_{im}^{*} > 0 \ and \ 0 \ otherwise \tag{9}$$

where y_{im}^{*} is the probability of purchasing variety m for observation i, X_{im} is a vector of explanatory variables, β_m represent the unknown parameters to be estimated, and ϵ_{im} represent error terms with a multivariate normal distribution (refer section 3.4.2.1 for variables to be included in multivariate probit). Since there were six varieties that were analyzed (i.e. m=1, …,6); the multivariate probit consisted of six equations. The multivariate probit was estimated via simulation likelihood. The log likelihood function for a sample of *n* independent observations is given by;

$$L = \sum_{i=1}^{n} w_i \log \Phi_m \, (\mu_i; \vartheta) \qquad (10)$$

where w_i is an optional weight for observation $i = 1, \dots, n$, and $\Phi_m(.)$ is a multivariate standard normal distribution function with arguments μ_i and ϑ. $\mu_i = (K_{im}' X_{im})$ with $K_{ik} = 2y_{ik} - 1$ for each $i, k = 1, .., m$. ϑ is a matrix with elements ϑ_{jk} where $\vartheta_{JJ} = 1$ for $j = 1, \dots, m$ and $\vartheta_{jk} = \vartheta_{kj} = \mathrm{k}_{ij}\mathrm{k}_{ik}\rho_{jk}$.

Since the multivariate model extends to more than two outcome variables, the practical obstacle to such a procedure is the evaluation of higher-order multivariate normal integrals (Greene, 2003). Standard linear methods can’t work in this case and favor falls to simulation based methods. Simulation allows us to estimate otherwise intractable models by higher dimension integrals (Cappellari and Jenkins, 2003). Greene (2003, p714) and Hajivassiliou and Ruud *(*1994) present a number of the simulation based methods that have emerged over the past years to evaluate higher order multivariate normal integrals. However, all these have fallen out of favor due to numerous estimation issues. The only reliable method of evaluating multivariate normal integrals by both papers is the Geweke–Hajivassiliou–Keane (GHK) smooth recursive conditioning simulator[26].

The merit of GHK simulator over its counterparts is that it exploits the fact that a multivariate normal distribution function can be expressed as the product of sequentially conditioned univariate normal distribution functions, which can be easily and accurately evaluated (Cappellari and Jenkins, 2003). Due to the aforesaid, this study employed the GHK simulator to evaluate the multivariate normal integral in multivariate probit model. One key issue during estimation was the choice of number of draws for the GHK estimator. The default is the square root of the sample size. However, Cappellari and Jenkins (2003) recommend higher number of draws for precision even though this increases the computation time of the multivariate probit model. They further offer a guide on the choice of number of draws, which is to keep increasing the number of draws until no further larger differences occur in the estimates. For this study, this point was achieved at 150. However, the study deliberately assigned 1000 draws to the GHK simulator to get the highest possible precision even though this increased the computation time of the multivariate probit model by 2500%. To further ensure accurate results, the study controlled for heteroscedasticity during estimation by adding the robust option in order to obtain the heteroskedastic-consistent standard errors. After estimating the multivariate probit, the Cumulative Distribution Function (Φ) and Probability Distribution Function (ϕ) were obtained and then used as arguments in the second stage of our analysis which involved estimation of the LA/AIDS and QuAIDS.

[26] GHK simulator is also encouraged by Cappellari and Jenkins (2003) who authored the multivariate probit model

Outcome model: QuAIDS and LA/AIDS

The AIDS model is derived from a cost function that belongs to a specific class of preference known as the PIGLOG class (for derivation see Deaton and Muellbauer 1980, pp 313-314). The cost function defines the minimum expenditure necessary to attain a specific utility level at given prices. Following Deaton and Muellbauer (1980), the AIDS in budget share form is specified as follows

$$w_j = \alpha_j + \sum_{k=1}^{m} \gamma_{jk} \ln(p_k) + \beta_j \text{In}(X/P) \qquad j, k = 1, \ldots, 6 \tag{11}$$

where in this study, w_j is the budget share of variety j of common bean and was computed as $p_j q_j / X$, α_j and γ_{jk} are parameters, p_k is the price of variety k and was derived implicitly from the respective quantities and expenditures, and X is the total expenditure on the six common bean varieties analyzed and equals $\sum_{m=1}^{6} p_m q_m$ section 3.4.2.1 describes variables that were included in the demand system estimation). P is a Translog price index defined by

$$\log P = \alpha_0 + \sum_j \alpha_j \log P_j + \frac{1}{2} \sum_j \sum_k \gamma_{jk}^* \log P_j \log P_k \tag{12}$$

where α_0 is the intercept parameter. The parameters γ_{jk} are defined as follows:

$$\gamma_{jk} = \frac{1}{2}(\gamma_{jk}^* + \gamma_{kj}^*) \tag{13}$$

The inclusion of the price index in the AIDS model helps to solve the non-reliability issues of income. However, the price index in equation (12) in non-linear in parameters and this brings about estimation difficulties in the AIDS model (Deaton and Muellbauer, 1980). In addition, the empirical plausible value of α_0 is not provided by any household theory (Taniguchi and Chern, 2000). To remedy this problem, Deaton and Muellbauer (1980) proposed a linear approximate of the Translog price index known as the Stone's price index. The Stone's price index is linear and easy to estimate and is specified as follows;

$$\ln(P^*) = \sum_{i=1}^{6} w_j \ln(p_j) \tag{14}$$

Equation (14) links with equation (11) in the way that $P = \varphi P^*$ and $E(\ln(\varphi)) = \alpha_0$. The use of the Stones price index in the AIDS model changes the model into the LA/AIDS. According to Moschini (1995), the Stone's price index is variant to changes in unit of measurement and this can result in unit measurement error. To remedy this problem, the Stone price index should be obtained by replacing w_j in equation (14) by mean budget shares, w_j^0 (as shown by Moschini, 1995 and also as applied by Khaliukova, 2014). This changes the stone price index into the Laspeyres price index and the LA/AIDS with the Laspeyres price index therefore becomes;

$$w_j = \alpha_j^* + \sum_k \gamma_{jk} \text{In}(p_k) + \beta_j(\text{In}(\text{X}) - \sum_{j=1}^{m} w_j^0 \ln(p_j)) + \mu_j^* \tag{15}$$

There are a number of price index formulae developed, each with its own merits and demerits. Two methods are often used to assess the suitability of a price index formula and these include the economic theory approach and the axiomatic approach. Both of these approaches suggest the Fischer and Tornqvist price indexes (which bear quadratic and translog structures, respectively) as the best (Asche and Wessels, 1997). However, to preserve the simplistic linear structure of the LA/AIDS model and truly approximate the original translog price of the AIDS model, this study limited its choice to the family of price indexes with simple linear production structures and chose the Laspeyres price index as described in equation (15). The study used Shonkwiler and Yen (1999) approach to remedy the censored response problem and the LA/AIDS with Φ and ϕ from the multivariate probit was estimated in the following form:

$$w_j = \alpha_j^* + \Phi\left(\sum_k \gamma_{jk} \operatorname{In}(p_k) + \beta_j(\operatorname{In}(\mathrm{X}) - \sum_{j=1}^{6} w_j^0 \ln(p_j))\right) + \delta\phi + \mu_j^* \quad (16)$$

The theory of demand requires that the above system be estimated under restriction of adding up[27], homogeneity[28] and symmetry[29]. The theoretical demand restrictions in terms of adding up, homogeneity in prices and income and the symmetry of cross effects of demand functions are expressed as:

Adding up: $\sum_j w_j = 1$, $\sum_{j=1}^{6} \alpha_j = 1$, $\sum_{j=1}^{6} \gamma_{jk} = 0$, $\sum_{j=1}^{6} \beta_j = 0$

Homogeneity: $\sum_k \gamma_{jk} = 0$

Symmetry: $\gamma_{jk} = \gamma_{kj}$

A generalization of the PIGLOG class of preference gives birth to the QuAIDS model. The QuAIDS model starts with an indirect utility function (for derivation see Banks, et

[27] All the budget shares must some up to one
[28] All the price coefficients for each equation should sum up to zero
[29] Coefficient of price A on budget share B should equal to coefficient of price B on budget share A

al.,, 1997). It assumes non-linear Engel curves and non-constant expenditure elasticity hence differing with the LA/AIDS model that assumes linear Engel curves and constant expenditure elasticities (Taniguchi and Chern, 2000). Thus, the difference in the functional form of the QuAIDS and LA/AIDS is the addition of the quadratic expenditure term in the QuAIDS to overcome the limitation of the inflexibility in expenditure. The QuAIDS model due to Banks, et al., (1997) is specified as follows in budget shsre form;

$$w_j = \alpha_j + \sum_{k=1}^{m} c_{jk} \log(p_k) + \beta_j \log\left(\frac{X}{\alpha(p)}\right) + \frac{d_i}{\beta(p)}\left\{log\left[\frac{X}{\alpha(p)}\right]\right\}^2 \quad (17)$$

where α_j is the intercept parameter. $\log a(p)$ is the price index and is defined as follows

$$\log \alpha(p) = \alpha_0 + \sum_j \alpha_j \log P_j + \frac{1}{2}\sum_j \sum_k c_{kj}^* \log P_j \log P_k \quad (18)$$

$\beta(p)$ and the parameters c_{jk} are defined as follows:

$$\beta(p) = \prod_{i=1}^{m} p_i{}^{b_i} \quad (19)$$

$$c_{jk} = \frac{1}{2}(c_{jk}^* + c_{kj}^*) \quad (20)$$

The study replaced the price index defined in equation (18) by the Laspyeres price index as it happened with the LA/AIDS. In addition, the QuAIDS was modified using the Shonkwiler and Yen (1999) approach, by multiplying all the predictors by Φ and adding ϕ as a predictor. Hence the estimated QuAIDS was as follows

$$w_j = \Phi\left(\alpha_j + \sum_{k=1}^{m} c_{jk} \log(p_k) + \beta_j \log\left(\frac{X}{I}\right) + \frac{d_i}{\beta(p)}\left\{log\left[\frac{X}{I}\right]\right\}^2\right) + \delta\phi + \mu_j \quad (21)$$

where I is the Laspyeres price index and Φ and ϕ are cdf and pdf, respectively, estimated in first step using multivariate probit. The theoretical demand restrictions in terms of adding up, homogeneity in prices and income and the symmetry of cross effects of demand functions are expressed as

Adding up: $\sum_j w_j = 1, \sum_{j=1}^{6} \alpha_j = 1, \sum_{j=1}^{6} c_{jk} = 0, \sum_{j=1}^{6} \beta_j = 0, \sum_{j=1}^{6} d_{jk} = 0$

Homogeneity: $\sum_{k=1}^{6} c_{jk} = 0$

Symmetry: $c_{jk} = c_{kj}$

The LA/AIDS and QuAIDS were estimated using Iterative Seemingly Unrelated Regression (ITSUR) and Iterative Full Generalized Non-Linear Square (IFGNLS) routines respectively. Theoretical restrictions of adding-up, homogeneity, and symmetry were imposed during estimation. In the estimation of the LA/AIDS, the ITSUR estimator is more efficient than OLS estimator and controls for error correlation across equations (Asche and Wessells, 1997). In fact, the ITSUR uses the correlations in the errors of the other equations to improve parameter estimates (Taylor, 2014). To satisfy the demand restrictions and to avoid singularity in the covariance matrix among residuals; one of the six demand equations was dropped from the system of equations in each model. The ITSUR and IFGNLS routines produce results that are invariant to the equation deleted. The parameters from the deleted equation in each model were calculated from the parameters of the other equations through the restrictions on parameters.

Choice of variables for analysis

Variables that were included in the multivariate probit model

Following Cappellari and Jenkins (2003), the dependent variables in the multivariate probit were the decisions to purchase each of the six common bean varieties analysed. Each dependent variable was binary with values of 1 if the household purchased the respective variety and 0 if the household did not purchase the variety during the survey period. The explanatory variables included in the multivariate probit were the same for each of the six equations.

The explanatory variables included were marital status of household head, household income, and household size, location of household and price of each variety. The variables were chosen based on consumer behaviour theory and the nature of the data set. Some important socioeconomic variables (e.g. gender, education, occupation and age of household head) and bean characteristics (e.g. grain size, colour, taste, and cooking time) specified in the conceptual framework of this study in section 2.2 were not included in the multivariate probit due to data limitations. Much as the common bean characteristics were not included in the selection equation, literature (Mishili et al., 2009; Chirwa and Phiri, 2007) explain that these are already reflected in the prices. The subsequent sections explain in detail each of the variables that were included in the multivariate probit and Table 3.1 gives a summary.

A. Marital status of household head

Marital status was included in the multivariate probit because it affects household consumption behaviour. This study measured marital status of the household head as a dummy with values of 1 for 'Married (Monogamous/Polygamous)' and 0 for ‘otherwise’. Theory provides no *a priori* expectation of the relationship between marital status and food purchase decisions. This being the case, the study expected a positive or a negative value for the coefficient on marital status in each of the probits. Lazaro (2014) reported both positive and negative coefficients of marital status on demand of high quality and average quality rice in Tanzania. In another study, Khaliukova (2014 noted that married households were more likely to buy onions in Nigeria due to high frequency of cooking meals.

B. Household size

Household consumption decisions are made reflecting on the number of people in the household. For this reason, household size was included in the multivariate probit and was measured as a continuous variable[30]. Theory does not provide any *a priori* expectations with regard to household size and choice of food. Therefore, the study did not expect any specific sign on the coefficient of household size in each equation of the multivariate probit. Agostini (2014) revealed that in Britain; an increase in household size increases the demand for dairy products, fat and sugar and cereals while it decreases the demand for meat, fish, vegetables and fruits. This is because the dairy products, fat, sugar and cereals belonged to a necessity food basket for most households while the meat, fish, vegetables, fruits varied as necessities or luxuries depending on type of household. However, Kawabata (2011) argues that in Malawi, an increase in household size turns foods like rice, wheat, beans, meat, and milk into luxuries thereby reducing their probability of purchase.

[30] The Amsterdam scale was used to compute adult equivalents for each member in the household. Based on the scale and the nature of the dataset: below 5 years equaled 0.52, between 5 and 12 years equaled 0.8, between 12 and 18 years equaled 0.94, and above 18 years equaled 1.

C. Household income

Based on theory of consumer behaviour, household income is one of the most important economic variables in a choice model. Essentially, household income determines consumption basket and level of spending. In this study, household income was measured as a continuous variable in local currency. The logarithm of household income was included in the multivariate probit. A number of economic theories on spending behaviour have been developed over the past decades and some examples include the absolute income hypothesis, relative income hypothesis, permanent income hypothesis, life cycle hypothesis and Engel law. These describe the link between consumption of food and income of the household. However, since the aforesaid description is at aggregate level and also because of mixed results on the relationship of income and consumption of legumes presented in section 2.3; this study has no *a priori* expectation on the relationship between household income and the decision to purchase a variety of common bean.

D. Household location

Availability of and access to food are often affected by location. Households in different locations tend to have different consumption patterns. Thus, it was necessary to include location as a variable in the multivariate probit. Location was measured as a dummy for each density. Thus 3 dummies were available to show whether a household was located in high-density (1=Yes, 0=No), medium-density (1=Yes, 0=No) or low-density (1- Yes, 0=No) area. However, to prevent a dummy variable trap, only two of the dummies were included in the multivariate probit. High-density dummy was dropped from all equations and acted as a comparison category because it contained the largest proportion of the sample than its counterparts. Since the definite relationship of location and choice of food is not provided by theory, this study had no *a priori* expectations on location. Several demand studies have reported significant results on location. For instance, Khaliukova (2014 noted that location was significant in influencing the decision to purchase a number of vegetables. On the other hand, Aidoo (2009) revealed that location is significant in influencing demand for yam. Both Khaliukova (2014) and Aidoo (2009) made their arguments based on differences that exist in income and preferences by location.

E. Price of each variety

Theory of consumer behaviour explains that utility ultimately depends on prices and income of the consumer. For each of the varieties of common beans analysed, the associated price reflects a bundle of attributes. Thus, meaningful choice analysis can be done based on prices. Prices in this study were measured as continuous variables in local currency. The logarithm of each price was included in the multivariate probit. In terms of *a priori* expectations, this study expected negative signs on the coefficients of prices in all the equations of the multivariate probit, reflecting the law of demand. Much of the relationship between price and demand of beans was presented in section 2.3.

Table 3-1: Summary of variables that were included in the multivariate probit

Variable	Description	Expected sign
DVs		
p_{w1}	Decision to purchase *Kabalabala*	+
p_{w2}	Decision to purchase *Nanyati*	+
p_{w3}	Decision to purchase *Napilira*	+
p_{w4}	Decision to purchase *Khaki*	+
p_{w5}	Decision to purchase *Phalombe*	+
p_{w6}	Decision to purchase mixed beans	+
IVs		
Marital status	Binary (1-Married, 0-otherwise)	+/-
Household size	Continuous (adult equivalent)	+/-
Household income	Logarithm of total income in MK	+/-
Household location		
Low-density	Dummy (1=Yes, 0-otherwise)	+/-
Medium-density	Dummy (1=Yes, 0-otherwise)	+/-
Prices	Logarithm of price for each variety in MK	-

Variables that were included in the LA/AIDS and QuAIDS

The variables included in the LA/AIDS and the QuAIDS were the same except for the square term of the logarithm of expenditure which was added in the QuAIDS model. The dependent variables for both models were the budget share for each of the six common bean varieties analysed. Based on the neoclassical micro theory of consumer demand, the independent variables included price of each variety and the total expenditure in logarithmic form. Table 3-2 summarizes the variables used in the demand system estimation, their definitions, and a *priori* expectations based on the theory of consumer demand. The study expected a negative coefficient on logarithm of prices since the law of demand states that demand is negatively related to price. The study also assumed common bean is a normal good and thus has a positive coefficient on logarithm of expenditure term which is the proxy for income. In the QuAIDS model, a negative coefficient was assumed in the logarithm of the squared expenditure term to reflect diminishing returns. Table 3.2 gives a summary of the variables that were included in the LA/AIDS and the QuAIDS model.

Table 3-2: Summary of variables included in the LA/AIDS and QuAIDS

Variable	**Description**	**Expected sign**	
		LA/AIDS	QuAIDS
DVs			
w_1	Budget share of *Kabalabala*	+	+
w_2	Budget share of *Nanyati*	+	+
w_3	Budget share of *Napilira*	+	+
w_4	Budget share of *Khaki*	+	+
w_5	Budget share of *Phalombe*	+	+
w_6	Budget share of mixed beans	+	+
IVs			
ln_{p1}	Log of price of *Kabalabala*	-	-
ln_{p2}	Log of price of *Nanyati*	-	-
ln_{p3}	Log of price of *Napilira*	-	-
ln_{p4}	Log of price of *Khaki*	-	-
ln_{p5}	Log of price of *Phalombe*	-	-
ln_{p6}	Log of price of mixed beans	-	-
ln_{exp}	Log of expenditure term	+	+
$ln_{(exp)^2}$	Log of squared expenditure term		-

Empirical models

With a series of equations presented in section 3.4.2, it is essential to present the models empirically in the way they were analyzed using the afore-presented variables. The study estimated both the LA/AIDS and QuAIDS models. Empirically, using variables presented in section 3.4.2.1 the LA/AIDS estimated six equations namely:

$$Budget\ share_{Kabalabala} = \alpha_0 + cdf_1\{\beta_{11} lnp_{kabalabala} + \beta_{12} lnp_{napilila} + \beta_{13} lnp_{phalombe} + \beta_{14} lnp_{Khaki} + \beta_{15} lnp_{nanyati} + \beta_{16} lnp_{mixed\ beans} + \delta_1 lnexp\} + ð_1 pdf_1 + \mu_1$$

$$Budget\ share_{Napilila} = \alpha_0 + cdf_2\{\beta_{21} lnp_{kabalabala} + \beta_{22} lnp_{napilila} + \beta_{23} lnp_{phalombe} + \beta_{24} lnp_{Khaki} + \beta_{25} lnp_{nanyati} + \beta_{26} lnp_{mixed\ beans} + \delta_1 lnexp\} + ð_1 pdf_2 + \mu_2$$

$$Budget\ share_{Phalombe} = \alpha_0 + cdf_3\{\beta_{31} lnp_{kabalabala} + \beta_{32} lnp_{napilila} + \beta_{33} lnp_{phalombe} + \beta_{34} lnp_{Khaki} + \beta_{35} lnp_{nanyati} + \beta_{36} lnp_{mixed\ beans} + \delta_1 lnexp\} + ð_1 pdf_3 + \mu_3$$

$$Budget\ share_{Khaki} = \alpha_0 + cdf_4\{\beta_{41} lnp_{kabalabala} + \beta_{42} lnp_{napilila} + \beta_{43} lnp_{phalombe} + \beta_{44} lnp_{Khaki} + \beta_{45} lnp_{nanyati} + \beta_{46} lnp_{mixed\ beans} + \delta_1 lnexp\} + ð_1 pdf_4 + \mu_4$$

$$Budget\ share_{Nanyati} = \alpha_0 + cdf_5\{\beta_{51} lnp_{kabalabala} + \beta_{52} lnp_{napilila} + \beta_{53} lnp_{phalombe} + \beta_{54} lnp_{Khaki} + \beta_{55} lnp_{nanyati} + \beta_{56} lnp_{mixed\ beans} + \delta_1 lnexp\} + ð_1 pdf_5 + \mu_5$$

$$Budget\ share_{Mixed\ beans} = \alpha_0 + cdf_6\{\beta_{61} lnp_{kabalabala} + \beta_{62} lnp_{napilila} + \beta_{63} lnp_{phalombe} + \beta_{64} lnp_{Khaki} + \beta_{65} lnp_{nanyati} + \beta_{66} lnp_{mixed\ beans} + \delta_6 lnexp\} + ð_1 pdf_6 + \mu_6$$

Empirically, using variables presented in section 3.4.2.1 the QuAIDS also estimated six equations namely;

$$Budget\ share_{kabalabala} = \alpha_0 + cdf_1\{\beta_{11}lnp_{kabalabala} + \beta_{12}lnp_{napilila} + \beta_{13}lnp_{phalombe} + \beta_{14}lnp_{Khaki} + \beta_{15}lnp_{nanyati} + \beta_{16}lnp_{mixed\ beans} + \delta_1 lnexp + \partial_1(lnexp)^2\} + ð_1 pdf_1 + \mu_1$$

$$Budget\ share_{Napilila} = \alpha_0 + cdf_2\{\beta_{21}lnp_{kabalabala} + \beta_{22}lnp_{napilila} + \beta_{23}lnp_{phalombe} + \beta_{24}lnp_{Khaki} + \beta_{25}lnp_{nanyati} + \beta_{26}lnp_{mixed\ beans} + \delta_1 lnexp + \partial_1(lnexp)^2\} + ð_1 pdf_2 + \mu_2$$

$$Budget\ share_{Phalombe} = \alpha_0 + cdf_3\{\beta_{31}lnp_{kabalabala} + \beta_{32}lnp_{napilila} + \beta_{33}lnp_{phalombe} + \beta_{34}lnp_{Khaki} + \beta_{35}lnp_{nanyati} + \beta_{36}lnp_{mixed\ beans} + \delta_1 lnexp + \partial_1(lnexp)^2\} + ð_1 pdf_3 + \mu_3$$

$$Budget\ share_{Khaki} = \alpha_0 + cdf_4\{\beta_{41}lnp_{kabalabala} + \beta_{42}lnp_{napilila} + \beta_{43}lnp_{phalombe} + \beta_{44}lnp_{Khaki} + \beta_{45}lnp_{nanyati} + \beta_{46}lnp_{mixed\ beans} + \delta_1 lnexp + \partial_1(lnexp)^2\} + ð_1 pdf_4 + \mu_4$$

$$Budget\ share_{Nanyati} = \alpha_0 + cdf_5\{\beta_{51}lnp_{kabalabala} + \beta_{52}lnp_{napilila} + \beta_{53}lnp_{phalombe} + \beta_{54}lnp_{Khaki} + \beta_{55}lnp_{nanyati} + \beta_{56}lnp_{mixed\ beans} + \delta_1 lnexp + \partial_1(lnexp)^2\} + ð_1 pdf_5 + \mu_5$$

$Budget\ share_{Mixed\ beans} = \alpha_0 + cdf_6\{\beta_{61} lnp_{kabalabala} + \beta_{62} lnp_{napilila} +$

$\beta_{63} lnp_{phalombe} + \beta_{64} lnp_{Khaki} + \beta_{65} lnp_{nanyati} + \beta_{66} lnp_{mixed\ beans} + \delta_6 lnexp +$

$\partial_1 (lnexp)^2\} + ð_1 pdf_6 + \mu_6$

Assumptions tested

The Shonkwiler and Yen (1999) two-step estimation method assumes normality in the residuals of both the selection and the outcome equation. Additionally, it also assumes no correlation between the residual of the selection equation and the outcome equation. These assumptions were tested during estimation and the results are given in chapter five. Additionally, several diagnostic tests were performed to assess the quality of the data, the equations formulated and the performance of the models.

Testing hypothesis associated with objective one

With regard to objective one, this study hypothesized that prices and income affect household budget share of common beans. In order to test this hypothesis; the study estimated the demand system which featured the budget shares of the six varieties of common beans as dependent variables and prices and income as explanatory varieties. Thereafter, the significance of prices and income in the budget share equations of the chosen demand system among LA/AIDS and QuAIDS was checked and decision of whether or not to reject the aforesaid hypothesis was reached. Chapter 5 presents the results from the demand system estimation and testing of the hypothesis.

3.4.3 Achieving the second specific objective

Price and expenditure elasticities show the responsiveness of demand to changes in prices and household expenditure, respectively. Taniguchi and Chern (2000) express the elasticity formulae for calculating demand elasticities from the LA/AIDS. Following their procedure, which starts by taking the derivative of equation (11) with respect to In(p_k), this research augments by including the Shonkwiler and Yen (1999) modification (see equation (16)) to their equation. Hence the correct formula for calculating uncompensated price elasticity of commodity *j* with respect to commodity *k* multiplies their formula by the cdf and is as follows

$$\eta_{jk} = \Phi_i \left(\frac{\gamma_{jk}}{w_j} - \frac{\beta_j w_k}{w_j} - \delta_{jk}\right) \qquad \forall j, k = 1, \ldots, 6 \qquad (22)$$

where Φ_i is the cdf for equation i and δ_{jk} is the Kronecker delta[31]. Under Slutsky model, the compensated price elasticities, s_{jk}^*, were computed from

$$s_{jk}^* = e_{jk} + e_j w_k \qquad (23)$$

Following Taniguchi and Yen (2000), the expenditure elasticity is derived by taking the derivative of equation (11) with respect to $\ln(x)$. This research multiplies their result by the cdf to apply the Shonkwiler and Yen (1999) correction and the formula becomes

$$e_j = \Phi_i \left(1 + \left(\frac{\beta_j}{w_j}\right)\right) \qquad (24)$$

[31] Kronecker delta equals one if $j = k$ and zero otherwise

The idea behind derivation of demand elasticities for the LA/AIDS is the same as for the QuAIDS, only differing because of the difference in the functional forms of the models, which form an initial step in the derivation. Following Maganga et al. (2014) who also applied the Shonkwiler and Yen (1999) approach, the expenditure, uncompensated and compensated price elasticities from the QuAIDS model are presented in equation 25-27, respectively

$$e_j = \Phi_i \left\{ \frac{\beta_j + \frac{2d_i}{\beta(p)}\left[\log\left(\frac{X}{I}\right)\right]}{w_j} + 1 \right\} \quad (25)$$

$$\eta_{jk} = \Phi_i \left\{ \frac{c_{jk} - \mu_j\left(\alpha_j + \sum_{k=1}^{m} c_{jk} \log(p_k)\right) + \frac{2d_i}{\beta(p)}\left[\log\left(\frac{X}{I}\right)\right]}{w_j} - \delta_{jk} \right\} \quad (26)$$

$$s_{jk}^{*} = e_{jk} + e_j w_k \quad (27)$$

It is an often practice in most empirical research to use expenditure elasticity as a proxy to income elasticity since the AIDS model and its variants do not produce an estimate of income elasticity directly[32]. However, much as this is acceptable; with research focused on more disaggregated food items, the deviation of the expenditure term from the true income term might be high and this can cause wrong inferences to be made about the nature of demand for products analysed. Thus, the estimation of the true income elasticity was necessary for this study. The study followed Chern et al. (2003, p16) suggestion which has also been applied by Sheng et al. (2010), of first estimating the following Engel function in order to calculate the income elasticity;

$$\log x = \alpha_0 + \alpha_1 \log X + \beta \log P + \varepsilon \quad (28)$$

[32] The AIDS model and its variants uses total expenditure for the group of food items analyzed in order to satisfy the adding up property

where x is total expenditure of the foods included in the model, X is total expenditure on food and non-food consumer goods and services and is synonymous to total income, and P is Laspeyres price index for the foods included in the model. The study estimated the Engel function within the framework of necessary assumptions[33] governing OLS. After estimating the Engel function, the income elasticity was found by multiplying the responsiveness of expenditure x to change in income by the expenditure elasticity. The responsiveness of expenditure on the items analysed (e_i) was calculated as follows:

$$e_i = \frac{dx}{dX} * \frac{X}{x} \tag{29}$$

Thus, the income elasticity[34] is calculated as follows;

$$e_y = e_i * e_j \tag{30}$$

3.4.3.1 Testing hypotheses associated with objective two

The elasticities calculated using equations (22) to equation (30) were compared against their hypothesized values (see section 1.5) according to theory. Decision of whether or not to reject hypotheses on the elasticities were made based on the aforesaid comparisons and chapter 5 presents and explains the results.

[33] Assumptions necessary for OLS estimator include homoscedasticity, no multicollinearity, normality, linearity, continuity, no endogeneity, no autocorrelation, and zero covariance.
[34] Results on the Engel function are presented in appendix A. 1

3.5 Data sources, sampling procedure, and sample size determination

This study used cross-section secondary data collected in the year 2015 in Lilongwe district of Malawi through the bean consumption survey conducted by the bean value chain research network; a collaborative research initiative comprising Lilongwe University of Agriculture and Natural Resources, Sokoine University of Agriculture, University of Zambia and Kansas State University with funding from the legume innovation laboratory, a USAID-funded initiative. Lilongwe was chosen because it is more populated and diverse as compared to the other cities in Malawi and thus represents a potential viable market for beans.

Lilongwe is the capital and largest city of Malawi with an estimated population of 1,077,116 people as of 2015 (National Statistical Office (NSO) Malawi, 2016). Located 1050m above sea level in the central region of Malawi; the district is an important economic and transportation hub for central Malawi. The district features a humid subtropical climate (temperature ranges 18-24 °C) with warm summers and mild winters. The economy of the district is dominated by government and public institutions and the main economic activities comprise of retail trade, finance, public administration, tourism, banking, transport and tobacco manufacturing (NSO Malawi, 2016). Agriculture is minimal in the district with much influx of crops on the markets from other districts. The major crops cultivated include maize, tobacco, beans, cassava, groundnuts and sweet potatoes.

A multistage stratified sampling procedure was employed to select the required sample. First, Lilongwe was stratified into high, medium and low density areas. The stratification was based on income levels with high to low density areas representing low to high income levels, respectively. This was done to ensure that all categories of people living in the city were represented in the sample. The second stage involved random selection of clusters within the strata. From high-density areas, area 8, 18, 21, 23, 36 and 49 were selected. Areas 6, 12, 14 and 47 were selected to represent medium-density areas. From low-density areas, area 3, 9, 10 and 43 were selected. The third and final stage involved randomly selecting households that formed the final sample.

A survey questionnaire with discrete choice experiment component was administered in the primary data collection. The questionnaire collected information on socioeconomic and demographic factors of the respondents. Additionally, the questionnaire collected information on household consumption decisions (i.e. the types of food products consumed in the past seven days and the quantities of different types of beans consumed, food expenditures, and decision making in the house) and bean consumer choices. In consumer choices, respondents were presented with a sequence of choice sets from which preferred alternatives were selected. Duly, the sample size determination made use of De Bekker-Grob et al. (2012) formula below, that took into account the design of the choice experiment.

$$n \geq 500 \times \frac{L^{max}}{JS} \times \ b \tag{31}$$

$$n \geq 500 \times \frac{3}{4 \times 8} \times 6$$

$$n \geq 276$$

where L^{max} is highest attribute level (which was 3), J is the number of alternatives (the experiment involved 4 dry common bean varieties), S is the number of choice tasks[35] for each respondent (which was 8), and b is the number of blocks[36] used in the survey (which was 6 blocks). The final sample calculated was 684 households. To arrive at this sample size, a 10% non-response rate was added and this brought the sample from 276 to 303 households. Furthermore, a design effect was accounted for and this brought the sample to 606. An additional 78 respondents were added to the sample to have adequate questionnaires to replace uncompleted questionnaires ending up with a sample size of 684. The sample size was then proportionally distributed to the strata and then to the clusters. A total of 584 households was sampled from high-density areas, 56 households from the medium-density areas, and 44 households from low-density areas.

The questionnaire did not collect information on prices, hence the prices were derived implicitly as expenditure divide by quantity of each of the common bean variety for each household. In order to obtain price data of households that did not consume a particular variety, the study followed Aepli (2014) by assuming that each household with zero consumption faces the mean price of the variety depending on the area of residence. The data was analysed in Stata Corp 14. Stata Corp 14 offered a rich and integrated environment for data analysis.

[35] A choice task represents a combination of attributes of a particular alternative which a respondent was asked whether they would buy or not buy.
[36] Blocks are sets of limited choice questions equal in size which are presented to survey respondents. Blocking is usually done to ensure efficiency of responses by minimising the number of choice questions presented to a survey respondent.

3.6 Summary

The modelling framework of this study was shaped by the neoclassical micro-theory of consumer demand. The QuAIDS model was used to represent consumer demand behaviour and achieve the objectives of the study. The model was estimated in a suite of addressing expenditure endogeneity and presence of zero consumption that often arise with cross-sectional data. Special attention was made to the assumptions governing the modelling framework. The study used secondary data collected in 2015 in Lilongwe, the capital city of Malawi.

CHAPTER FOUR

HOUSEHOLD CHARACTERISTICS

4.1 Introduction

Before any statistical analysis, an analyst needs to understand the data that is used. This becomes very pertinent when discussing model results. Accordingly, the purpose of this chapter is to present information on characteristics of households that composed the sample. The chapter gives a summary of household characteristics and other pertinent variables featured in the models that are presented in chapter five.

4.2 Socioeconomic characteristics of the households

The socioeconomic characteristics considered in this section include household size, household age structure, marital status of household head, employment status of the household head and spouse, and household income. The subsequent subsections describe the aforementioned socioeconomic characteristics in detail. The variable of analysis for all socio-economic characteristics is location categorised into high, medium and low-density areas.

4.2.1 Household size

Household size refers to number of people living in a household. Economists use adult equivalent scales to measure household size for both theoretical and applied reasons. Henceforth, this study used the Amsterdam scale[37] to measure household size. The motive behind the choice of the Amsterdam scale among its counterparts was the nature

[37] Adapting the Amsterdam scale to our data: less than 5 years equals 0.52 adults, between 5 and 12 equaled 0.8 adults, between 12 and 18 years equals 0.94 adults, and above 18 years equals 1 adult. The study used average of both sexes because gender of each household member was not present in the dataset.

of the dataset[38]. Table 4-1 gives a summary of household size by location. The overall average household size for the sample is 4.50 adults and there are no significant differences by location. NSO Malawi (2012) reported a similar average household size (4.6 adults) for Malawi, even though it is unclear of the adult equivalent scale that was used. The highest maximum (24.19 adults; probably a case of extended family) household size for the sample was found in the high-density areas while the lowest maximum (10.05 adults) was found in the low-density areas. This is in line with the fact that high-density areas contained more people per unit area and composed of relatively poor households (NSO Malawi, 2012).

Table 4.1 Household size by location

Location	***Household size (Adult equivalents)***				*Significance test*	
	Mean	Median	Min	Max		
High	4.49	4.17	0.95	24.19	High vs Medium	0.50
Medium	4.68	4.52	1.47	13.21	Medium vs low	0.58
Low	4.44	3.80	1.89	10.05	High vs low	0.88
Overall	4.50	4.17	0.95	24.19		

4.2.2 Household age structure

Age structure defines the distribution by age of household members. The data used in this study was collected categorically in segments of less than or equal to 5 years, above 5 years and less than or equal to 12 years, between 12 and 18 years, and above or equal to 18 years. Evidence from Table 4-2 shows that households in Lilongwe District are

[38] Age data was collected in <5, 5-12, 12-18, >18 years denominations and gender for each household member was not collected hence the Amsterdam scale best qualified with the nature of the dataset.

mostly composed of adults. GoM (2012) reports that age structure of Malawi varies by household wealth. Poorest households are more youthful with median age of 13 as compared to wealthiest households with median age of 17. This partly explains the trend in Table 4-2 since the sample was largely composed by urban households which are considered wealthier than their rural counterparts.

Table 4.2 Household age structure

Location	***Proportion of people by age (years) segments***				**N**
	≤5	>5 and ≤12	>12 and <18	>18	
High	13.08	18.69	10.27	57.96	2959
Medium	11.41	18.46	9.40	60.74	298
Low	10.17	18.22	8.47	63.14	236
Overall	12.74	18.64	10.08	58.55	3493

4.2.3 Marital status of household head

Marital status of the household head is an important variable since it determines a range of other household characteristics. Among the households interviewed, majority (72.34%) of the household heads are married/cohabiting while 13. 39% of the household heads are single/never married. Among the single household heads, about 4% and 10% are divorced and widowed, respectively. In the models presented in chapter five, marital status was treated as a dummy and thus single/never married, divorced, and widowed household heads were placed in one category namely single-headed households, which represented 27.66% of the sample. A comparison of the location shows that married household heads dominate in all the location and the percentage is highest (73.94%) in the high-density areas and lowest (55.36%) in the medium-density areas. Table 4-3 summarizes marital status of household heads in Lilongwe District of Malawi.

Table 4.3 Proportion of households by marital status and location

Density	*Sample proportions by marital status*				N
	Single/never married	Divorced	Widowed	Married	
High	12.61	4.09	9.37	73.94	584
Medium	23.21	7.13	14.29	55.36	56
Low	11.36	4.55	11.36	72.73	44
Overall	13.39	4.37	9.90	72.34	684

4.2.4 Employment status of the household head and spouse

Employment status of the household head and spouse is an important socioeconomic characteristic since it exhibits the financial position of the household. Results from Table 4-4 show that among the single household heads in all the locations, majority were employed (this confirms findings of NSO Malawi (2012)) and most were in salaried type of employment. An inter-comparison of the location shows that high-density areas contain the least (27.45%) proportion of single household heads in 'salaried employment' category. This was expected since high-density areas contain relatively poorer households. A similar trend is seen among the married household heads and in the pooled sample. Among the married household heads, there is no case where both the household head and spouse are unemployed in the medium and low-density areas and this is not odd since it is expected that these locations contained relatively wealthier households.

Table 4.4 Proportion of households by employment status

Density							N
	Single household heads						
	Salary		Self		Unemployed		
High	27.45		49.67		22.88		153
Medium	72.00		8.00		20.00		25
Low	66.67		8.33		25.00		12
	Married household heads						
	Salary		Self		Unemployed		
	Both	One	Both	One	Both	One	
High	9.91	7.14	14.98	19.59	3.69	44.70	431
Medium	45.16	12.90	6.45	25.81	0.00	9.68	31
Low	31.25	31.25	12.5	18.75	0.00	6.25	32

4.2.5 Household income

Income distribution is an important socioeconomic characteristic since it relates to household welfare and socioeconomic position. In Table 4-5; the mean and median monthly incomes for the sample were MK109,861.00 and MK65,000, respectively. An inter-comparison of the location shows that high-density area households have the lowest mean and median monthly incomes while low-density areas have the highest incomes. A significant test reveals that significant differences occur between the high density mean income is significantly lower than medium and high density areas. These results are consistent with the fact that the lower density areas contained relatively wealthier households and high-density areas contained relatively poorer households.

Results from Table 4-5 further show high variation in income across the sample as evidenced by the standard deviation which is higher than the overall mean and median incomes. Furthermore, income is positively skewed[39] (skewed to the right). This according to Nicholson and Snyder (2008) means that majority of the households have their incomes below the mean income. Kurtosis of the income distribution (51.72 in Table 4-5) shows that the distribution is leptokurtic[40], and this according to Greene (2003) and Maddala, (1983) means that more of the variance of income is the result of infrequent extreme deviations, as opposed to frequent modestly sized deviations.

Table 4.5 Household income by location

Location	***Household income (MK)***				***Significance test***	
	Mean	Median	Min	Max		
High	85,103.95	60,000.00	2500.00	700,000.00	High vs medium	0.00***
Medium	215,584.00	150,000.00	12,000.00	900,000.00	Medium vs low	0.14
Low	305,586.00	180,000.00	30,000.00	2,000,000.00	High vs low	0.00***
Overall	109,861.00	65,000.00	2,500.00	2,000,000.00		

Measures for spread and variation in income distribution

Std. deviation: 145,764.00

Skewness: 5.37

Kurtosis: 51.72

NB: Asterisks; ***significant at 1%

[39] Skewness is when data points cluster more toward one side of the scale than the other or the right and left side of the distribution are shaped differently from each other

[40] Leptokurtic is kurtosis that is above 3

4.3 Household food consumption patterns

This section presents household food consumption patterns. The section starts by discussing reported rank of importance of different food categories to household food security. Thereafter, the section describes food expenditure patterns, food expenditure-income ratio and food expenditure shares. The aforementioned overall food preference analysis is essential as it provides useful insights on food preference.

4.3.1 Importance of different food categories to household food security

Household food preferences differ by household characteristics (Moscati, 2007). In a demand study featuring food commodities, it is essential to assess the importance of different food categories to household food security. This knowledge helps to understand food decisions. In the dataset for this study, respondents were asked to rank, on a scale of one to six (one-highest, six-lowest), the importance of different categories of food to their household food security. Table 4-6 presents the proportions of the sample that gave the highest rank to each food category. Cereals are the most important food category when it comes to household food security, reflecting their position as staples. Second from cereals are fruits and vegetables in the high and low-density areas, and animal products in the medium-density areas and in the pooled sample. Legumes are the third, fifth and fourth from cereals in the high, medium and low-density areas, respectively (Table 4-6). Overall, legumes occupy the mid position in household food security and this result concurs with findings of Kumar et al. (2011), Maganga et al. (2014) and Leserer (2010) who concluded that legumes occupy a mid-position amongst household food groups in terms of household food security.

Table 4.6 Proportion of households reporting importance of different food categories to household food security by location

Food category	Sample proportions by location (%)			
	High	Medium	Low	Pooled
Legumes	47.60	41.07	31.82	45.99
Fish and seafood	36.01	46.43	34.09	36.68
Fruits and vegetables	59.56	67.86	65.91	60.55
Cereals	83.79	76.79	81.82	82.97
Roots and tubers	22.56	35.71	9.09	22.71
Animal products	36.02	53.57	46.51	38.01
N	584	56	44	684

4.3.2 Food expenditure patterns

Food expenditure patterns help among others reveal household food preference[41] (Moschini, 1995) and it is essential for demand studies on food to understand the whole spectrum of household food preferences. The overall mean and median monthly food expenditures for the households in Lilongwe District were MK45,993.51 and MK35,000.00, respectively (Table 4-7). Based on the high standard deviation, there is considerable variability in food expenditure in the district. Food expenditure is positively skewed indicating that most households spend below the mean expenditure. The kurtosis (Table 4-7) of the expenditure distribution is 14.17 showing that more of the variance in expenditure is the result of infrequent extreme deviations, as opposed to frequent modestly sized deviations. Households in low-density areas spend a lot on food when compared to the other areas. This is evident by the higher mean and median monthly food expenditures (Table 4-7).

[41] The researcher perfectly recognizes that food expenditure may not entirely mean preference, but also availability and affordability.

Table 4.7 Monthly food expenditures by food category and location

Food category	***Monthly food expenditure (MK) by location***			
	High	Medium	Low	Pooled
Legumes	2609.90	3656.30	3996.59	2784.01
	(2000.00)	(2400.00)	(3000.00)	(2000.00)
Fish and seafood	5573.43	15080.36	24747.73	7576.42
	(4000.00)	(10000.00)	(15000.00)	(5000.00)
Fruits and vegetables	4155.40	10044.86	13471.59	5232.14
	(3000.00)	(7000.00)	(10000.00)	(3500.00)
Cereals	7900.72	16307.14	16058.18	9108.41
	(6000)	(13750.00)	(12500.00)	(7000.00)
Roots and tubers	2617.51	5491.96	5837.50	3050.05
	(2000.00)	(4900.00)	(5000.00)	(2000.00)
Animal products	8646.85	24975.00	30837.50	11399.06
	(5840)	(20000.00)	(24000.00)	(7000.00)
Other foods	5871.11	11826.61	13347.73	6835.42
	(4000.00)	(7000.00)	(10000.00)	(5000.00)
Total food expenditure	37374.92	87382.23	108296.82	45993.51
	(31600.00)	(73350.00)	(84700.00)	(35000.00)
N	584	56	44	684

Measures for spread and variation of household food expenditure

Standard deviation: 40349.22

Skewness: 2.77

Kurtosis:14.17

Note: Values in brackets are median values

The nominal expenditures presented in Table 4-7 are less informative from an Engel law perspective since they do not relate to household income. To comply with Engel law, this study additionally estimated the food-expenditure-income-ratio. Table 4-8 presents the food-expenditure-income ratio for households in all the location and the pooled sample. The overall mean and median food expenditure-income ratio are 0.51 and 0.54, respectively. Households in high-density areas have high mean and median food-expenditure-income ratios as compared to households in low-density areas. Since households in higher density areas are poorer than the low-density areas, the aforementioned result is consistent with Engel law. According to Varian (1987), Engel law states that as income rises, the proportion of income spent on food falls, even if actual expenditure on food rises. This negative relationship is also confirmed by the highly significant negative correlation coefficient (Table 4-8) of food expenditure-income ratio and household income.

Table 4.8 Food expenditure-income ratio by location

Density	**Food expenditure-income ratio**	
	Mean	Median
High	0.52	0.54
Medium	0.45	0.49
Low	0.48	0.47
Pooled	0.51	0.54

Correlation coefficient of food expenditure and income: -0.3678***

Note: Asterisks: *** means significant at 1%

The use of food expenditures presented in Table 4-7 to reveal household food preference is less intuitive unless accompanied by food expenditure shares. The latter is more informative since it relates expenditures on different foods to the overall food expenditure and/or household income. Table 4-9 presents the mean food expenditure shares calculated in relation to total household food expenditure.

Among all food groups, cereals have the highest expenditure shares (maintaining their position as dominant staples) seconded by animal products while roots and tubers have the lowest. Moving from high to low-density areas, expenditure shares for animal products, fruits and vegetables, and fish and seafood increase. On the other hand, the expenditure shares for cereals, legumes, roots and tubers, and other foods decrease as we move from high to low-location. In Malawi, animal products and fish and sea-foods are more expensive than cereals, legumes and root tubers. This is why the expenditure shares on animal products and fish and sea-food increases as we move from higher to lower location while the expenditure shares on cereals, legumes and root tubers decreases. These results are consistent with Bennet's law which according to Bopape (2006), states that households switch from less to more expensive calorie consumption as their incomes rise.

Table 4.9 Monthly food expenditure shares by food category and location

Food category	***Food expenditure shares by location***			**Pooled**
	High	Medium	Low	
Legumes	0.085	0.045	0.046	0.079
Fish, seafood	0.146	0.165	0.201	0.151
Fruits, vegetables	0.117	0.117	0.122	0.117
Cereals	0.225	0.200	0.159	0.220
Roots and tubers	0.072	0.066	0.060	0.070
Animal products	0.203	0.272	0.267	0.213
Other foods	0.151	0.137	0.144	0.149

4.4 Common bean consumption patterns

This section focuses on consumption of common beans. First, the section describes the purchase behaviour of consumers and sources of common beans. Thereafter, the section explains household expenditure on common beans; frequency of consumption, quantities consumed, typical roles of common beans in household food consumption, and pairing options of common beans. The last section describes factors that affect the purchase decision of common beans. This analysis is essential for understanding demand for common beans.

4.4.1 Common bean purchase

Understanding household purchase behavior of common beans is an important step towards understanding demand for common beans. Consumers first make choice/purchase decisions before allocating a budget to common beans (see the conceptual framework in section 3.2). Among the six common bean varieties analyzed, *Phalombe* variety was purchased by majority of the households across all the location (Table 4-10). Muthoni et al. (2008) and Chirwa and Phiri (2007) explain that *Phalombe* variety is highly preferred in the Central Region because of its short cooking time, good taste, color and emission of nice aroma during and after cooking. Overall, *Kabalabala* is the least purchased. Muthoni et al. (2008) explains that *Kabalabala* is mostly preferred in the Northern Region than in the other regions. The northerner loves its taste.

A comparison of zero purchase across the location shows that the high-density areas had the lowest (4.94%) zero purchase during the survey period. This was expected because common beans are regarded as a cheap source of protein as compared to meat (Larochelle et al., 2015) and the high-density areas contained households that are relatively poorer than the low-density areas hence a higher consumption of common beans in the high-density areas.

Table 4.10 Proportion of households by variety of common bean purchased and location

Common bean variety	Proportion purchased (%)			
	High	Medium	Low	Pooled
Kabalabala	4.64	8.93	11.36	5.68
Napilira	19.25	19.64	20.46	19.36
Phalombe	84.67	71.43	56.82	81.80
Khaki	24.02	23.21	34.09	24.60
Nanyati	22.15	30.36	22.73	22.85
Mixed beans	6.30	7.14	9.09	6.55
None	4.94	10.71	9.09	5.68
N	558	49	38	645

4.4.2 Common bean expenditure

Analysing expenditures on common beans is an important step to demand system estimation. Expenditure patterns among others reveal preferences which inform the demand elasticities (Yen and Lin, 2004). Table 4-11 shows the mean and median weekly expenditures on common beans by the households in Lilongwe. The median expenditures are zeros except for *Phalombe* beans reflecting high zero consumption of the other varieties analyzed during the survey period.

Phalombe has the highest mean weekly expenditure of MK2,632.70. High expenditure may be as a result of more purchase or high price. Even though Phalombe is not the cheapest (see Table 4-18), results in Table 4.14 shows it was purchased in largest quantities and this reflects a higher position in the preference spectrum. This concurs with Chirwa and Phiri (2007) and Muthoni et al. (2008) who argued that *Phalombe* and

Napilila variety are the most preferred and purchased variety in the central region of Malawi broadly due to visual, cooking and eating habits. There is no clear pattern when it comes to inter-comparison of the location in terms of magnitude of the expenditures. However, majority (83.33%) of the varieties have high expenditures in the high-density areas. This concurs with the conclusion, reported in section 4.3.2, that expenditure on legumes is highest in the high-density areas since legumes are a cheaper source of proteins as compared to meat.

Table 4.11 Weekly expenditures on different varieties of common bean by location

Variety	**Weekly expenditures (MK) by location**			
	High	Medium	Low	Pooled
Kabalabala	56.08	128.57	132.95	66.91
	(0.00)	(0.00)	(0.00)	(0.00)
Napilira	702.02	240.67	2396.23	772.92
	(0.00)	(0.00)	(0.00)	(0.00)
Phalombe	2632.70	1511.26	1701.82	2481.66
	(1400.00)	(1000.00)	(725.00)	(1400.00)
Khaki	354.35	300.60	706.86	372.55
	(0.00)	(0.00)	(0.00)	(0.00)
Nanyati	284.59	771.31	475.00	336.46
	(0.00)	(0.00)	(0.00)	(0.00)
Mixed beans	63.88	95.15	231.82	77.18
	(0.00)	(0.00)	(0.00)	(0.00)

Note: Values in brackets are median values; base year for prices is 2015

Table 4-12 presents budget shares calculated in relation to weekly expenditure on common beans. Overall, the results indicate that *Phalombe* has the highest budget share.

Khaki and *Nanyati* varieties rank second and third, respectively. The aforesaid varieties are among the four most preferred varieties[42] in the central region as discussed earlier.

Table 4.12 Budget shares of different common bean varieties by location

Common bean variety	Mean and median budget shares by location			
	High	Medium	Low	Pooled
Kabalabala	0.02	0.02	0.06	0.02
	(0.000)	(0.00)	(0.00)	(0.00)
Napilira	0.07	0.10	0.10	0.07
	(0.00)	(0.00)	(0.00)	(0.00)
Phalombe	0.64	0.48	0.39	0.61
	(0.69)	(0.50)	(0.33)	(0.667)
Khaki	0.11	0.10	0.16	0.11
	(0.00)	(0.00)	(0.00)	(0.00)
Nanyati	0.09	0.17	0.13	0.09
	(0.00)	(0.00)	(0.00)	(0.00)
Mixed beans	0.03	0.02	0.06	0.03
	(0.00)	(0.00)	(0.00)	(0.00)
Budget share of beans in household food expenditure	0.24 (0.07)	0.05 (0.03	0.49 (0.04)	0.21 (0.06)

Note: Values are in brackets are median values; base year for prices is 2015

Besides analyzing budget shares, it was essential to see how the budget shares correlate with household food expenditure. This reveals sensitivity of preference of common beans to changes in household welfare. Table 4-13 presents the correlations of the budget shares with total food expenditure. Significant correlations exist, showing both negative and positive correlations across the location and the pooled sample. A number

[42] Chirwa and Phiri also explain that consumers prefer these varieties mainly due to their attributes. In some locations these varieties are expensive yet highly preferred. This is also evident in Table 4-18 and Table 4.14 which shows that Phalombe variety is a bit expensive relatively yet purchased in largest quantities

of explanations can be made pertaining to the correlation between one type of food and the overall food expenditure. However, the most notable point is that since a budget share and food expenditure can be used as a proxies for preference and household welfare, respectively; positive and negative correlations between varietal budget shares and food expenditures indicate increase and decrease in preference, respectively, when household welfare improves. For instance, *Phalombe* variety has an overall negative correlation with food expenditure and this implies a reduction in its preference as household welfare improves. In other words, the *Phalombe* variety is highly preferred among low-income households. This concurs with the findings in Table 4-11 and Table 4-12 that the highest expenditure and budget share of the *Phalombe* variety was in the high-density areas.

Table 4.13 Correlation of budget shares of common beans and total food expenditure by location

Common bean variety	**Correlations of budget shares and total food expenditure by location**			
	High	Medium	Low	Pooled
Kabalabala	0.08*	0.22	-0.21	0.07*
Napilira	0.02	-0.05	0.48***	0.10***
Phalombe	0.00	-0.24*	0.06	-0.12***
Khaki	0.05	0.01	-0.06	0.04
Nanyati	-0.02	0.26**	-0.34**	0.03
Mixed beans	-0.0291	0.04	0.05	0.01

Note: Asterisks: *** means significant at 1%, ** means significant at 5%, and * means significant at 10%.

4.4.3 Quantities of common bean purchased

Quantity purchased of a product is an integral part of a demand function (as shown in section 3.3.1) hence the need to thoroughly understand this variable. Table 4-14 presents the quantities of common bean that were purchased weekly during the survey period. Quantity purchased per week was higher for *Phalombe* beans and lowest for Kabalabala and mixed beans implying to some degree that most consumers in the Lilongwe have a strong preference for *Phalombe* variety[43]. The mean purchased quantities for the variety are far much higher than the median quantities thereby revealing the presence of few extreme values for the survey. Table 4-14 also shows that the median values for most varieties are zero signifying the problem of zero purchase by majority of the sample during the survey period.

[43] Quantity purchased may also be affected by availability on the market. However, since prices of Phalombe were higher and the quantities purchased were highest, this reflects a high position in the preference spectrum

Table 4.14 Quantity purchased of different varieties of common bean by location

Variety	Quantity (Kg) purchased by location			
	High	Medium	Low	Pooled
Kabalabala	0.09	0.25	0.30	0.12
	(0.00)	(0.00)	(0.00)	(0.00)
Napilira	1.14	0.36	0.64	1.04
	(0.00)	(0.00)	(0.00)	(0.00)
Phalombe	8.49	2.13	2.41	7.58
	(2.00)	(1.50)	(1.00)	(2.00)
Khaki	1.08	0.44	1.07	1.03
	(0.00)	(0.00)	(0.00)	(0.00)
Nanyati	0.39	0.86	0.75	0.45
	(0.00)	(0.00)	(0.00)	(0.00)
Mixed beans	0.12	0.13	0.36	0.13
	(0.00)	(0.00)	(0.00)	(0.00)

Note: Values in brackets are median values

4.4.4 Frequency of consumption of common beans

When studying household demand for food products, it's essential to analyze the consumption frequency as this among others also informs about preference. Overall, the presence of zero consumption was highest (84.43% of households) for Kabalabala (Table 4-15). Muthoni et al. (2008) explains that this variety is highly preferred in the Northern Region unlike Phalombe, Nanyati and Napilira that are highly preferred in the Central Region. In each sample segment, *Phalombe* variety has the lowest zero consumption maintaining its position as the most preferred variety in the Central Region (Chirwa and Phiri 2007).

Table 4-16 presents proportion of households by frequency of consumption of different varieties of common bean and location. *Nanyati* and *Khaki* were the most frequently (more than once per week) consumed varieties of common bean in all the location. This supports Chirwa and Phiri (2007) and Muthoni et al. (2008) who concluded that these varieties are among the top four most preferred varieties in the Central Region of Malawi.

Table 4.15 Zero consumption of different varieties of common bean by location

Variety	**Sample proportion (%) of zero consumption by location**			
	High	Medium	Low	Pooled
Kabalabala	86.20	66.07	84.09	84.43
Napilira	61.84	42.86	75.00	61.14
Phalombe	7.76	12.5	25.00	9.17
Khaki	57.41	46.43	59.09	56.62
Nanyati	62.52	42.86	63.64	60.99
Mixed beans	83.99	78.57	84.09	83.55
N	584	56	44	645

Table 4.16 Proportion of households by frequency of consumption of different varieties of common bean and location

Variety	Proportion of households (%)			
	High	Medium	Low	Pooled
Kabalabala	6.30	17.86	4.55	7.13
	0.34	*1.79*	*6.82*	*0.14*
	0.17	**0.00**	**0.00**	**0.15**
	<u>1.87</u>	<u>5.36</u>	<u>0.00</u>	<u>1.02</u>
Napilira	20.78	26.79	2.27	20.09
	4.60	*10.71*	*0.00*	*4.80*
	3.92	**1.79**	**9.09**	**4.08**
	<u>4.26</u>	<u>8.93</u>	<u>9.09</u>	<u>4.95</u>
Phalombe	4.26	7.14	9.09	4.80
	6.64	*10.71*	*9.09*	*7.13*
	11.41	**3.57**	**6.82**	**10.48**
	<u>66.27</u>	<u>60.71</u>	<u>45.45</u>	<u>64.48</u>
Khaki	10.56	10.71	0.00	9.90
	9.71	*12.50*	*4.55*	*9.61*
	7.50	**5.36**	**11.36**	7.57
	<u>10.05</u>	<u>17.86</u>	<u>20.45</u>	<u>11.35</u>
Nanyati	9.03	12.50	2.27	8.88
	8.18	*3.57*	*6.82*	*7.71*
	4.94	**7.14**	**2.27**	**4.95**
	<u>10.90</u>	<u>26.79</u>	<u>20.45</u>	<u>12.81</u>
Mixed beans	5.79	1.79	2.27	5.24
	2.39	*1.79*	*2.27*	*2.33*
	0.85	**1.79**	**2.27**	**1.02**
	<u>2.21</u>	<u>7.14</u>	<u>4.55</u>	<u>2.77</u>
All beans	3.02	5.66	4.55	3.35
	3.91	*11.32*	*2.27*	*4.42*
	12.46	**1.89**	**15.91**	**11.89**
	<u>80.60</u>	<u>8.11</u>	<u>70.45</u>	<u>80.34</u>
N	558	49	38	645

Note: normal, italic, bold, and underlined values in a cell represent less than once per month, once per month, once every two weeks, and once or more per week respectively.

4.4.5 Pairing options of common beans

The preceding section looked at the consumption patterns of common beans. It is thus essential to look at the pairing options available for common beans in households in order to have a full picture of the preference spectrum associated with common beans. Table 4-17 presents the pairing options that were available for the sample. Majority (above 93%) of the households paired common beans with cereals. Cereals (especially maize) are dominant staples in Malawi. As such it is not surprising to see majority of the households pairing common beans with cereals. Second option for pairing is meat. Apart from beans serving as vegetables, a lot of households pair common beans and meats to meet their preferences and nutritional requirements. Akerele et al. (2013) and Kumar et al. (2011) reported similar findings in Nigeria and India, respectively. The proportion of the households that paired beans with either meat or fish is lowest in the high-density areas since most of these households can't afford to consume more than one source of protein per meal.

Table 4.17 Proportion of households by pairing options of beans and location

Pairing option	Proportion of household by location			
	High	Medium	Low	Pooled
Cereal (maize, rice, millet, and sorghum)	94.89	94.59	93.21	94.76
Cereal products (e.g. bread)	43.44	55.32	72.76	46.29
Plantains and bananas	15.16	14.23	36.38	16.45
Roots and tubers (cassava, potatoes, and yams)	27.08	35.70	45.44	43.52
Groundnuts	35.26	30.30	38.57	35.08
Meat (beef, chicken, and pork)	44.98	64.28	72.76	48.33
Fish (fresh, dried, and tinned)	27.94	42.81	43.25	30.13
N	558	49	38	645

4.5 Common bean prices

Prices are an ingredient in the budget share equations estimated through a demand system. Thus, an understanding of pricing patterns is essential as it informs both the estimation of a demand model and the price elasticities calculated from it. Table 4-18 presents the mean and median prices of the varieties analyzed. Overall, *Nanyati* has the highest prices seconded by *Khaki* variety. The inter-comparison of mean prices by location shows no clear-cut pricing pattern in terms of magnitude; each location density has a variety that is priced highest than the other location. Nanyati variety is still the most expensive variety in the high and medium-density areas while *Napilira* reigns in the low-density areas. The price of Napilira is unusually high in the low-density areas. This may be the result of value-addition and packaging. Overall, the cheap variety is mixed beans. Mixed bean is the cheapest since it contains a mixture of different varieties while consumers mostly prefer sorted and graded beans (Mkanda, 2007).

Table 4.18 Prices of different varieties of common beans by location

Variety	Price by location			
	High	Medium	Low	Pooled
Kabalabala	660.32	544.27	579.89	645.71
	(650)	(525.00)	(579.89)	(650.00)
Napilira	590.60	735.31	978.55	627.24
	(615.56)	(780.00)	(1217.63)	(630.00)
Phalombe	716.68	727.10	753.23	719.87
	(700.00)	(700.00)	(800)	(700.00)
Khaki	760.67	726.04	679.82	752.67
	(721.54)	(600.00)	(575.00)	(721.54)
Nanyati	768.28	842.05	611.93	764.28
	(792.86)	(718.58)	(650.00)	(792.86)
Mixed beans	547.61	742.86	619.32	568.12
	(537.50)	(742.86)	(619.32)	(547.61)

Note: Values in brackets are median values

4.6 Factors influencing quantity demanded of common beans

This section discusses factors that influence quantity demanded of different common beans. Non-consuming households (39% households of the sample) provided the general factors that deter them from purchasing the beans, and the factors that can persuade them to eat common beans.

4.6.1 Factors that deter purchase of common beans by non-consuming households

Table 4-19 shows that most non-consuming households did not buy common beans because of health reasons. Second from health reasons is preparation inconvenience. Beans take time to prepare and demand more cooking energy than other relish hence in cases where time is crucial or cooking energy is limited they bring a lot of preparation inconveniences.

Table 4.19 Proportion of non-consuming households by non-consuming factors and location

Factor	**Proportion of non-consuming households by location**			
	High	Medium	Low	Pooled
It is an inferior product	0.00	0.00	0.00	0.00
Health reasons	37.93	16.67	50.00	35.90
Religion/taboo	0.00	0.00	0.00	0.00
Expensive	3.44	0.00	0.00	2.56
Risk of social embarrassment	0.00	0.00	0.00	0.00
Preparation inconvenience	31.03	16.67	25.00	28.21
One household member does not like beans	0.00	16.67	0.00	2.56
Beans not considered as a meal option	10.34	0.00	0.00	7.69
Do not know how to prepare it	0.00	0.00	25.00	2.56
Do not believe beans are healthy	0.00	0.00	0.00	0.00
Do not like them	0.00	16.67	0.00	2.56
N	**26**	**7**	**6**	**39**

4.6.2 Factors that persuade people to eat common beans

Among the non-consuming households, majority in each density and the pooled sample valued good health benefits associated with consumption of beans as a factor that can persuade them to eat common beans (refer Table 4-20). Larochelle et al. (2015) explains that some of the health benefits associated with consumption of common beans include low salt and fat content, and zero cholesterol besides richness in protein and improved iron absorption. This explains why richness in protein and improved iron consumption also gained a lot of prominence (refer Table 4-20) among the factors that can persuade consumers to buy common beans. It is eminent to see that factors to do with health and nutrition gained a lot of prominence with the current increased awareness of ‘health eating’ at national and global levels. Fast cooking time also gained a lot of prominence and this was expected with the current scarcity and costly nature of cooking energy in Malawi (NSO Malawi, 2012).

Table 4.20 Factors that persuade consumers to buy common beans by location

Source of persuasion	**Proportion of non-consuming households by location**			
	High	Medium	Low	Pooled
High nutritive value	24.14	16.64	0.00	20.51
Rich in protein	27.58	16.64	50.01	28.21
Cheap source of proteins than animals	17.24	0.00	0.00	12.82
Faster cooking time	24.14	16.64	50.01	25.64
Good health benefits	34.48	16.64	50.01	33.33
Reduce risk of getting cancer	24.14	0.00	50.01	23.08
Enhance social status	17.24	16.64	24.96	17.95
No social embarrassment	24.14	33.35	50.01	28.21
Increase iron absorption	31.04	33.35	50.01	33.33
Improved pairing options in diets	13.80	0.00	0.00	10.26
If income increased by 10%	13.80	0.00	0.00	10.26
If income increased by 30%	6.90	0.00	0.00	5.13
From people they respect	6.90	0.00	0.00	5.13
N	**26**	**7**	**6**	**39**

4.7 Summary

The aim of this chapter was to provide a descriptive summary of different household characteristics and other variables that were used in the estimation of demand model presented in the subsequent chapter. Dwelling on the household characteristics, the average household size is 4.15 adult equivalent and households are composed mostly of adults. Majority of the household heads are married and employed. The mean monthly income for the household is MK109861 and the mean monthly food expenditure is MK45993. In terms of consumption of common beans, majority of the households consume common beans more than once per week and prefer the *Phalombe* variety. Common bean is consumed as main relish or as a vegetable especially when paired to meats. The factors that contribute to its purchase include its associated nutrition and health benefits. On the other hand, non-consumers don't purchase of common beans because of preparation inconvenience, preference, health reasons and expensiveness. In terms of pricing, prices of common beans average about MK700. The is no clear-cut pattern in magnitude of prices as we move from high to low-density areas, but the Khaki variety was the most expensive in the high and medium-density areas while Napilira was observed the highest priced in the low-density areas.

CHAPTER FIVE

HOUSEHOLD DEMAND FOR COMMON BEANS

5.1 Introduction

This chapter presents the results on the household demand for common beans in Lilongwe district of Malawi. The study objectives were: (1) to assess the sensitivity of budget shares for common beans to price and income effects; and (2) to assess the responsiveness of quantity demanded to price and income changes. To achieve the first objective, the study estimated a QuAIDS model which was estimated in a two-stage modelling framework in order to address the problem of zero consumption (refer Chapter three). The first stage estimated a multivariate probit model from which the pdfs and cdfs were calculated. The pdfs and cdfs fed in the second stage which estimated the QuAIDS model in order to address the problem of zero consumption. Results from all the stages are reported and discussed in the subsequent sections. For each stage, diagnostic tests are first presented before the model results. In order to achieve the second objective, this study calculated demand elasticities (see section 3.4.3) using the QuAIDS results. Overall, the study tested four hypotheses namely: (1) income of the household and prices affect the budget share of a common bean variety; (2) a unit percentage increase in own-price of a common bean variety decreases the quantity demanded by a less than unit percentage; (3) a unit percentage increase in price of a common bean variety increases the quantity demanded of other common bean varieties by more than a unit percentage; and (4) a unit percentage increase in income of the household increases the quantity demanded of a common bean variety by less than a unit percentage.

5.2 Multivariate probit model

5.2.1 Diagnostic tests for the multivariate probit model

This study estimated a multivariate probit model as the selection equation of the Shonkwiler and Yen (1999) two-stage demand system estimation. As discussed in chapter three, the purpose of estimating the multivariate probit was to predict the cdf and pdf. The cdf and pdf fed into the demand system as shown by Shonkwiler and Yen (1999) and later by a variety of studies (see section 2.4). Greene (2003), Gujarati (2004) and Maddala (1983) advise that special attention should be paid to multicollinearity, heteroskedasticity, and non-normality when estimating a probit regression. The Shonkwiler and Yen (1999) approach requires normality of residuals from the selection equation before estimating the outcome equation. All the above tests were done and the result are reported below.

5.2.1.1 Multicollinearity

Multicollinearity exists when two or more predictors are correlated in the model. When it occurs, it is difficult to separate the independent effect of each parameter estimate on the dependent variable and there is limited confidence in any results based on the estimates (Maddala, 1983). There are basically two types of multicollinearity namely structural[44] and data–based[45] multicollinearity. This study used the Variance Inflation Factors (VIF) to detect the overall multicollinearity. VIF is superior to other modes of testing multicollinearity (such as pairwise correlations[46]) because it shows both the presence and severity of multicollinearity in a simpler manner. Table 5-1 shows the

[44] Structural multicollinearity results from the equations we create i.e. a regression containing a predictor and its square.

[45] Data-based multicollinearity results from poor data which often is a result of poorly designed surveys and experiments.

[46] For more on measures of association refer Greene (2003), Gujarati (2004), and Maddala (1983).

results on the VIF. A mean VIF of 10 and above signals intolerable multicollinearity (Maddala, 1983). Based on the results in Table 5-1, this study concludes that there was no intolerable multicollinearity.

Table 5.1 Multicollinearity test of variables in multivariate probit model

Variable	VIF
Marital status	1.05
Location	
Low-density	1.89
Medium-density	1.85
Food expenditure	1.48
Household size	1.11
Prices	
Kabalabala	1.33
Napilira	1.56
Phalombe	1.17
Khaki	1.26
Nanyati	1.35
Mixed beans	1.51
Mean VIF	1.42

5.2.1.2 Heteroskedasticity

Heteroskedasticity (opposite of homoskedasticity) occurs when there is non-constant variance for any independent observations or explanatory variables. When it is present, it leads to inconsistent parameter estimates and wider confidence intervals (Gujarati, 2004). Due to lack of a post-estimation check for heteroskedasticity in Stata after multivariate probit model, this study estimated the robust-regression[47] form of the multivariate probit model. This self-cures heteroskedasticity and allows for heteroskedasticity-consistent standard errors.

5.2.1.3 Normality

Normality in the residuals is an inherent requirement for a probit regression (Greene, 2003; Klein and Spady 1993; Shonkwiler and Yen 1999). This study checked normality in the residual of the multivariate probit model using both graphical and statistical methods. The graphical and numeric method used were kernel-density plot and Shapiro-Wilk test of normality, respectively. All the methods employed showed a normal distribution of the residual from the multivariate probit model. Table 5-2 shows the results from the Shapiro-Wilk normality test (with null hypothesis of normality). The associated p-value from the hypothesis testing is not significant (leads to failure to reject the null hypothesis of 'there is no non-normality') showing that the multivariate probit model residual is normally distributed. Figure 5-1 shows the kernel-density plot of the multivariate probit model residual. The distribution is almost the same as that of its normal counterpart.

Table 5.2 Shapiro-Wilk normality test of multivariate probit model residual

Variable	**Z**	**P-value**
Mvprobit_Residual	1.204	0.1142

[47] Robust regression is a kind of regression estimated when there is belief of presence of heteroskedasticity and non-normality in the error terms. It involves\transforming the data to minimize the effect of extreme observations. Its generates robust standard errors and enables proper significance testing (Greene, 2003; Maddala, 1983)

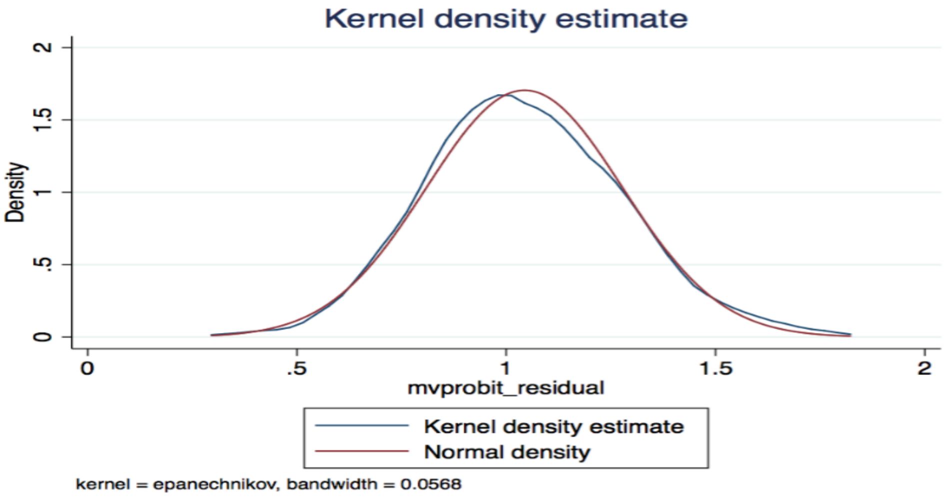

Figure 5.1 Stata's kernel-density plot of multivariate probit model residual

5.2.2 Results from the multivariate probit model

Food choice is a process that results from the competing, reinforcing and interacting influences of a variety of factors. The factors may range from the sensory, psychological and physiological responses of consumers to the interaction between social, environmental and economic influences (Moscati, 2007). As presented in section 3.4.2.1, the multivariate probit model was used as the selection equation of the Shonkwiler and Yen (1999) approach in order to obtain the cdf and pdf for the estimation of the outcome equation. The model included socioeconomic factors, location and prices, and featured six varieties of common beans.

An application of Lancaster (1966) utility theory clearly stipulates that the choice of common beans does not only depend on the aforementioned factors only, but also on bean attributes. However, due to data limitation, the attributes were not available. As such, this study only included prices as identifying variables in the model. The use of prices as identifying variables in the multivariate probit model is not a very weak procedure considering that prices are also impacted by the attributes (Muthoni et al., 2008) and thus reflect on the attributes.

Besides data limitation problems, a multivariate probit model cannot handle specific bean attributes in modelling choice behaviour because of its theoretical background (refer to Maddala, 1983), yet it is the only parametric selection equation that can be used to calculate the cdf and pdf in the Shonkwiler and Yen (1999) approach[48]. Thus if inferences on choice of beans were to be made, this study would (beside the Shonkwiler and Yen (1999) approach) have considered models such as mixed logit model[49] that can properly handle specific product attributes. Consequently, this study estimated the multivariate probit for cdf and pdf calculations only.

Table 5-3 presents result from the multivariate probit model. It is evident from the results that the choice of common beans is affected by socioeconomic factors, location, and own and cross prices besides specific bean attributes (as stipulated by literature) that were not included in the model. The constant terms are larger than any other terms in the model signalling that the model has left out important factors. This is so because critical bean attributes were not included in the model due to data limitation and the inherent weakness[50] of the multivariate probit model. Even though no inferences are made on the results from the multivariate model, the significance of the factors concur

[48] This is the sole reason why when normality in the residuals of the multivariate probit model fails to hold, semiparametric and non-parametric procedures are considered.
[49] Mixed logit model is built on foundations of Lancaster (1966) utility framework which notices that choice of a product such as a variety of common bean emanates from utility which results from an interaction of variety of factors that include consumer and product attributes.
[50] The micro foundation of multivariate probit model is not based on Lancaster (1963) utility theory which is the sole ingredient if choice of common beans is to be modelled and proper inference be made.

with Akerele et al. (2013) and Temitrope and Haruna (2013) and paves way for the subsequent demand system estimation by showing that demand of beans is sensitive to prices.

Table 5.3 Results from the multivariate probit model

	Choice					
	Kabalabala (1)	Napilira (2)	Phalombe (3)	Khaki (4)	Nanyati (5)	Mixed (6)
Constant	7.5369	13.0633***	3.6634	9.5565***	5.0390	8.8673*
	(4.9225)	(3.6634)	(3.4053)	(3.6759)	(3.9239)	(4.8092)
MAR	0.3648*	-0.1712	0.1958	0.1113	0.2564*	0.2851
	(0.2040)	(0.1238)	(0.1302)	(0.1258)	(0.1315)	(0.1921
HH	0.1023	0.1991	0.3964***	0.0649	0.3230***	0.2666
	(0.1747)	(0.1412)	(0.1428)	(0.2152)	(0.1218)	(0.1692)
$Ln(EXP_{food)}$	0.1587	0.0379	0.0059	0.1135	0.0401	0.0214
	(0.1332)	(0.0894)	(0.1000	(0.0892)	(0.0812)	(0.1096)
LD	0.1573	0.3312	-0.8289***	0.0629	-0.4101	0.4746
	(0.3610)	(0.3196)	(0.2874)	(0.3055)	(0.3507)	(0.3818)
MD	0.0023	0.1137	-0.5540**	-0.1847	0.2177	0.2609
	(0.3423)	(0.2783)	(0.2638)	(0.2786)	(0.2928)	(0.3882)
$Ln (P_1)$	-0.9227	-0.7897**	-0.0598	-0.5066	-0.1852	-0.5858
	(0.6763)	(0.3935)	(0.3902)	(0.3841)	(0.3743)	(0.5505)
$Ln (P_2)$	-0.0288	-0.8423**	-0.0193	0.1458	-0.0779	-0.7126*
	(0.3121)	(0.3337)	(0.2388)	(0.2702)	(0.2642)	(0.2952)
$Ln (P_3)$	-0.4222*	-0.2250	-1.2073***	0.4364**	0.0563	-0.3055
	(0.2306)	(0.1713)	(0.1761)	(0.1858)	(0.1761)	(0.2213)
$Ln (P_4)$	-0.4763**	-0.0644	0.4040*	-1.7674***	0.2934	-0.0455
	(0.2301)	(0.2368)	(0.2158)	(0.2760)	(0.2398)	(0.2630)
$Ln (P_5)$	0.1154	-0.1892	0.0457	0.6989**	-1.6552***	0.1458
	(0.3084)	(0.2618)	(0.3075)	(0.2777)	(0.3806)	(0.3627)
$Ln (P_6)$	-0.0029	-0.8435	0.3345	-0.8100**	0.54764	-0.2433)
	(0.4290)	(0.3817)	(0.3845)	(0.3627)	(0.4795)	(0.6020)
n=684						

Note: Values in brackets are standard errors MAR, marital status; EXP_{food} , food expenditure; HH, household size; LD, low-density; MD, medium-density; CST, constant; *** means significant at 1%, ** means significant at 5%, * means significant at 10%.

5.2.3 Calculation of cdf and pdf from the multivariate probit model

The Shonkwiler and Yen (1999) approach requires that the cdf and pdf should be obtained from the selection equation and used in the estimation of the outcome equation which is the second stage of the approach. As shown in section 3.4.2.1, the cdf multiplies all the right-hand variables in the outcome equation while the pdf is treated as an added regressor. This helps to correct for the problem of zero expenditures on some products analysed during the survey period. Duly, the cdf and pdf were predicted from the multivariate model. Thereafter, the cdf was multiplied to all the regressors in the demand system estimation while the pdf was treated as an added regressor.

5.3 Complete demand system estimation

As shown in chapter three, both the QuAIDS and LA/AIDS were estimated. Thereafter, one model was chosen among the two to represent consumer demand for common beans in Lilongwe District. The choice of the model was based on diagnostic tests. The diagnostic tests involved testing the approach, the data, and the performance of the models. The subsections below present the results from the diagnostic tests, model output and the associated demand elasticities.

5.3.1 Diagnostic tests for the QuAIDS and the LA/AIDS model

5.3.1.1 Approach-based diagnostic tests

The Shonkwiler and Yen (1999) two-stage approach has two critical assumptions namely normality of the residual from the selection equation and the correlation of the errors from the selection and outcome equations. Normality of the errors from the multivariate probit model was tested in section 5.2.1.3.and the results rejected the null hypothesis of no normality. Table 5-4 presents the correlation test (with null hypothesis of no correlation) of the residual from the multivariate probit model and the QuAIDS and LA/AIDS independently. The test fails to reject the null hypothesis of no correlation among the aforementioned residuals for both models. Thus the two-stage modelling framework was deemed valid for the analysis. On a further note, this study made efforts to test for significance of the pdfs in the QuAIDS and LA/AIDS through the Wald tests as this confirms the validity of the Shonkwiler and Yen (1999) two-stage approach. Table 5-5 presents the results of the test in each equation of each model and the pooled sample. The results indicate overall significance of the pdfs in the equation, implying the importance of the Shonkwiler and Yen (1999) two-stage approach to the estimation of the demand systems.

Table 5.4 Multivariate probit and demand system error correlation

	Error correlations	
	Multivariate probit and LA/AIDS	Multivariate probit and QuAIDS
Correlation coefficient	0.0386	-0.0102
P-value	0.3120	0.7993

Table 5.5 Testing for significance of pdfs in the QuAIDS and LA/AIDS

Equation	χ^2	
	LA/AIDS	QuAIDS
KB	3.61*	5.60**
NP	0.03	2.69
PH	5.74**	90.85***
KH	0.52	1.12
NY	6.40**	9.64***
MB	3.78	147.99***
Overall	24.91***	237.72***

Note: KB, *Kabalabala*; NP, *Napilira*; PH, *Phalombe*; KH, *Khaki*; NY, *Nanyati*; MB, *mixed beans*; ***, significant at 1%; **, significant at 5%; *, significant at 10%.

5.3.1.2 Data-based diagnostic tests

Test for multicollinearity

As discussed in section 5.2.1.1, multicollinearity is a serious issue that affects the accuracy of results in a regression framework. Duly, this study made efforts to test for multicollinearity of the equations included in both the QuAIDS and the LA/AIDS. VIFs were used to test for multicollinearity and Table 5-6 presents the results. The estimated equations in each model used different sets of cdf and pdf thereby making the predictors to differ across the equations as illustrated in section 3.4.2.1. This being the case, mean VIFs are reported for each equation in each model. All the mean VIFs are far much less than 10, making the study fail to reject the hypothesis of no intolerable multicollinearity for both models.

Table 5.6 Multicollinearity test of variables in LA/AIDS and QuAIDS

Equation	***Mean VIF***	
	LA/AIDS	QuAIDS
KB	1.30	1.27
NP	1.54	1.48
PH	1.52	1.47
KH	2.37	2.22
NY	1.60	1.54
MB	1.26	1.23

Note: KB, *Kabalabala*; NP, *Napilira*; PH, *Phalombe*; KH, *Khaki*; NY, *Nanyati*; MB, *mixed beans*

Test of restrictions

The AIDS model gives a researcher the possibility to impose and/or test the validity of underlying consumer demand theory in economics (Deaton and Muellbauer, 1980). Thus, the underlying theory of the AIDS can be tested if it is supported by the data at hand. Results from such tests do not affect the model estimation and validity, but rather it is worth noting if the data supported the validity of the restrictions imposed. In this regard, this study tested the symmetry and homogeneity restrictions to see if they were supported by the data. Following Abdulai (2002), Alviola (2010), and Leserer (2010); this study tested the restrictions using the likelihood ratio test[51]. Three types of restriction tests were performed namely homogeneity[52], symmetry[53], and homogeneity and symmetry all together. Table 5-7 shows that the calculated chi-square statistics are

[51] In likelihood ratio test: based on the restriction at hand, the unrestricted and restricted models are estimated and the maximum log-likelihoods L_0 and L_1 obtained. The ratio L_0 / L_1 is calculated and it follows a chi-square distribution. χ^2=-2Log (L_0 / L_1) is compared to the tabulated χ^2 with k-degrees of freedom

[52] The null hypothesis of homogeneity test is that the data supports the homogeneity restriction

[53] The null hypothesis of symmetry test is that the data supports the symmetry restriction

highly significant implying that in all cases the data rejected the validity of the restrictions imposed.

Table 5.7 Testing restrictions of demand theory

Restriction	QuAIDS		LA/AIDS	
	χ^2	Verdict on null hypothesis	χ^2	Verdict on null hypothesis
Homogeneity	34.97***	Rejected	35.50***	Rejected
Symmetry	20.14***	Rejected	15.64***	Rejected
Homogeneity and symmetry	53.26***	Rejected	51.97***	Rejected

Note: Asterisks; *** means significant at 1%, ** means significant at 5%, * means significant at 10%.

5.3.1.3 Model-based diagnostic tests

Test of specification

Misspecification of a model comes due to the functional form used or due to the effect of omitted variables. When estimating the AIDS model, it is very essential to check for misspecification. In chapter three, we showed that the most competing variants of the AIDS model for this study are the QuAIDS and LA/AIDS models. For precision's sake, both models were estimated and specification tests based on functional form[54] and effect of omitted variables[55] were performed. Table 5-8 and Table 5-9 present the results. The overall result in Table 5-8 rejects the hypothesis that there is no need for

[54] The null hypothesis of the functional form test is that there is no need for the quadratic expenditure term

[55] The null hypothesis of the specification test based on effect of omitted variables is that there is no effect of the omitted variables

the quadratic expenditure term in the demand system estimation. Hence QuAIDS is superior to LA/AIDS. However, for both models as shown in Table 5-9, the effect of omitted variables is significant across majority (83.33%) of the equations estimated suggesting that both could have been improved by adding more variables.

Table 5.8 Specification test based on functional form

Equation	χ^2	**Verdict on null hypothesis**
KB	0.22	Accepted
NP	0.00	Accepted
PH	42.80***	Rejected
KH	0.00	Accepted
NY	0.12	Accepted
MB	108.57***	Rejected
Overall	121.80***	Rejected

Note: KB, *Kabalabala*; NP, *Napilira*; PH, *Phalombe*; KH, *Khaki*; NY, *Nanyati*; MB, *mixed beans*; ***, significant at 1%; **, significant at 5%; *, significant at 10%.

Table 5.9 Specification test based on the effect of omitted variables

Equation	**H_0: Model has no omitted variables**			
	QuAIDS		LA/AIDS	
	F	Verdict	F	Verdict
KB	17.74***	Rejected	17.90***	Rejected
NP	7.88***	Rejected	7.96***	Rejected
PH	5.95***	Rejected	6.03***	Rejected
KH	3.54**	Rejected	3.54**	Rejected
NY	5.85***	Rejected	5.83***	Rejected
MB	0.85	Accepted	0.50	Accepted

Note: KB, *Kabalabala*; NP, *Napilira*; PH, *Phalombe*; KH, *Khaki*; NY, *Nanyati*; MB, *mixed beans*; ***, significant at 1%; **, significant at 5%; *, significant at 10%.

Test of heteroscedasticity

The study estimated the QuAIDS model using robust regression framework (for more on robust regression see section 5.2.1.2). This self-heals heteroskedasticity and allow for estimation of heteroskedastic-consistent standard errors. However, for the LA/AIDS, STATA's SUR framework does not have an option for robust regression. Hence, efforts were made to test for heteroskedasticity[56] after the LA/AIDS. Table 5-10 presents result from the test. The study tested for heteroskedasticity in each equation and the results indicate presence of heteroskedasticity in the LA/AIDS model and this necessitating for correction for heteroskedasticity if inferences are to be made from the model.

Table 5.10 Test of heteroskedasticity

Equation	χ^2
KB	532.06***
NP	71.34***
PH	4.15**
KH	64.41***
NY	38.06***
MB	0.90

Note: KB, *Kabalabala*; NP, *Napilira*; PH, *Phalombe*; KH, *Khaki*; NY, *Nanyati*; MB, *mixed beans*; ***, significant at 1%; **, significant at 5%; *, significant at 10%.

Test of expenditure endogeneity

Expenditure endogeneity occurs when the expenditure variable in the budget share equation is correlated with the equation errors. Disregarding expenditure endogeneity

[56]The null hypothesis of the test of heteroscedasticity is that the is no heteroskedasticity

and treating expenditure as exogenous in demand models often results in inconsistent and biased estimates (Agostini, 2014). Hence a formal test of expenditure endogeneity[57] is important to any demand system estimation. This study adopted the augmented regression approach (Blundell and Robin, 1999) to test for expenditure endogeneity. In the first step, the study regressed all the cdf-multiplied price variables, the pdfs, and income and its square on the expenditure variable for each product category. Residuals were predicted from each equation and included in the budget share equations. The significance of the residual (which is test for endogeneity) for each equation was tested using the Wald test. The overall significance of the residuals was also tested using the Wald test. Table 5-11 reports the results for both the QuAIDS and LA/AIDS models. The overall result suggests no expenditure endogeneity for both models.

Table 5.11 Test of expenditure endogeneity

Equation	**QuAIDS**		**LA/AIDS**	
	χ^2	Verdict on null hypothesis	χ^2	Verdict on null hypothesis
KB	1.42	Accepted	3.75*	Rejected
NP	0.16	Accepted	0.10	Accepted
PH	0.27	Accepted	0.06	Accepted
KH	4.71*	Rejected	1.38	Accepted
NY	0.02	Accepted	0.20	Accepted
MB	0.19	Accepted	0.05	Accepted
Overall	0.47	Accepted	6.87	Accepted

Note: KB, *Kabalabala*; NP, *Napilira*; PH, *Phalombe*; KH, *Khaki*; NY, *Nanyati*; MB, *mixed beans*; ***, significant at 1%; **, significant at 5%; *, significant at 10%.

[57] The null hypothesis of expenditure endogeneity is that there is no expenditure endogeneity

5.3.2 Choice of demand model

Based on the model performance diagnostic tests namely specification tests, heteroscedasticity, restriction test, and expenditure endogeneity, this model chose the QuAIDS over the LA/AIDS as the best representation of consumer demand for beans. The specification tests supported the assertion that the QuAIDS is superior to the LA/AIDS model. Based on the heteroskedasticity test, the LA/AIDS suffered from heteroscedasticity. The restriction tests were not in favor of theory of any of the models. However, since restriction tests only help to test if the data supports the underlying demand theory and not reject/accept a demand model, both the QuAIDS and LA/AIDS can still be estimated. The expenditure endogeneity tests for both models supported the null hypothesis of no expenditure endogeneity for both models. Based on the aforesaid tests, this study chose the QuAIDS over the LA/AIDS as a representation of consumer demand. Table 5-12 summarizes the model-based diagnostic tests performed and the verdict on the choice between the QuAIDS and the LA/AIDS.

Table 5.12 Choice between LA/AIDS and QuAIDS

Test	**Results based on model**	
	QuAIDS	LA/AIDS
Specification	Valid	Not valid
Restriction	Not valid	Not valid
Heteroscedasticity	Valid	Not valid
Expenditure endogeneity	Valid	Valid
Verdict	Valid	Not valid

5.3.3 Results from the QuAIDS model

Results from the QuAIDS model are reported in Table 5-13. The model included five budget share equations. The estimates of the sixth budget share recovered as explained in section 3.4.2.1. For all the estimated equations, it is consistently observed that a small proportion of the price and income (proxied by expenditure) effects are significant. This implies little sensitivity of the budget shares of different types of common beans to price and income changes and suggests that there is more to household budget allocation of beans than price and income changes only. In compliment, Muthoni *et al*. (2008) noted that only a small variation in expenditure on beans is explained by prices and income because demand is largely affected by product attributes. On another note, Khaliukova (2014) explains that it is a common phenomenon and not surprising to find a small proportion of significant price and income effects in demand analysis involving cross-sectional data.

Table 5.13 IFGNLS estimates of the QuAIDS parameters

	Expenditure share					
	Kabalabala (1)	Napilira (2)	Phalombe (3)	Khaki (4)	Nanyati (5)	Mixed (6)
Constant	-0.8191**	-0.2040**	-1.6410***	0.1834	-0.5191	3.9999***
	(0.3765)	(0.0983)	(0.5788)	(0.3133)	(0.3374)	(0.5529)
ln_{p1}	-0.0323					
	(0.0404)					
ln_{p2}	0.0354	-0.0243				
	(0.0436)	(0.0774)				
ln_{p3}	-0.0302	-0.0944	-0.0729			
	(0.0391)	(0.0947)	(0.3636)			
ln_{p4}	0.0041	0.0625	0.0870	-0.2532***		
	(0.0285)	(0.0509)	(0.1058)	(0.0966)		
ln_{p5}	0.0160	-0.0482	-0.0614	0.1596***	-0.0688	
	(0.0385)	(0.0511)	(0.0854)	(0.0557)	(0.1396)	
ln_{p6}	0.0071	0.0690	0.1719	-0.0599	0.0029	-0.1909
	(0.0413)	(0.0860)	(0.2172)	(0.0972)	(0.1475)	(2409)
ln_{exp}	-0.0017	0.0199	-0.3661***	0.0476	0.0209	0.2794***
	(0.0156)	(0.0458)	(0.1117)	(0.0431)	(0.0370)	(0.0744)
ln_{exp^2}	0.0005	0.0001	0.0484***	0.0001	-0.0009	0.0497***
	(0.0010)	(0.0032)	(0.0074)	(0.0030)	(0.0025)	(0.0048)
PDF	-1.0405***	-0.5024	-3.0403***	-0.2767	0.8931***	-5.7530***
	(0.4398)	(0.3066)	(0.3190)	(0.2611)	(0.2877)	(0.4729)
n=680						

Note: Values in brackets are standard errors corrected for heteroskedasticity using robust regression; *** means significant at 1%, ** means significant at 5%, * means significant at 10%

The presence of significant price and expenditure effects in the QuAIDS model makes us fail to reject the overall hypothesis that income and price changes affect the budgets allocated to different types of common beans, and further indicate the importance of considering income and prices in demand analysis as suggested by neoclassical micro theory of consumer demand. However, the interpretation of price and income effects is best discussed in terms of elasticities (Aepli 2014; Bopape, 2006; Taniguchi and Chern, 2000). As shown in chapter two, demand elasticities show the proportionate change that arises in quantity demanded of a product as a result of a unit percentage change in the analysed variable. Table 5-14 presents the calculated demand elasticities from the QuAIDS model using equations 22 to 30.

5.3.4 Elasticities calculated from the QuAIDS result

The elasticities[58] calculated from the QuAIDS results include expenditure, income, price (both Marshallian and Hicksian) elasticities. Income elasticities were calculated from expenditure elasticities using Chern et al. (2003) formula shown in section 3.4.3. A comparison of expenditure and income elasticities shows substantial differences, suggesting that demand studies especially for more disaggregated commodities should shun away from using expenditure elasticities as proxies for income elasticities, as this can be misleading for inferences. The comparison of Marshallian and Hicksian price elasticities shows noticeable differences in signs and magnitude, signifying importance of expenditure effects in consumer demand decisions (Aepli 2014; Abdulai, 2002). However, both need to be estimated since Marshallian price elasticities are excellent at own-price (give own-price net of utility) effects while Hicksian price elasticities give a more accurate picture of cross-price substitution between commodities, since they are a measure of substitution effects net of income (Abdulai, 2012: Nicholson and Snyder, 2008)

[58] All the elasticities were approximately normally distributed and thus were calculated at their mean values. Quantile-quantile, Kernel density, and histogram plots were employed to check the distributions of the elasticities. Results on the distributions of all the elasticities have not been included but are available upon request. If the elasticities were skewed the mean would not provide the best measure of central tendency and the median would have been used in such case. The standard errors and p-values for the elasticities were calculated using the delta method.

It can be observed that all the income elasticities are positive except for mixed beans and all range from 0.0005 to 0.0090 (Table 5-14). This indicates that except for mixed beans, all the varieties of common beans analyzed are normal and necessity goods, consumption of which will increase with rising income. This result concurs with demand theory and also concurs Akerele *et al.* (2013) and Gonzalez and Wieck (2014) who concluded that beans are normal and necessity goods[59]. Their argument based on the elasticities was that beans are part and parcel of most traditional meals across the world, they act as cheap source of protein as compared to meat hence affordable to many and are also consumed as vegetables. Contrary to expectation; mixed beans have negative income elasticity and this suggests that mixed beans are inferior goods, of which consumption decreases with rising incomes. However, this result is not significant, but probably it may be because mixed beans have the lowest price (see section 4-5), and by Bennet's law as explained by Bopape (2006), households switch from less to more expensive calorie consumption as their incomes rise.

To provide a picture of demand for common beans due to prices, Table 5-14 presents the Marshallian and Hicksian price elasticities. The Marshallian own-price elasticities are larger in absolute values than the Hicksian. This is because the formulation of Marshallian price demand function allows for change in purchasing power, unlike the Hicksian demand function for which purchasing power is maintained (Nicholson and Snyder, 2008). Both the Marshallian and Hicksian own price elasticities are negative for all the varieties of common beans analyzed except for mixed beans. Negative own-price elasticities conform to the law of demand which states that demand for a

[59] It has to be noted that results in table 4.14 may seem contradictory. This is because figures in Table 4.14 are nominal and don't control for price effects unlike elasticities which are real.

commodity increases with a decrease in prices. Mixed beans have a positive own-price elasticity and this is contrary to literature and violates the law of demand. The result suggests that mixed beans belong to the category of Giffen goods. Quantity demanded of a good depends on both income and substitution effects and Varian (1987) explains that the case of a Giffen good is when a good has a negative income elasticity (more demand at low income) and this income effect is so large than the price effect (high price, less purchase of a good, more purchase of its substitute), causing the net effect of increase in demand due to rise in price. Thus the negative income elasticity found on mixed beans is in line with the finding that mixed beans belong to Giffen goods.

Through a comparison of Marshallian own-price elasticity estimates, we can also see that the *Napilira* variety has a closer to unity own-price elasticity in absolute values than the other varieties. Since demand theory (Moscati, 2007) recognizes that the optimal own-price elasticity for any commodity is unity; therefore, our results suggest that the own-price elasticity of the *Napilira* variety is more optimal than those for the other varieties analyzed, and this essentially implies higher revenue generation capacity. This result is not surprising especially with the finding that this variety has high prices in the medium and low-density areas (see section 4-5). Chirwa and Phiri (2007) also noted that it is among the top three most preferred varieties of common bean in Malawi.

Table 5.14 Demand elasticity estimates for the QuAIDS model

	Price						Expenditure	Income
	Kabalabala	*Napilila*	*Phalombe*	*Khaki*	*Nanyati*	Mixed beans		
Marshallian								
Kabalabala	-1.0386***	0.0656***	-0.0400***	0.0028***	0.0336***	-0.0038***	1.0151***	0.0089***
	(0.0003)	(0.0005)	(0.0002)	(0.0000)	(0.0003)	(0.0001)	(0.0001)	(0.0004)
Napilila	0.0782***	-1.0132***	-0.0520***	0.0711***	-0.0504***	0.0263***	1.0126***	0.0089***
	(0.0013)	(0.0002)	(0.0009)	(0.0013)	(0.0009)	(0.0005)	(0.0001)	(0.0004)
Phalombe	0.1606***	0.1305***	-1.0424***	-0.0105***	0.1180***	-0.4967***	1.2445***	0.0115***
	(0.0039)	(0.0031)	(0.0030)	(0.0009)	(0.0028)	(0.0118)	(0.0080)	(0.0008)
Khaki	0.0445***	0.0946***	0.2213***	-1.2860***	0.1873***	-0.1749***	1.0149***	0.0089***
	(0.0006)	(0.0013)	(0.0031)	(0.0041)	(0.0027)	(0.0025)	(0.0001)	(0.0004)
Nanyati	0.0524***	-0.0231***	0.0143***	0.1547***	-1.0519***	-0.0968***	1.0265***	0.0090***
	(0.0014)	(0.0007)	(0.0004)	(0.0042)	(0.0015)	(0.0026)	(0.0004)	(0.0004)
Mixed beans	-0.8369***	-0.3031***	0.4803***	-1.5715***	-0.9215***	1.2048***	-0.0312	-0.0005
	(0.0143)	(0.0062)	(0.0196)	(0.0289)	(0. 173)	(0.0368)	(0.0199)	(0.0002)
Hicksian								
Kabalabala	-1.0182***	0.0860***	-0.0195***	0.0231	0.0540***	0.0165		
	(0.0038)	(0.0037)	(0.0038)	(0.0037)	(0.0037)	(0.0037)		
Napilila	0.1506***	-0.9408***	0.0204***	0.1435***	0.0220***	0.0987***		
	(0.0070)	(0.0072)	(0.0074)	(0.0071)	(0.0073)	(0.0071)		
Phalombe	0.8926***	0.8924***	-0.3104***	0.7214	0.8500***	0.2352***		
	(0.0154)	(0.0155)	(0.0167)	(0.0159)	(0.0155)	(0.0221)		
Khaki	0.1555***	0.2057***	0.3324***	-1.1749***	0.2984***	-0.0639***		
	(0.0088)	(0.0087)	(0.0088)	(0.106)	(0.0088)	(0.0098)		
Nanyati	0.1525***	0.0770***	0.1144***	0.2548***	-0.9518***	0.0035***		
	(0.0084)	(0.0085)	(0.0084)	(0.0090)	(0.0087)	(0.0090)		
Mixed beans	-0.8305***	-0.2966***	0.4871*	-1.5651***	-0.9150***	1.2113***		
	(0.0163)	(0.0083)	(0.0188)	(0.0307)	(0.0351)	(0.0351)		

Note: *Values in brackets are standard errors, *** significant at 1%, ** significant at 5%, * significant at 10%*

Pertaining to cross-varietal relationships, results on the Marshallian and Hicksian cross-price elasticities provide a mixture of complementary and competitive varietal relationships (Table 5-14). There is no clear difference in magnitude and signs between the Marshallian and Hicksian cross-price elasticities. All the cross-price elasticities are inelastic and accordingly imply little sensitivity of demand for a variety due to cross price effects. It was difficult to compare these results with other findings since no study amongst the reviewed papers has studied beans at a more disaggregated level. However, Agostini (2014), Khaliukova (2014) and Taylor (2014) specify that competitive relationships are expected for varieties of the same commodity due to differences in prices and other attributes which define the preference spectrum of consumers. In addition, expenditure on one variety may mean no expenditure on the others. Thus the complementary relationships are somehow odd. A possible explanation to this phenomenon according to Katungi et al. (2009) and Mkanda (2007) is that households may prefer/buy two or even more varieties during a designated time period. Digging dip, it is possible for the preference spectrum of a consumer to be spread across two or more varieties, eventually making consumers purchase more than one variety (Katungi et al., 2009).

In a demand study, it is possible to compare cross price elasticities in order to see the strongest complementary/competing relationships among the products analysed. In a more disaggregated study, for instance studying different varieties of a crop, this knowledge is essential for different players and processes in the crop’s value chain. Using estimated results on cross-price elasticities in this study, we can compare the competitiveness among the varieties and come up with the most competing varieties of common beans. The highest positive cross price elasticity signifies the highest competition/substitution power. Thus according to the Hicksian cross-price elasticities presented in Table 5-14, the greatest competition occurs between the *Phalombe* and *Kabalabala* varieties, seconded by *Phalombe* and *Napilira*. In a nutshell, the *Kabalabala* and *Napilira* varieties are the two greatest competitors of the Phalombe variety, and can substitute the *Phalombe* variety at a rate higher than any other substitution among the varieties analysed. Muthoni et al. (2008) explains that in the Central Region of Malawi, *Kabalabala*, *Napilira*, *Nanyati*, and *Phalombe* are the most preferred varieties and often have the largest market shares. Thus, it is not surprising to see (Table 5-14) that *Kabalabala*, *Napilira*, and *Nanyati* are the three greatest competitors of the *Phalombe* variety, and can substitute it at a rate higher than any other substitution between two varieties among the varieties analysed.

5.4 Results on hypothesis testing

Four hypothesis were tested in this study. Table 5-1 summarizes the results on hypothesis testing. The results led to the rejection of the second and third hypotheses which, respectively stipulated that a unit percentage increase in own-price of a common bean variety decreases the quantity demanded by a less than unit percentage and a unit percentage increase in price of a common bean variety increases the quantity demanded of other common bean varieties. This is because all the own-price elasticities were elastic and cross-varietal relationships were both competitive and complementary. On the other hand, the study failed to reject the first and fourth hypotheses which stated that income of the household and prices affect the budget share of a common bean variety and a unit percentage increase in income of the household increases the quantity demanded of a common bean variety by less than a unit percentage. This is because price and income effects were significant in the QuAIDS and all the income elasticities were positive and below unity.

Table 5.15 Hypothesis testing

Hypothesis	Accepted/Rejected
1. Income of the household and prices affect the budget share of a common bean variety	Accepted
2. A unit percentage increase in own-price of a common bean variety decreases the quantity demanded by a less than unit percentage	Rejected
3. A unit percentage increase in price of a common bean variety increases the quantity demanded of other common bean varieties	Rejected
4. A unit percentage increase in income of the household increases the quantity demanded of a common bean variety by less than a unit percentage	Accepted

5.5 Concluding remarks

Each model estimated in this study was estimated with great attention to precision, assumptions and known estimation issues. The study methodology followed a two-stage approach in which multivariate probit model and QuAIDS were estimated. Diagnostic tests were conducted for each model to ensure correct inference making. The multivariate probit model was estimated as the selection equation and first stage model of the two-stage approach. Thereafter the study predicted the cdf and pdf from the multivariate probit results. These fed into the QuAIDS model which was estimated as the outcome equation and second stage model. In the second stage, the study estimated both the QuAIDS and the LA/AIDS and based on diagnostic tests; the QuAIDS was chosen as the best representation of consumer demand for common beans.

Results from the multivariate probit model suggested that besides specific bean attributes (such as cooking time, grain size, grain colour, grain damage, flavour, taste, familiarity, and flatulence) stipulated by literature, household preference of common beans in Lilongwe depends on socioeconomic factors, location, and prices. Results from the QuAIDS model showed that price and income changes affect budget shares of different varieties of common beans. In terms of the elasticities computed from QuAIDS results: all the significant income elasticities were positive and far below unity; income elasticities were far much less than expenditure elasticities; all the own-price elasticities were found to be above unity; the own-price elasticities were all negative except for mixed beans which had a positive own-price elasticity; the *Napilira* variety had a closer to unity own-price elasticities than the rest of the varieties; the cross-price elasticities were both positive and negative; and the price elasticities were larger in magnitude than income elasticities. Chapter six provides conclusions and recommendations based on the aforementioned results.

CHAPTER SIX

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary and Conclusions

Consumer demand information can signal the relative scarcity of goods and services, guide the decision of economic agents, and ensure mobility of commodities over time and across space. Henceforth, this study was conducted to analyse the demand of different varieties of common beans to price and income changes among households of different locations in Lilongwe District of Malawi. Specifically, the study (1) assessed the sensitivity of budget shares for six commonly traded varieties of common beans to price and income effects; and (2) assessed the responsiveness of the quantities demanded to price and income changes. The study made four hypotheses namely: (1) prices and income affect household budget share of common beans; (2) a unit percentage increase in own-price of a common bean variety decreases the quantity demanded by a less than unit percentage; (3) a unit percentage increase in price of a common bean variety increases the quantity demanded of other common bean varieties; and (4) a unit percentage increase in income of the household increases the quantity demanded of a common bean variety by less than a unit percentage.

To assess the sensitivity of budget shares of common beans to price and income effects, the study used household survey data and estimated a complete demand system that featured six commonly traded varieties of common beans. The household survey data was collected in Lilongwe district by the bean value chain research network and included information on 10 varieties of common beans. The QuAIDS model was employed for the analysis and featured six commonly traded varieties of common beans namely *Kabalabala*, *Napilira*, *Phalombe*, *Khaki*, *Nanyati*, and mixed beans. The model was estimated in a censored fashion were a two-stage approach was used in order to account for zero consumption of some of the varieties during the survey period. In the first stage a multivariate probit model was estimated. Thereafter, the cdf and pdf were predicted and fed into the second stage where the QuAIDS was estimated. The study estimated both the QuAIDS and its competitor LA/AIDS and based on the performance of each model chose the QuAIDS over the LA/AIDS as the best representation of demand for common beans among households in Lilongwe. In order to assess the responsiveness of quantity demanded of the varieties analysed to price and income changes, the study computed price and income elasticities from the QuAIDS results. Later, expenditure elasticities as proxies of income elasticities were used in calculation of income elasticities.

The descriptive analysis showed that households in Lilongwe district are mostly composed of adults and average about 4.15 adult equivalent. Majority consume common beans more than once per week and prefer the *Phalombe* variety. This variety is highly preferred in all areas (high, medium, and low-density areas). Results from the multivariate probit model suggest that besides specific bean attributes (such as cooking time, grain size, grain colour, grain damage, flavour, taste, familiarity, and flatulence) stipulated by literature, household preference of common beans in Lilongwe depends on socioeconomic factors, location, and prices. Further to this, the descriptive analysis revealed that purchase of common beans is very likely to be influenced by perceived nutrition and health benefits. A scrutiny of the non-consuming households showed that high prices, preparation inconvenience, preference, and health reasons are the factors that deter the purchase of common beans. In terms of pricing, no clear-cut pattern in magnitude of prices exist as we move from high to low-density areas, but the *Khaki* variety is the most expensive variety in the high and medium-density areas while *Napilira* is the most expensive in the low-density areas.

The QuAIDS model was estimated on the pooled sample (combining households in high, medium and low-density areas) and the results showed that the price and income effects were significant in some of the budget share equations. This means that the budget shares of common beans are sensitive to price and income effects. This necessitated the calculation of price and income elasticities from the QuAIDS results in order to assess the responsiveness of quantity demanded to price and income changes in Lilongwe district. All the significant income elasticities were positive and below unity suggesting that the common bean varieties analysed belong to the category of necessity goods. All the own-price elasticities were found to be above unity implying that the common bean varieties analysed have price-elastic demand. The own-price elasticities were all negative except for mixed beans which had a positive own-price elasticity suggesting that it belongs to the category of Giffen goods of which consumption increases with increase in price. The cross-price elasticities were both positive and negative suggesting both competitive and complementary cross-varietal relationships.

The inter-comparison of the own-price elasticities showed that the *Napilira* variety has a closer to unity own-price elasticity than the rest of the varieties. This implies that *Napilira* variety has the highest revenue generation capacity in Lilongwe. The comparison of magnitude income and expenditure elasticities showed that the income elasticities were far much less than the expenditure elasticities and this negates the use of expenditure elasticities to proxy income elasticities. The comparison of magnitude of price and income elasticities showed that price elasticities are larger hence quantity demanded is more sensitive to price than income effects.

The results led to the rejection of the second and third hypotheses which, respectively stipulated that a unit percentage increase in own-price of a common bean variety decreases the quantity demanded by a less than unit percentage and a unit percentage increase in price of a common bean variety increases the quantity demanded of other common bean varieties. This is because all the own-price elasticities were elastic and cross-varietal relationships were both competitive and complementary.

The study failed to reject the first and fourth hypotheses which stated that income of the household and prices affect the budget share of a common bean variety and a unit percentage increase in income of the household increases the quantity demanded of a common bean variety by less than a unit percentage. This is because price and income effects were significant in the QuAIDS and all the income elasticities were positive and below unity.

6.2 Recommendations

Based on the study results and conclusions, the following recommendations are made:

1. Revenue maximization from common beans traded by retailers can be maximized through reduction of prices

   Since all the own-price elasticities were above unity and highly significant, consumer demand for common beans is highly sensitive to price changes and retailers would generate more revenue by reducing prices. By reducing price, the proportionate change in price would be lower than the associated proportionate change in quantity demanded, thereby enabling higher revenue. Thus, retailers are encouraged to sell at lower prices for more revenue generation. However, special attention should be paid to customer segment and production costs.

2. If meaningful changes are to occur in common bean consumption, policy makers should pay more attention to pricing rather than to income-related policy instruments

   Comparison between income and price elasticities shows that price elasticities are larger. This implies high sensitivity in demand associated with price rather than income effects. Thus if meaningful changes are to occur in common bean consumption, as a direction, policy makers should pay more attention to pricing (focus on price-adjustments[60]) rather than to income-related (focus on changing incomes of consumers) policy instruments. However, a proper balance must be made between improving livelihood conditions of the farmers and taking care of consumer demand preferences.

6.3 Further Research

Further research on demand for common beans needs to extend from cross-sectional data to time series data in order to capture the effect of seasonal variations. Prices of common bean, just like any agricultural commodity, are affected by seasonality and it is essential to use time series data especially monthly data. In addition, the data has to be exhaustive of the variables in order to allow for advanced modelling. For instance, the data must contain all necessary information (such as expenditure and quantity of all

[60] This may include price floors and price ceilings

food commodities consumed) to allow for multi-stage consumer demand systems. Moreover, the data must contain information on availability and affordability by consumers since expenditure isn't purely a demand issue, it may also be an issue of market availability and affordability by consumers. On the other note, the sample size should be exhaustive enough to all estimation of demand models by location and income groups.

Further research also needs to extend from estimation of demand elasticities to calculation of welfare indicators such as change in consumer and producer surpluses, change in government revenue, net economic loss in consumption, net economic loss in production, and net effect of trade barriers and/or trade policies. Moreover, the elasticity estimates should feed into multimarket and CGE models that aim at explaining production and consumption, price formation, trade flows, welfare impacts of policies, income levels, and government fiscal revenues. In the estimation of the demand elasticities, expenditure elasticities should not be used to proxy income elasticities since in this study the expenditure elasticities showed extreme deviations in signs and magnitude from the income elasticities.

With the shift towards a demand driven economy, addressing the aforementioned gaps will help the policy making framework strike the balance between retaining the countries food security and the dual objective of improving livelihood conditions and taking care of consumer demand preferences with little impact on prices in order to protect the poor vulnerable population of the country.

REFERENCES

Abdulai, A. 2002. Household demand for food in Switzerland. A Quadratic Almost Ideal Demand System Approach. *Swiss Journal of Economics and Statistics*, 138(1):1-18. Available at: http://EconPapers.repec.org/RePEc:ses:arsjes:2002-i-1 [Accessed on 15 December 2016].

Aepli, M. 2014. Consumer demand for alcoholic beverages in Switzerland: a two-stage quadratic almost ideal demand system for low, moderate, and heavy drinking households. *Agricultural and Food Economics* 138(3):1-43. DOI:10.1186/s40100-014-0015-0. [Accessed on 10 November 2016].

Agostini, A. 2014. *The effect of food prices and household income on the British diet*. Institute for Social and Economic Research Working Paper No. 2014-10.Available at: https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2014-10.pdf. [Accessed on 10 November 2016].

Aidoo, R. 2009. *An Analysis of Yam Consumption Patterns in Ghanaian Urban Communities*. PhD dissertation. Kwame Nkrumah University of Science and Technology, Kumasi, Ghana. Available at: http://ir.knust.edu.gh/bitstream/123456789/252/1/PhD%20Thesis%20Robert%20Aidoo.pdf. [Accessed on 4 November 2016].

Akbulutgiller, K. 2008. *Application of Almost Ideal Demand System For A Pharmaceutical*. Master's thesis. Hoskolan Dalana University, Falun, Sweden. Available at: https://www.diva-portal.org/smash/get/diva2:518584/FULLTEXT01.pdf. [Accessed on 19 October 2016].

Akerele, D., Tiffin, R., and Srinivasan, C.S. 2013. *Household food demand in Nigeria: an application of Multivariate Double-Hurdle Model*. Paper prepared for presentation at the 87th Annual Conference of the Agricultural Economics

Society, University of Warwick, United Kingdom Association. Available at: https://ideas.repec.org/p/ags/aesc13/158700.html. [Accessed on 15 October 2015].

Akibode, C.S., and Maredia, M. 2011. *Global and Regional Trends in Production, Trade and Consumption of Food Legume Crops*. Available at: http://ispc.cgiar.org/sites/default/files/images/ Legumetrendsv2.pdf. [Accessed on 15 October 2015].

Alviola, P. 2010. Micro-demand systems analysis of non-alcoholic beverages in the United States: an application of econometric techniques dealing with censoring Available at: http://ageconsearch.umn.edu/bitstream/60462/2/MicroDemand%20Systems%20Analysis%20of%20NonAlcoholic%20in%20the%20United%20States%20An%20Application%20of%20Econometric%20Techniques%20Dealing%20with%20Censoring.pdf. [Accessed on 2 June 2016].

Amemiya, T. 1974. Multivariate regression and simultaneous equation models when the dependent variables are truncated normal. *Econometrica* 42:999-1012. Available at. https://conservancy.umn.edu/bitstream/handle/11299/54864/1976-79.pdf?sequence=1. [Accessed on 2 June 2016].

Asche, F., and Wessells, C.R. 1997. On price indices in the Almost Ideal Demand System. *American Journal of Agricultural Economics* 79:1182-1185. Available at: https://academic.oup.com/ajae/article-abstract/79/4/1182/64607/On-Price-Indices-in-the-Almost-Ideal-Demand-System?redirectedFrom=PDF. [Accessed on 11 July 2016].

Bareja, G. 2016. *List of pulses or grain legumes and their various names*. Available at: http://www.cropsreview.com/grain-legumes.html [Accessed on 20 October 2016].

Barslund, M. 2011. Censored demand system estimation with endogenous expenditures in clustered samples: an application to food demand in urban Mozambique. LICOS Discussion Paper Series No. 280/2011. Available at: https://papers.ssrn.com/sol3/ papers.cfm?abstract_id=1850525. [Accessed on 11 July 2016].

Banks J., Blundell R., Lewbel A. 1997. Quadratic Engel curves and consumer demand. *The Review of Economics and Statistics* 79(4): 527-39. Available at: http://www.jstor.org/stable/2951405. [Accessed on 24 July 2015].

Binger, B., and Hoffman, E. 1998. *Microeconomics with Calculus*. 2nd ed. Reading, Massachusetts: Addison-Wesley.

Birachi, E. 2012. Value chain analysis of beans in eastern and southern: building partnership for impact through research on sustainable intensification of farming system. International Institute of Tropical Agriculture. Available at: https://cgspace.cgiar.org/handle/10568/24878?show=full. [Accessed on 3 July 2015].

Blundell R, and Robin, J.M. 1999. Estimation in large and disaggregated demand systems: An estimator for conditionally linear systems. *Applied Economics* 14(3):209–232. DOI:10.1002/(SICI)1099-1255(199905/06)14:3<209::AID-JAE532>3.0.CO;2-X. [Accessed on 3 January 2016].

Bopape, L. E. 2006. The influence of demand model selection on household welfare estimates: an application to South African food expenditures. PhD thesis, Department of Agricultural economics, Michigan State University, East Lansing, Michigan. Available at: http://s3.amazonaws.com/zanran_storage/www.aec.msu.edu/ContentPages/344989.pdf. [Accessed on 3 July 2016].

Cappellari, L., and Jenkins, S. 2003. Multivariate probit regression using simulated maximum likelihood. *The Stata Journal* 3(3):278–294. Available at: http://www.bc.edu/repec/usug2003/Capelari_Jenkins-UKSUG2003.pdf. [accessed on 10 February 2016].

Chern, W.S., Ishibashi, K., Taniguchi, K., and Yokoyama, Y. 2003. *Analysis of Food Consumption Behavior by Japanese Households*. Food and Agricultural Organization. Availability at. https://www.researchgate.net/profile/Wen_Chern/publication/5021699_Analysis_of_Food_Consumption_Behavior_by_Japanese_Households/links/0046353ae502c8f97c000000/Analysis-of-Food-Consumption-Behavior-by-Japanese-Households.pdf. [Accessed on 9 December 2016].

Chirwa, R. M., and Phiri, M. A. 2007. *Factors that influence demand for beans in Malawi*. CIAT. Highlights series. No. 37. Available at: http://ciat-library.ciat.cgiar.org/Articulos_Ciat/Factors_affecting_demand_for_beans.pdf. [Accessed on 3 December 2015].

Centro Internacional de Agricultural Tropical, 2012. *Leveraging legumes to combat poverty, hunger, malnutrition and environmental degradation*. CGIAR Research program on grain legumes.

Centro Internacional de Agricultural Tropical.2013. *Tropical legume farming in Malawi*. Bulletin of the tropical legumes.

De Bekker-Grob, E.W., Ryan, M. and Gerard, K. 2012. "Discrete choice experiments in health economics: a review of the literature". *Health Economics*. 21: 145-172. doi:10.1007/s40273-014-0170-x. [Accessed on 12 December 2015].

Deaton, A. 1986. *Demand Analysis. New York*: Elsevier Science Publisher. Available at: https://scholar.princeton.edu/deaton/publications/demand-analysis. [Accessed on 12 December 2015].

Deaton, A., and Muellbauer, J. 1980. An Almost Ideal Demand System. *American Economic Review* 70(3):312-326. Available at: https://scholar.princeton.edu /deaton/publications/demand-analysis. [Accessed on 12 December 2015].

Eales, J., and Unnevehr L. 1991. The inverse almost ideal demand system. Available at: http://www.farmdoc.uiuc.edu/nccc134. [Accessed on 18 June 2015].

Food and Agricultural Organization of the United Nations. Dry Beans production in Malawi. Available at: http://faostat3.fao.org/analysis/Q/QC/E [Accessed on 3 April 2015].

Food and Agricultural Organization of the United Nations. Population in Malawi. Available at: http://faostat3.fao.org/analysis/Q/QC/E [Accessed on 3 April 2015].

Gonzalez, A., and Wieck, C. 2014. *Food and nutrient demand in the context of the conditional cash transfer in Mexico*. Poster paper prepared for presentation at the EAAE 2014 Congress 'Agri-Food and Rural Innovations for Healthier Societies'. Available at: https://ageconsearch.umn.edu/record/182834/files/Poster_Paper._Meza._Food_and_Nutrient_Demand_in_the_context_of_the_Conditional_Cash_Transfer_Program.pdf. [Accessed on 3 November 2015].

Government of Malawi. 2013. *Annual agricultural statistical bulletin*. Lilongwe, Malawi: Ministry of Agriculture Irrigation and Water Development.

Greene, W.H. 2003. *Econometric Analysis*. 5th ed. Upper Saddle River, N.J. : Prentice Hall.

Gujarati, D.N. 2004. *Basic Econometrics*. 4th ed. New York: McGraw-Hill Company.

Hajivassiliou, V., and Ruud, P. 1994. Classical estimation methods for LDV models using simulation. *Handbook of Econometrics* 4(4):2383–2441. DOI:10.1016/S1573-4412(05)80009-1. [Accessed on 9 December 2015].

Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47(1):153-161. Available at: https://faculty.smu.edu/millimet/classes/eco7321/papers/heckman02.pdf. [Accessed on 8 December 2015].

Heien, D., and Wessells, C.R. 1990. Demand systems estimation with microdata: a censored regression approach. *Journal of Business and Economic Statistics* 102(8):365−371. Available at: http://www.jstor.org/stable/1391973. [Accessed on 10 December 2015].

Hikaru, H., and Kozumi, H. 2001. Bayesian analysis on Engel curves estimation with measurement errors and an instrumental variable. *Journal of Business and Economic Statistics* 19(8):292–298. http://www.jstor.org/stable/1392031. [Accessed on 10 December 2015].

Huber, C. C. 2012. *An estimated system of demand for specialty potato varieties: Marketing Implications from a multi-city comparison*. Master's thesis, Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, Colorado. Available at: https://dspace.library.colostate.edu/bitstream/handle/10217/68118/Huber_colostate_0053N_11249.pdf?sequence=1. [Accessed on 8 December 2015].

Kankwamba, H., Mangisoni, H., Simtowe, J., Mausch, F., and Siambi, M. 2012. *Improved legume seed demand systems in central Malawi: what do farmers' seed expenditures say about their preferences*? Paper provided by International Association of Agricultural Economists in its series 2012 Conference, August 18-24, 2012, Foz do Iguacu, Brazil with number 123945. Available at: http://ageconsearch.umn.edu/bitstream/131684/2/H_kankwamba_IAAEx.pdf. [Accessed on 8 December 2015].

Katungi, E., Farrow, A., Chianu, J., Sperling, L., and Beebe, S. 2009. *Common bean in Eastern and Southern Africa: a situation and outlook analysis*. CIAT. Highlights series. No. 12. Available at: https://www.researchgate.net/publication/228601612_Common_bean_in_Eastern_and_Southern_Africa_a_situation_and_outlook_analysis. [Accessed on 3 December 2015].

Kawabata, H. 2011. *Labor Market Entrants and Food Consumption Patterns by Agricultural Households in Malawi: Econometric Application of Agricultural Household Models*. PhD dissertation. Graduate School of International Cooperation Studies, Kobe University, Japan. Available at: http://www.lib.kobe-u.ac.jp/repository/thesis/d1/D1005590.pdf. [Accessed on 5 December 2015].

Kebede, B. 2003. Intra-household Distribution of Expenditures in Rural Ethiopia: A Demand Systems Approach. The Centre for the Study of African Economies Working Paper Series. Available at: https://www.csae.ox.ac.uk/workingpapers/pdfs/2003-08text.pdf. [Accessed on 5 December 2015].

Kenkel, B., and Signorino, C. 2012. An alternative solution to the Heckman selection problem: selection bias as a functional from misspecification. *University of*

Roadster 15(3):100-121. Available at: https://www.cambridge.org/core/membership/spmfiles/polmeth/selection_1.pdf. [Accessed on 5 December 2015].

Khaliukova, O. 2014. Demand Analysis for Tomato, Onion, Peppers, and Fresh Okra in Nigeria. Master's thesis, University of Tennessee, Knoxville, Tennessee, United States. Available at: http://trace.tennessee.edu/cgi/viewcontent.cgi?article=2839&context=utk_gradthes. [Accessed on 5 December 2015].

Kinsey, J.D. 2001. The new food economy-consumers, farms, pharms, and science. *American Journal of Agricultural Economics* 83(5):1113-1130. Available at: http://www.management.wharton.upenn.edu/raff/documents/Jean%20Kinsey%20EOI%20background%20reading%20No.%202.pdf. [Accessed on 5 December 2015].

Klein, R. L., and Spady, R.H. 1993. An efficient semiparametric estimator for binary response models. *Econometrica* 61:387-421. Available at: http://www.crest.fr/ckfinder/userfiles/files/pageperso/bertail/smi/semi_probit.pdf. [Accessed on 6 December 2015].

Kumar, P., Kumar, A., Parappurathu, S., and Raju, S.S. 2011. Estimation of demand elasticity for food commodities in India. *Agricultural Economics Research Review* 24(1):1-14. Available at: http://dairyknowledge.in/sites/default/files/estimation_of_demand_elasticity_for_food_commodities_in_india.pdf. [Accessed on 6 December 2015].

Lancaster, K.J. 1966. A new approach to consumer theory. *Econometrica 2*:132-157. DOI:10.1086/259131. [Accessed on 6 December 2015].

Larochelle, C., Katungi, E., and Beebe, S. 2015. *Disaggregated analysis of bean consumption demand and contribution to household food security in Uganda*.

CIAT. Highlights series. No. 46. Available at: https://www.researchgate.net/publication/279203381_disaggregated_analysis_of_bean_consumption_demand_and_contribution_to_household_food_security_in_uganda. [Accessed on 3 December 2015].

Lazaro, E. 2014. An Empirical Analysis of Rice Demand in Tanzania. *Agricultural Economics*, 48: 187–196. DOI:10.1111/agec.12325. [Accessed on 6 December 2015].

Lewbel, A. 1996. Demand estimation with expenditure measurement errors on the left and right hand side. *The Review of Economics and Statistics* 78(4):718–725. Available at: http://econpapers.repec.org/article/tprrestat/v_3a78_3ay_3a1996_3ai_3a4_3ap_3a718-25.htm. [Accessed on 6 December 2015].

Liu, K. 2006. *A Quadratic Generalization of the Almost Ideal and Translog Demand*. Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Long Beach, California. Available at: http://ageconsearch.umn.edu/bitstream/21387/1/sp06li02.pdf. [Accessed on 6 December 2015].

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Econometric Society Monographs No. 3. New York: Cambridge University Press.

Maganga, A.M., Phiri, M.A.R., Mapemba, D.L., Gebremariam G.G and Dzanja, K.J. 2014. *A Food Demand System Estimation for Rural Malawi: Estimates Using Third Integrated Household Survey Data*. Selected Paper presented at the Agricultural and Applied Economics Association annual meetings in Minneapolis, MN, July 28-30, 2014. DOI: 10.13140/2.1.3072.8485. [Accessed on 9 December 2015].

Mangisoni, J., and Bokosi. J. 2004. Socio-economic impact assessment of Kalima and Napilira bean varieties on smallholder farmers in a Smallholder Seed Security Multiplication Scheme in Dedza District of Central Malawi. *Bunda Journal of Agriculture, Environmental Science and Technology* 2(1):1-12.

Mishili, J.F., Fulton, j., Shehu, M., Kushwaha, S., Marfo, K., Jamal, M., Kergna, A., and DeBoer, J. 2009. Consumer preferences for quality characteristics along the cowpea value chain in Nigeria, Ghana, and Mali. *Agribusiness* 25(1):16–35. DOI: 10.1002/agr.20184. [Accessed on 9 December 2015].

Mkanda, A.V. 2007. Relating consumer preferences to sensory and physicochemical properties of dry beans (Phaseolus vulgaris L.). Master's thesis. University of Pretoria, Pretoria, South Africa. DOI: 10.1002/jsfa.3046. [Accessed on 8 December 2015].

Moscati, I. 2007. History of consumer demand theory 1871–1971: A Neo-Kantian rational reconstruction. *The European Journal of the History of Economic Thought* 14(1):119–156. DOI:10.1080/09672560601168504. [Accessed on 8 December 2015].

Moschini, G. 1995. Units of measurement and the Stone index in demand system. *American Journal of Agricultural Economics* 77:63-68. Available at: http://www.jstor.org/stable/1243889. [Accessed on 8 December 2015].

Munthali, G. 2013. Assessing farmer's willingness-to-pay for improved common bean seed varieties in Malawi: a case study of Kasungu and Dedza Districts. Master's thesis. University of Pretoria, Pretoria, South Africa. Available at: https://repository.up.ac.za/bitstream/handle/2263/40344/Munthali_Assessing_2013.pdf?sequence=1. [Accessed on 8 December 2015].

Musyoka, M.P., Kavoi, M.M., and Omiti, J.M. 2007. Food consumption patterns and distributional welfare impact of import tariff reduction on cereals in Kenya. *African Journal of Agricultural and Resource Economics* 9:183-199. Available at:http://ageconsearch.umn.edu/bitstream/183892/2/2.%20Musyoka%20 Kavoi%20%20Omiti.pdf. [Accessed on 9 December 2015].

Muthoni, R., Barungi, M., Chirwa, R., Chianu, J., and Birachi, E. 2008. *Determining the market share of improved common bean varieties traded in selected markets in Malawi*. FAO: AGRIS. Available at: http://agris.fao.org/agris-search/search.do?recordID=GB2013201808. [Accessed on 9 December 2015].

Mwale, V.M., Bokosi, J.M., Masangano, C.M., Kwapata, M.B., Kabambe, V.H., and Miles, C. 2008. Yield performance of dwarf bean (Phaseolus vulgaris L.) lines under researcher designed farmer managed system in three bean agro-ecological zones of Malawi. *Africa Journal of Biotechnology* 7(16): 2847-2853. Available at: https://www.ajol.info/index.php/ajb/article/view/59179/0. [Accessed on 9 December 2015].

Nicholson, W., and Snyder, C. 2008. *Microeconomic Theory: Basic Principles and Extensions.* 10th ed. London: South-Western (Thomas Learning).

National Statistical Office Malawi. 2012. *Household socioeconomic characteristics report.* Integrated Household Survey 2011-2012. Available at: http://microdata.worldbank.org/index.php/catalog/1003. [Accessed on 17 January 2016].

National Statistical Office Malawi. 2016. Demographic and Health Indicators. Available at: https://dhsprogram.com/pubs/pdf/PR73/PR73. [Accessed on 17 January 2016].

Parraga, I. 1990. Determinants of Food Consumption. *Journal of the American Dietetic* 18:312-323. Available at: https://www.ncbi.nlm.nih.gov/pubmed/2335679. [Accessed on 9 December 2015].

Raunikar, R., and Huang C.H. 1987. *Food demand analysis, problems, issues, and empirical evidence.* Ames: Iowa State University Press. Available at: http://trove.nla.gov.au/work/12170124?q&versionId=14356406 [Accessed on 4 June 2015].

Ravallion, M. 1986. Testing market integration. *American Journal of Agricultural Economics* 68:102-109. Available at: http://www.jstor.org/stable/1241654. [Accessed on 4 August 2015].

Rusike, j., Brand, G., Boahen, S., Dashiell, K., Kantengwa, S., Ongoma Mongane, J.D., Kasongo, G., Jamagani, Z.B., Aidoo, R., and Abaidoo, R. 2013. Value chain analyses of grain legumes in N2Africa: Kenya, Rwanda, eastern DRC, Ghana, Nigeria, Mozambique, Malawi and Zimbabwe. Available at: https://agriknowledge.org/downloads/47429915b. [Accessed on 4 June 2015].

Sadoulet, E. and Janvry, A. 1995. *Quantitative development policy analysis*. Baltimore MD: The Johns Hopkins University Press. Available at: http://hdl.handle.net/10.2307/1243800. [Accessed on 3 July 2015].

Sheng, T.Y., Shamsudin, M.N., Mohamed, Z., Abdullah, A.M., and Alias, R. 2010. Demand analysis of meat in Malaysia. *Journal of Food Products Marketing* 6(2):4-14. DOI:10.1080/10454440903415105. [Accessed on 3 June 2015].

Shephard, R.W. 1953. *Cost and Production function*. Princeton University Press, Princeton.

Shonkwiler, J.S., and Yen, S.T. 1999. Two-step estimation of a censored system of equations. *American Journal of Agricultural Economics* 81:972−982. Available at https://academic.oup.com/ajae/article-abstract/81/4/972/68457/Two-Step-Estimation-of-a-Censored-System-of?redirectedFrom=PDF. [Accessed on 21 August 2015].

Sichilima, T., Mapemba, L., and Tembo, G. 2015. What determines expenditure allocation to beef among Lusaka residents in Zambia? Evidence from household Survey. *Modern Economy,* 6:411-422. DOI:10.4236/me.2015.63039. [Accessed on 21 August 2015].

Sivaramane, N. 2012. Demand Models. L.A.S.R.I, Library Avenue, Pusa, New Delhi.

Sola, O. 2013. Demand for food in Ondo state, Nigeria: Using quadratic almost ideal demand system. *Journal of Business Management and Economics* 4(1):001-019. Available at: http://e3journals.org/cms/articles/1358596271_Olorunfemi.pdf. [Accessed on 5 August 2015].

Leserer, M.P., 2002. *An Almost Ideal Demand System for food: based on cross-sectional data; Rural and Urban east java, Indonesia*. PhD thesis, Georg-August Universitaet Goettingen, Indonesia. Available at: http://hdl.handle.net/11858/00-1735-0000-0006-B062-C. [Accessed on 5 August 2015].

Taniguchi, K., and Chern, W. (2000). *Income elasticity of rice demand in Japan and its implications: cross-sectional data analysis.* Paper prepared for 2000 AAEA meeting. Available at: http://econpapers.repec.org/paper/agsaaea00/21755.htm. [Accessed on 5 August 2015].

Taylor, C. 2014. "An Almost Ideal Demand System Analysis of Orange and Grapefruit Beverage Consumption in the United States". *All theses.* Paper 1913. Available at: http://tigerprints.clemson.edu/all_theses/1913 [Accessed on 3 July 2015].

Temitrope, E., and Haruna, A. 2013. Economic implications of food demand in Akoko south west, Ondo state, Nigeria: AIDS approach. *American journal of humanities* *and social sciences* 1(2):80-86. Available at: https://www.google.com /url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjqisCg0pTWAhVL6iYKHThQClQQFggnMAA&url=http%3A%2F%2Fwscholars.com%2Findex.php%2Fajhss%2Farticle%2Fdownload%2F308%2Fpdf&usg=AFQjCNFc6IKz6FE1bqVPNmzbsaMuPVeRMg. [Accessed on 3 July 2015].

Theil, H. 1975. *Theory and measurement of consumer demand.* Amsterdam: North-Holland Publishing Company.

Tobin, J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica* 26:24-36. Available at: https://web.sonoma.edu/users /c/cuellar/econ411/Tobin.pdf. [Accessed on 5 July 2015].

United State Agency for International Development. 2011. Feed the Future 2011-2015 Multi-year strategy.

Varian, H. 1987. *Intermediate microeconomics: A modern approach.* 8th ed. New York: W.W. Norton & Co.

Walker, J., and Akiva, M. 2002. Generalized random utility model. *Mathematical Social Sciences* 2(43):303–343. DOI:10.1016/S0165-4896(02)00023-9. [Accessed on 4 July 2015].

Wetzstein, M. 2004. *Microeconomic theory: concepts and connections with economic applications.* Kentucky: South-Western College Publishers.

Yen, S., and Lin, B. 2004. *A sample selection approach to censored demand systems*. Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting. Available at: ttp://www.jstor.org/stable/3697762. [Accessed on 1 July 2015].

APPENDIX

Results from the estimation of the Engel function

Table A. 1: Results from the estimation of the Engel function

Variable	Coefficient	T
Lnhhincome	0.1090**	2.08
	(0.0909)	
LnP^L	0.1130	0.54
	(0.2101)	
Constant	4.4666***	1.37
	(1.3674)	

Note: Values in brackets are standard errors, *** significant at 1%, ** significant at 5%, * significant at 10