CHAPTER 3

ANALYTICAL FRAMEWORK

3.1 INTRODUCTION

This study employs a macro-micro approach to meet its set objectives. This chapter presents the technical details and the theoretical foundations on which macro-micro modelling/analysis is based. This is achieved by firstly providing detailed theoretical foundations for the micro-component of the macro-micro model developed for this study, followed by a general overview of the macro-component. An in-depth description of the latter is provided in Chapter 5. The last section addresses how linkages are established and how consistency is achieved between the macro and micro components.

In any macro-micro study, it is important to contextualise the terms "micro" and "macro", as they can have different meanings in different contexts. Macro, in this study, does not refer to macro-economic policies; but refers to the integrated national maize market in Malawi; which is a country level "market" with many components that make up the Malawi maize sector. The micro component is the household level which consists of individual rural farming households engaged in the production and marketing of maize. The rural household economy modelled in Chapter 5 can be regarded as the meso-level found between the national maize market and rural farm household.

3.2 THE MICRO-COMPONENT

A full structural estimation of household behaviour that takes into account different aspects of a household's complete activities, such as consumption, production, labour supply and other social economic activities, is not econometrically possible (Cogneau & Robillard, 2004) – although theoretically possible. Instead, common practice is to focus on modelling the aspects of the household's behaviour that is relevant for the study. This study is mainly concerned with assessing the impact of macro-economic policy changes and agricultural prices policies on household incomes. Therefore, at the micro level, the focus is on estimating incomes for different rural household groups.

In rural Malawi, household income is not synonymous with cash wage income, but is rather a computed value that consists of wage income, income earned from self-employment and the imputed value of crops and livestock kept for home consumption. In this study, aggregate income constituted ten different income sources (Chapter 4). In such cases, household income estimates comprise a system of equations that consists of the earnings equation for household members who are of working age and who therefore contribute economically to the household; self-employment equations that estimate the potential income that can be earned from self-employment; and occupational choice equations that describe the allocation of time by individuals within the household who are of working age between wage employment, self-employment and non-farm/non-labour time.

The methodology for estimating rural household income is given below. This is adapted from the work of Bourguignon *et al*. (2001) and Robilliard *et al*. (2001), which was further adapted by Alantas and Bourguignon (2004):

Let y_{it} denote household income for household i at time t. It is assumed that income (y_{it}) depends through a function $Y(\ )$ and there are four main arguments governing the relationship:

i) The observed and unobserved socio-economic characteristics of different household members denoted x and ε, respectively.
ii) The set of remuneration prices at which these characteristics are given in the labour market (β).
iii) A set of parameters that define the participation and occupational choices of different household members (λ).

Given this, household income can be expressed as follows:

$$y_{it} = Y\left(x_{it}, \varepsilon_{it}, \beta_t, \lambda_t\right) \qquad (3.1)$$

The following four equations provide a decomposition of Equation 3.1:

$$Logw_{mi}^{t} = X_{mi}^{t}\beta^{t} + u_{mi}^{t} \;\; ; \;\; i = 1, \ldots k_m \qquad (3.2)$$

$$L_{mi}^{t} = Ind\left[X_{mi}^{t}\lambda_{1X}^{t} + Z_{mi}^{t}\lambda_{iZ}^{t} + v_{mi}^{t}\right]; \; i = 1, \ldots k_m \qquad (3.3)$$

$$L_{mi}^{At} = Ind\left[X_{mi}^{t} \lambda_{2X}^{t} + Z_{mi}^{t} \lambda_{2Z}^{t} + v_{mi}^{Ai} \right];\ i = 1, \dots k_m \quad (3.4)$$

$$L_{mi}^{NAt} = Ind\left[X_{mi}^{t} \lambda_{3X}^{t+} + Z_{mi}^{t} \lambda_{3Z}^{t} + v_{mi}^{NAt} \right];\ i = 1, \dots k_m \quad (3.5)$$

Equation 3.2 is a typical Mincer earnings equation and it gives the log of the wage earnings[3] for individual household members (i) for a specific household (m) expressed as a function of a set of personal characteristics (X). u_{mi}^{t} is the residual term and it describes the effects of unobserved determinants of wage earnings. Earnings are modelled as a function of personal characteristics of only those household members who are able to earn an income. This enables the capturing of the heterogeneity of earnings within different income groups which may arise due to differences in socio-economic characteristics (Ahmed & Donoghue, 2010).

Equations 3.3, 3.4 and 3.5 represent the occupational choices available to each household member who contributes towards household income. In this study, there are ten different sources of income that were regrouped into these three occupational choices. Firstly, salaried employment included both on and off-farm income falling under Equation 3.3 (L_{mi}^{t}) represents the labour supplied by a household member i as a wage earner outside the household. Second, the real and imputed income from all upland crops except maize, livestock income, income from wetlands and the real and imputed value of maize falling into Equation 3.4 (L_{mi}^{At}). This represents all labour supplied by a member of the household i on the farm that the household owns and cultivates. Third, non-agro based commercial enterprises and agro-based commercial enterprises of the household falls under Equation 3.5 (L_{mi}^{NAt}), which represents the labour supplied by a household member i on the family's non-farm/non-labour income-generating activities. The v_{mi}'s are the residual terms and they represent the unobserved factors that determine household members occupational choices.

The co-efficient estimates for the occupational choices of different households members (λ) are represented by the X_{mi}'s and Z_{mi}'s and, as can be observed from the three occupational choice equations, these are common for all individuals within a household. However, these

[3] The Mincer earnings equation is commonly a log-linear function because it is an empirical operationalisation of the human capital model. The human capital model stipulates that individuals invest in human capital up to a point where investment costs just equal the present value of schooling gains. This implies that during the post school phase, levels of investment in human capital decline monotonically until it reaches zero at retirement. Since potential earnings are proportional to human capital stock, it implies that potential earnings also increase at a decreasing rate over a life cycle. The log-liner earnings wage function is therefore the empirical application of the linear declining post-school investment theory. See Lemieux (2006) and Polackek (2007) for details.

may differ across demographic groups' i.e. between male and female household members or between older and younger household members. The vector Z captures the household characteristics that include mainly the productive assets such as land; while X is a vector of the standard human capital variables such as years of schooling and access to trainings.

A review of the literature shows that many studies include in the vector for household characteristics, different variables such as area of residence, demographic composition of the household, resource endowments, the incomes of other household members, and human capital as well as consumers preferences (Robillard, *et al.* 2001; Kuepie, *et al.* 2007; Ahmed & Donoghue, 2010). However, other estimations of labour income include additional variables that have the potential to lead to household heterogeneity such as market access (Lay, 2006), labour market experience, occupation of the household head and other income source variables (Ferreira, *et al.* 2003). De Hoyos (2005) goes further to create subsets of the Z vector that take into account gendered differences of the male and female household members. Thus, in addition to the standard variables, he includes in the subset for female household members the number of children in the household and the sex of the household head.

This study goes beyond any of these studies in that the vector for household characteristics, differentiated households further by capturing their participation in the ERI initiative. This is achieved by creating subsets for participating and non-participating households that contain identical household characteristic variables, with the only difference being participation in the ERI initiative (the case study innovative agricultural research intervention for this research). The inclusion of participation in the ERI initiative in Z captures heterogeneity of households that may be created by changes at the household level arising from participating. Since the ERI initiative is rather recent, it is unlikely that other studies in literature have incorporated such a parameter in the estimation of occupational choices. The subsets in Z are presented by the notation e and Equations 3.3 to 3.5 can be rewritten as follows:

$$L_{mi}^{t} = Ind\left[X_{mi}^{t}\lambda_{1X}^{t} + Z_{mie}^{t}\lambda_{iZ}^{t} + v_{mi}^{t} \right]; i = 1,...k_m; \quad e = \begin{cases} 1 ERI-participation \\ 0 ERI-nonparticpation \end{cases} \tag{3.6}$$

$$L_{mi}^{At} = Ind\left[X_{mie}^{t}\lambda_{2X}^{t} + Z_{mie}^{t}\lambda_{2Z}^{t} + v_{mi}^{Ai} \right]; i = 1,...k_m; \quad e = \begin{cases} 1 ERI-participation \\ 0 ERI-nonparticpation \end{cases} \tag{3.7}$$

$$L_{mi}^{NAt} = Ind\left[X_{mie}^{t}\lambda_{3X}^{t+} + Z_{mie}^{t}\lambda_{3Z}^{t} + v_{mi}^{NAt} \right]; i = 1,...k_m; \quad e = \begin{cases} 1 ERI-participation \\ 0 ERI-nonparticpation \end{cases} \tag{3.8}$$

The labour supply parameter λ has three subsets (λ_1 for wage labour; λ_2 for self-employment; and λ_3 for non-farm/non-labour income) that differ depending on the household member under consideration.

Occupational choices are all modelled as a discrete logit model in which the household member has only two choices to either supply labour or not to supply labour. Labour is modelled as a discrete variable because the actual number of working hours per household member in each occupation is not captured (Robilliard, *et al.* 2001; Alantas & Bourguignon, 2004). This modelling technique is represented in the three occupational choice models by the notation *Ind*.

The aggregate household income is then obtained using an accounting identity that sums the wage income, earnings from self-employment (value of own production) and non-labour income as well as an exogenous variable for incomes earned from remittances and other cash transfers.

Aggregate rural household income can therefore be presented as follows:

$$y_m^t = \sum_{i=1}^{k_m} L_{mi}^t w_{mi}^t + \prod\nolimits_A^t \left[Z_m^{Tt}, \sum_{j=1}^{k_m} L_{mi}^{At}, s_A^t, \beta_A^t \right] + \prod\nolimits_{NA}^t \left[Z_m^{Tt}, \sum_{j=1}^{k_m} L_{mi}^{NAt}, s_{NA}^t, \beta_{NA}^t \right] + y_{0m}^t \qquad (3.9)$$

Where:

$\prod_A^t$ Profit function for self-employment income (value of own production)

$\prod_{NA}^t$ Profit function for non-farm/non-labour income

y_{0m}^t Exogenous variable representing transfers of income (remittances)

s_A^t and s_{NA}^t are the residuals of the profit functions which can represent either the unobserved factors determining profits or measurement errors

Equation 3.9 shows that aggregate household income is a non-linear function of wage income, profits generated from self-employment, profits generated from non-farm/non-labour income, various occupational choices for different household members and an exogenous parameter from transfers (y_{0m}^t), and the observed household characteristics of different household members that are of income-earning age.

In this study, both the wage income and profits were observed and captured at the household level, and not on an individual household member basis. This therefore is equivalent to modelling household income for a household with only one member, i.e. the household head. As such, the household income model was reduced to an arithmetic function that involved the summation of the different sources of household income. The arithmetic estimation of income reduced the household income generation model to a simple micro-accounting model. This is feasible for this study as the shock to be implemented does not alter the occupational choices of household members. Micro-accounting is fully consistent with micro-economic behaviour but the estimated models can only be used for short-term and/or medium-term analysis, due to the lack of behavioural responses that make them unsuitable for long-term analysis (Bourguignon, *et al.* 2008).

3.3 THE MACRO-COMPONENT

In literature, the most commonly-used tool for macro-economic modelling in macro-micro analysis is the Computable General Equilibrium (CGE) model. This is evidenced by numerous macro-micro studies assessing macro-economic policy shocks on household incomes and income distributions (Thurlow & Van Seventer, 2002; Robilliard, *et al.* 2001; Bourguignon, *et al.* 2008; Ferreira, *et al.* 2003; Ahmed & Donoghue, 2010; Diao, 2010). However, any other type of macro-economic modelling techniques (e.g Vector Auto Regression (VAR) models, Dynamic Stochastic General Equilibrium (DSGE) models, Partial Equilibrium Models (PEM), Structural cointergration VAR models (SVAR)); can be used (Bourguignon, *et al.* 2010).

For this study, a partial equilibrium framework was chosen as a tool for modelling the national (macro-level) maize market in Malawi, as it is capable of reasonably approximating the general effects of agricultural policy changes where weak links between commodities and their supplier or output sectors exist (Perali, 2003). In addition the approach allows for dynamism in the analysis as one can simulate the impact of policies over time. Due to the complexities of constructing and solving dynamic CGE model; partial equilibrium approaches are more appropriate for Malawi as the latter provides a less generalised, more disaggregated and more realistic picture of the general effects of macro-economic policy changes. Partial equilibrium models can be either single-market or multi-market models that represent a system without linkages with the rest of the economy (Britz & Heckelei, 2008; Van Tongeren, *et al.* 2001). The effects of and/or effects on the rest of the economy and/or on other sectors can be included in a top-down approach (Van Tongeren, *et al.* 2001).

3.3.1 General structure of model and economic theory

A typical partial equilibrium framework consists of four blocks that are the supply, demand, trade and price blocks. Domestic supply in partial equilibrium models is the summation of domestic production and beginning stocks. According to economic theory, production or supply of an agricultural commodity is dependent upon farmers' willingness to produce the commodity based on expected prices (Ferris, 1998). Expected prices influence agricultural supply because of the biological nature of agricultural production, which leads to time lags between when decisions are made and when output is attained. In the Malawi case, the absence of a futures market and limited forward contracting, which both allow farmers to establish price at the time when production decisions are made (Ferris, 1998), entail that price expectations are formed on the basis of past prices. In rural areas of Malawi where markets are thin and where there is poor access to markets due to high transaction costs and poor transport logistics, producers receive prices that are lower than prices prevailing in the national maize market or in other more lucrative urban markets. The dependency on past prices entails that maize prices received by rural households are consistently lower.

Apart from expected own prices, economic theory further stipulates that the production of an agricultural commodity is influenced by the prices of competing and/or complementary commodities, the price of inputs or technology and climatic changes. In Malawi, the majority of smallholder maize producers are subsistence farmers who do not substitute maize for other crops regardless of the market prices. In addition, the maize-based farming system in Malawi is mainly a mixed farming system in which maize is intercropped with various grain legumes that are planted in much smaller amounts. These two characteristics of the maize-based farming system in the country entail that the inclusion of prices for competing or complementary commodities in the modelling of maize supply is computationally difficult. Changes in climatic conditions, especially those pertaining to variations in water availability such as rainfall and droughts, play a major role in maize production in the country; as agricultural production is highly dependent on rain-fed farming with little or no irrigation, thus making it highly susceptible to climatic variability. Dynamism in production is captured in the partial equilibrium model, as production is calculated as an identity that is dependent upon the area of maize planted multiplied by crop yields; thus incorporating the biological lags associated with agricultural production and climatic variability.

The domestic demand block is composed of domestic human consumption, seed and industrial use, and ending stocks. Economic theory stipulates that consumption of a commodity is affected by many factors with the price of the commodity, the price of substitutes, changing tastes over time and the income of the consumer being key (Ferris, 1998). In rural Malawi the majority of smallholder producers consume their own production with supplementation from the market to meet any shortfall in subsistence needs. This implies that maize consumption may not be very responsive to market prices. In addition, food baskets of rural households in the country are non-diversified, with maize being the main food crop with little or no substitutes.

Over time, empirical evidence shows that societies tend to move away from the consumption of grains such as maize. For Malawi, with a rising population and where the majority of the people are food insecure, any improvements in welfare over time leads to an increase in maize consumption, as consumers use any additional income to meet their subsistence food requirements. This is mainly the case for the rural population who form the majority of the country's population. As such, the effect of the trend variable on maize consumption in Malawi may not be in line with the existing empirical evidence. As mentioned earlier, incomes are also known to affect the demand for a commodity. In Malawi, smallholder farmers are both producers and consumers of maize. This implies that household incomes may not impact upon consumption as stipulated by economic theory. This is because an increase in the income of rural households may be the result of a rise in the amount of a crop that the household markets which may be done at the expense of home consumption. Hence in the case of Malawi, the duality of smallholder farmers as both producers and consumers of maize imply that maize consumption is confounded both by supply-side and demand-side dynamics.

The trade block consists of imports and exports with imports being estimated as a function of net exports and the summation of imports and net exports providing exports. This treatment of the trade block makes trade explicit but exogenous to the model. In addition, imports are also determined by a government's trade policy and parity prices which take into account the costs associated with importing a good from one country to another as well as exchange rates. In Malawi, maize trade is controlled by government, with public policies dictating the amounts of maize to be exported or imported. The trade block is closely linked to the price block as domestic prices are influenced by a country's trade policies (Meyer, 2006) as well as other government policies and the domestic demand and supply dynamics. In Malawi, maize prices are mainly determined by government policy, as prices have been under some form of control since Malawi gained independence. Despite this, regional and international maize prices have some influence on the domestic maize price as maize is often imported to meet domestic food shortages. This implies that regional and international prices have an effect on domestic pricing.

3.3.2 Equilibrium in an imperfect market

As stated earlier partial equilibrium analysis is the determination of equilibrium in a single market. According to economic theory, market equilibrium is achieved by the stabilizing effect of a competitively set price. In Malawi however, the government sets floor and selling prices for maize. Maize price interventions are not new to Malawi as the government has always controlled the marketing and trade of maize. Various policy options with direct price interventions have been part of the Malawi maize market since independence (Chirwa, 2009). After independence maize pricing was based on a pan-territorial and pan-seasonal parity price regime which was shifted to a price band in 1996. This price band which was implemented through ADMARC remained in place until 2000 (Chirwa, 2005). In the 2007/08 season, another maize price band was put in place in an attempt to curb maize price swings. The price band restricted maize trade to a range of MK45 to MK52 per kilogram (USD150 to USD163 per ton in real terms).

In such a market, there is a reduction in output because price controls lead to lower incentives for producers. Price controls change the way in which equilibrium is reached, as they lead to consumers competing for the goods whose price is controlled. This leads to consumers incurring higher transactions costs associated with searching and queuing (Devarajan, *et al.* 1989). In a controlled market, the summation of the demand for a good from the parallel and official markets is equal to the demand in a control-free market. However, supply in the controlled market is less than the supply in a control-free market. Hence, equilibrium is reached when the transaction costs associated with accessing the price-controlled goods equals the difference between the price of the commodity on the parallel market and the controlled price (Devarajan, *et al.* 1989). The use of partial equilibrium models in such a market is, however, still possible, but it requires the making of assumptions pertaining to consumers and producers of the commodity under analysis. These assumptions are critical as they affect the market outcomes and, in a partial equilibrium framework, they determine the nature in which supply, demand and price are determined.

On the consumer side, there are two alternative assumptions pertaining to consumers' access to the good that is under price control. These are the assumption of costless access/rationing and the assumption of endogenous transaction costs (Devarajan, *et al.* 1989). Implicit in both assumptions is that a parallel market will emerge as a result of government implementing price controls. This is the case for Malawi, as evidence exists which shows that after the implementation of a price band for maize, 70% of all private markets continued to sell maize at prices that were above the government set prices (Chirwa, 2009).

The assumption of costless access/rationing implies that households that purchase the good that is controlled are able to access it without incurring any additional costs associated with searching and queuing. This implies that these households are able to buy the good either in the official or unofficial markets at the official prices or at prices that are above the control prices respectively. For simplicity, this assumption assumes that consumers buy from the unofficial market but are not fined for doing so. Costless rationing essentially increases consumer incomes and, as such, can lead to increased demand and therefore production if the good is a normal good; or to lower demand and therefore reduced production if the good is an inferior good (Devarajan, *et al.* 1989). The assumption of costless rationing would apply in cases where the government provides rationed access of the price-controlled goods to consumers who are legally entitled to access it. For the Malawi maize sector, this assumption would not hold, as government does not ration the amount of maize that consumers can access, despite the implementation of price controls.

The alternative assumption of endogenous transaction costs assumes that consumers of a good compete to access the good that is under price control. This entails that they incur higher transaction costs associated with searching and queuing. Equilibrium in such a market is reached when consumers become indifferent to buying the good in the official market and the parallel markets. Consumer indifference only occurs when the transaction costs that consumers incur equal the difference between the price of the good in the official market and the price in the parallel market (Nguyeni *et al.* (1989) cited by Devarajan, *et al.* 1989). The modelling of the Malawi maize market assumed endogenous transaction costs in both the national and local maize markets. This is because this assumption holds in cases where the government sets the control price below the free market equilibrium price (Devarajan, *et al.* 1989). This is the case for Malawi, as studies conducted by Jayne *et al*. (2008) showed that the ceiling price for maize trade for 2008 set by government was 15.3 % lower than what maize producers had anticipated.

Given the endogenous transaction cost assumption, the following has been taken on board in modelling the Malawi maize market in a partial equilibrium framework:

- *Maize demand or consumption at either the national or local level remains unaffected.* This is because economic theory has demonstrated that a combination of the good demanded in a price-controlled market and the parallel market equals the demand in a control-free market. Thus, demand for a good does not change with price control under the assumption of endogenous transaction costs.
- *Maize supply in both the national and local maize markets is lower*. This is because in the face of price controls, producers market their output in both the official and parallel market. This is based on empirical evidence which shows that in markets with price controls, producers do not completely abandon the official market, even if the controlled price is below the free market price (Devarajan, *et al.* 1989). In such cases, the market outcome is that the aggregate amounts of the price-controlled good supplied is always less than the amount of the good supplied in the market that is free of price controls. Therefore, for the maize market for Malawi, this was taken into account in forecasting maize production at both the national and the local economy level.

3.4 MACRO-MICRO LINKAGES AND MODEL CONSISTENCY

The macro-economic partial equilibrium maize model and the household-level income data were linked in a one-way top-down fashion using micro-accounting techniques. One-way top-down linkages are those in which macro-economic level effects are fed into the household level, thus creating unidirectional links from the macro-component to the micro-component of the model. The micro-accounting linkage method is a non-parametric arithmetic approach that assumes a stable within household group distribution and employment structure. Households are classified into different categories and following a macro-economic policy shock, growth rates in incomes or other variables such as per capita consumption are obtained from the macro-component. Household group specific growth rates are then applied separately to each household category, thus providing the post-shock level of income or any other variable that is under consideration.

The micro-accounting method was selected as it is computationally suitable for the household data that was available at the primary level in this study. In addition, micro-accounting methods are advantageous as they are relatively straight forward to use and hence are highly suitable for developing country analysis in which there is often the lack of financial resources and data (Agénor, *et al.* 2005). In addition micro-accounting techniques are capable of capturing the heterogeneity in households (Lay, 2006).

The micro-accounting method has two main short falls. First, the method does not completely take into account the differences amongst individual households within each household category. This is because the method applies category-specific growth rates instead of household level ones; thus assuming unrealistically that the intra-category distribution of income remains constant after a shock. Second, there is the assumption that macro-economic shocks will not change the initial sector of activity in which households are involved (Agénor, *et al.* 2005). Despite this, micro-accounting methods still remain relevant although the policy shocks to be simulated have the potential to alter the employment structure of the rural economy in Malawi. The applicability of the results are however relevant only in the short to medium; and not in the long term.

In linking models at different levels, it has to be assured that the aggregate information in the macro-component of the model is disaggregated to the micro-level in a manner which is consistent (Peichl, 2009). Inconsistency between the two components can occur due to either data measurement errors in each component or due to the difficulties of deriving theoretical concepts at the micro-level (Kavonius, 2010). The use of the top-down linkage approach, in which price changes from the macro-component are translated into income changes at the household level, removes the concerns pertaining to inconsistencies between the macro and micro components of the macro-micro model. This is because the top-down linkage approach has the advantage that modelling of the macro and micro components is done separately with changes in equilibrium prices providing a link between the two components. In such cases, there is no need to reconcile data of the macro and micro components, as the two components are separately consistent (Vos, *et al.* 2004:13).

3.5 CHAPTER SUMMARY

This chapter has provided an in depth description of the theoretical foundations on which the analytical framework of this study is based. The key emphasis of the chapter was on the theoretical foundations of the micro-component of the analysis. This was followed by a brief overview of the macro-economic partial equilibrium maize model that will be used for simulating policy shocks in Chapter 6, the technical details of which will be provided in greater detail in Chapter 5.

CHAPTER 4

QUANTIFYING THE IMPACT OF AGRICULTURAL INNOVATION SYSTEMS ON RURAL LIVELIHOODS

4.1 INTRODUCTION

The purpose of this chapter is to provide evidence of the impact of AIS driven research interventions on rural livelihoods. In so doing, the chapter demonstrates that the paradigm shift in agricultural research from a top-down linear approach to a holistic approach driven by innovation systems concepts has contributed towards changing the rural household economy by creating greater linkages between rural households and the market economy. The first part of the chapter provides a descriptive summary of some demographic and socio-economic characteristics of the sampled households to better understand the community under study. The second part presents the results of a logistic regression model of participation in the ERI intervention and the results of the single differencing analysis to determine the impact of AIS driven research interventions on rural livelihoods. The last part of this chapter presents a description of the different household typologies found in the study area and the description of their income portfolios. The results of the last section will be used in Chapter 6 to test the hypothesis that macro-economic policy shocks affect rural households differently as a result of differences in their income portfolios and household typology.

4.2 METHODOLOGY

4.2.1 Place of study and data collection

The study was conducted in the Ukwe Extension Planning Area (EPA) in Lilongwe District in the Central Region of Malawi. Households were sampled from Katundulu, Mphamba and Kango villages. Katundulu and Mphamba villages formed the intervention communities where the Enabling Rural Innovation (ERI) was piloted (as described in Chapter 2); while Kango village was the area where a counterfactual was established. Purposive random sampling was used to select study participants from the intervention community while simple random sampling was used to select study participations from the counterfactual community. A semi-structured questionnaire was used to sample a total of 303 households in the study area, with the counterfactual community sample size being double that of the intervention community sample size in order to allow for better matching of households (Ravallion, 2003). Households from the counterfactual community that did not match with those in the intervention community in terms of pre-existing observable social economic and farming systems characteristics were dropped from the analysis in order to reduce bias and to increase robustness.

4.2.2 Data analysis

This section provides a description of the analytical tools employed to meet the objective of this chapter. This includes a discussion of quasi-experimentation; propensity score matching and logistic regression modelling that have been used in the study to establish a valid counterfactual to overcome sample selectivity bias and to overcome the problems associated with attributing changes in observable livelihood outcomes to specific interventions.

4.2.2.1 Impact evaluation of livelihood outcomes

The choice of evaluation technique in micro-economic impact studies depends on the nature of the question to be answered, the available data and the way in which the participants were selected for the programme (Blundell & Costa, 2007 cited by Bourguignon, *et al.* 2008). In studies where the intervention has already occurred, the evaluation technique has to be one that is able to compare the outcomes of those that were part of the programme and those that were not part of the programme (counterfactual), but who are otherwise similar to the programme participants; and in so doing, establishing the effects or impact of the programme.

Hence, the key to a good impact evaluation is the estimation of what would have occurred in the absence of the intervention (Martinez, 2009). Since impact evaluations are carried out after the programme has started or finished, as is the case in this study, *ex-post* changes in outcome variables are used as a measure of impact. The problem with this is that there are many other observable and non-observable time variant characteristics which may alter outcome variables for participants. As such, it becomes difficult to attribute changes in the outcome variables to a specific intervention. This is because comparison of the before and after changes in the outcome variable can lead to either over or under estimation of programme impacts. To overcome this problem, commonly called the attribution problem, it therefore becomes necessary to use data on outcome variables from the counterfactual. A valid counterfactual must have very similar observable pre-intervention characteristics to the participants with the only difference being programme participation.

The availability of data from non-participants is, however, in itself also insufficient for attributing differences in outcome variables to a programme, as changes in the outcome variables for participants may also arise from "selection bias" in that participants may have been purposefully selected (Ravallion, 2003; Ravallion, 2005). This entails that those non-participants who are used for comparison purposes must, in addition to having near-identical pre-intervention characteristics, be those who would have had an equal chance of being selected for participation in the intervention, hence overcoming selectivity bias. In the absence of randomisation, which equalises the probability of participation in an intervention thus removing selection bias, matching techniques, specifically Propensity Score Matching (PSM), becomes the solution to the establishment of a valid counterfactual (Baker, 2000; Ravallion, 2003).

According to Ravallion (2003), the underlying concepts of PSM are that two groups are identified, one that took part in the intervention denoted $H_i = 1$ for household *i* and another that did not participate in the intervention demonstrated $H_i = 0$. Intervention households are matched to non-intervention households on the basis of the probability that the non-participants would have participated in the intervention and this probability is called the propensity score. It is given mathematically as follows:

$$P(X_i) = Prob(H_i = 1 | X_i) \quad (0< P(X_i) < 1) \qquad (4.1)$$

Where

X_i is a vector of pre-intervention control variables

These pre-intervention control variables are those which are based on knowledge of the programme under evaluation and on the social, economic and institutional theories that may influence participation in the intervention. The vector can also include the pre-intervention values of the outcome variables. Propensity score matching is not able to reproduce the results of randomisation if the variables that influence participation in the intervention are not properly defined.

PSM is driven by two main assumptions:

- The H_i's are independent over all i's
- The assumption of "conditional independence" or "strong ignorability" which says that outcomes are independent of participation given the variables that determine participation (X_i). In addition, outcomes are also independent of participation given $P(X_i)$ as they would be in a randomised experiment.

PSM equalises the probability of participation across the population just as in randomisation. However, the difference is that PSM achieves this based on conditional probabilities which are conditional on the variables determining participation (X_i).

In this study, propensity scores for each household in the sample were estimated using logistic regression modelling. Using the estimated propensity scores, matched pairs of households were established on the basis of the proximity of propensity scores of the probability of participation in the ERI initiative between the intervention and counterfactual samples. Unmatched counterfactual households were dropped from the analysis in order to remove bias and to increase robustness (Rubin & Thomas, 2000 in Ravallion, 2003). The best matched or "nearest neighbour" to the j^{th} intervention household is the counterfactual household that minimises $[P(X) – P(X_j)]^2$ over all j's in the set of counterfactual households.

A typical PSM estimator of the average impact of any intervention takes the following form (Ravallion, 2003):

$$\Delta\overline{Y} = \sum_{j=1}^{T} \omega_j (Y_{j1} - \sum_{i=1}^{C} W_{ij} Y_{ij0}) \qquad (4.2)$$

Where:

Y_{j1} is the post intervention outcome variable for the j^{th} household in the intervention

Y_{ij0} is the outcome indicator of the i^{th} counterfactual household matched to the j^{th} intervention household

T is the total number of interventions/treatments

C is the total number of counterfactual households sampled

W_{ij} 's are the weights applied in calculating the average outcomes of the matched counterfactual households

ω_j are the sampling weights used to construct the mean impact estimator

To avoid contamination by endogeneity of access to the ERI program, the regression model for ERI program participation (which was estimated to generate PSM scores) was run only for the matched comparison group. Hence the estimator in such cases becomes as follows:

$$\Delta\overline{Y} = \sum_{j=1}^{T} \omega_j \left[\left(Y_{ij} - X_j \hat{\beta}_0 \right) - \sum_{i=1}^{C} W_{ij} \left(Y_{ij} - X_i \hat{\beta}_0 \right) \right] \qquad (4.3)$$

Where $\hat{\beta}_0$ is the Ordinary Least Squares (OLS) estimate for the counterfactual matched group.

The impact estimator is approximated without any arbitrary assumptions about functional forms and error distributions, as PSM does not require a parametric model linking programme participants to outcomes (Ravallion, 2003). This makes PSM superior to non-experimental regression-based approaches.

4.2.2.2 ERI programme participation model

A logistic regression model of participation in the ERI initiative was estimated in order to determine the probability of a household participating in the intervention by generating propensity scores. Participation in ERI was therefore modelled as a dichotomous dependant variable determined by a set of exogenous variables that were crucial for determining participation in the ERI intervention as determined at the onset of the program.

In this case, the innovation platform which was established in Ukwe Extension Planning Area worked together to select an appropriate community within the EPA for piloting the ERI initiative based on a criteria that included all year-round road accessibility; availability of a motivated local-level extension agent; willingness of other development partners working in the community to take an active role in the initiative; and the existence of interest in the community for further agricultural research and development (Sangole, *et al.* 2003). In addition to this, individual households within a selected community also had the opportunity to decide whether they wanted to participate or not.

Given these considerations, the exogenous variables that were included in the model for ERI program participation included the frequency of contact with extension agents prior to the ERI initiative, the sex of the household head, and level of participation in other development initiatives prior to the ERI initiative as well as the size of the household. The functional form of the model is given below:

$$PART = f\left(EXT\_CONT, SEX\_HHEAD, PRE\_PART, HHSIZE\right) \quad (4.4)$$

Where:

PART	Participation in ERI (0 = Non-participant, 1= Participant)
EXT_CONT	Frequency of contact with extension agents
SEX_HHEAD	Sex of the household head (0= Female, 1= Male)
PRE_PART[4]	Index of previous participation of the household in other development initiatives (0=Low, 1=Intermediate, 2= High)
HHSIZE	Size of the household (people eating from the same harvest)

A major assumption of logistic regression modelling is that the data has a binominal distribution taking the following form (Gujurati, 1992):

$$Y_i \sim B(n_i, p_i) \text{ for } i = 1\ldots\ldots n \quad (4.5)$$

Where

Y_i is participation in the intervention

p_i is the unknown probability of participation

n_i are the observable outcomes of participation for each household

[4] This is a categorical variable whose results will be presented separately for the different types of previous participation level in order to show that the odds of participating in ERI may differ depending on the level of previous participation. This is a common and acceptable way of reporting categorical variables in logistic regression results. See http://128.97.141.26/stat/stata/webbooks/logistic/chapter2/default.htm

According to Gujarati (1992), the logistic regression model assumes that there is a set of explanatory variables that can inform the final probability of participation. Because of this assumption, the explanatory variables can be thought to be in a *k* vector X_i. If we model the natural log of the odds of the unknown binomial probability p_i as a linear function of the X_i's, we get the following:

$$In\left(\frac{p_i}{1-p_i}\right) = \mathrm{B}_0 + \mathrm{B}_1 x_i + \mu_i; i = 1 \ldots n \qquad (4.6)$$

Where B_0 and B_i are the intercept and the unknown parameters respectively.

4.3 DESCRIPTIVE ANALYSIS OF SAMPLED HOUSEHOLDS

This section presents the empirical findings of the impact evaluation of agricultural research interventions that are driven by innovation systems on rural livelihoods. It, however, starts by presenting a descriptive analysis of the sampled households; in order to provide greater contextual understanding. An independent samples t-test was carried out to statistically compare the difference in means for various socio-economic characteristics between households in the intervention community and counterfactual community.

4.3.1 Household characteristics

Table 4.1 shows that the majority of the total respondents interviewed were women, with 67.3 % of all respondents being female while only 32.3 % of the respondents were male. The counterfactual community had more female respondents (77.7 %) as compared to the intervention community (46.5 %). Despite this, the majority of households in the study area were male headed, with about 82.2 % of the total sample being male headed and only 17.8 % of the households being female headed. These findings are in line with national demographics which show that nearly 75 % of all households in Malawi are male headed (NSO, 2008a). In addition, it can be seen that the majority of female-headed households were found in the counterfactual community (20.3 %), while the intervention community had fewer households that were headed by females (12.9 %).

Table 4.1: Summary of household respondents and headship

	Total sampled households[5]		Intervention community		Counterfactual community	
Percentage and number of respondents interviewed						
	N	%	N	%	N	%
Female	204	67.3	47	46.5	157	77.7
Male	99	32.3	54	52.5	45	22.3
Total	303	100.0	101	100.0	202	100.0
Headship of household: Percentage and number						
	N	%	N	%	N	%
Female	54	17.8	12	12.9	41	20.3
Male	249	82.2	88	87.1	161	79.7
Total	303	100.0	101	100.0	202	100

N = number of respondents % = percentage

The average age of the household heads in the study area was about 40 years, while the average age for spouses was 35 years of age (Table 4.2). The independent samples t-test showed that there was a significant difference between the ages of spouses in the intervention community and counterfactual community, with the spouses in the intervention community being significantly older. Table 4.2 further shows that the average household size in the study area was 4.8 people. The size is similar to the national household size of 4.4 people per household (NSO, 2008b). The t-test further indicated that there are statistically significant differences in the marital status of sampled households in the two communities, with more sampled households in the intervention community (87.1 %) being in legally-binding and socially-acceptable marriages than households in the counterfactual (78.7 %). This may explain the larger number of female-headed households in the counterfactual community. In addition, the counterfactual community had a significantly higher number of households that were in polygamous marriages (14.4 %) as compared to the intervention community (10.1 %); and this difference was statistically significant at the 10 % confidence level.

Table 4.2: Household characteristics

Household characteristics	Intervention community	Counterfactual community	t-value
Average age of household head	40.93	39.10	0.004
Average age of spouse	34.85	31.57	2.92*
Average household size	5.02	4.54	1.27
% of households that are married	87.1	78.7	12.15*
% of households in polygamous marriages	10.9	14.4	13.54*

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

An analysis of the level of community engagement and leadership (Table 4.3) showed that more male household members had active membership in farmer groups as compared to

[5] For the socio-economic characterisation, any discussion referring to the sampled households or unmatched households, refers to the total 303 households that were interviewed. For the impact assessment, households in the intervention community were matched to the households in the counterfactual community and these are referred to as the matched sampled size/matched households.

female household members. This can be attributed to Malawian male household members participating more in community activities as compared to females, whose participation is limited by both cultural and social factors (Care Malawi, 2010).Table 4.3 further shows that more household heads (26.7 %) and spouses (6.3 %) in the intervention community had membership in farmer groups as compared to household heads (16.3 %) and spouses (2.8 %) in the counterfactual community. These differences in group membership were f statistically significant at the 10 % confidence level, implying that there was more membership in farmer groups in the intervention community than in the counterfactual community.

Table 4.3: Community engagement and leadership of sampled households

Community engagement and leadership	Intervention community	Counterfactual community	t-value
% of HH heads – membership in farmer groups	26.7	16.3	11.77*
% of spouses – membership in farmer groups	6.3	2.8	14.72*
% of HH – membership of more than one farmer group	5.9	2.5	9.29
% of household head with leadership position	40.6	18.4	59.89***
% of spouses with leadership position	18.4	10.2	0.621

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

Table 4.3 further shows that there were highly statistically significant differences between the numbers of sampled household heads who held leadership positions in the intervention community (40.6 %) and those in the counterfactual community (18.4 %). Leadership positions included traditional posts, community positions and positions arising in farmers' organisations. The difference in leadership position between the two communities arises from the leadership positions that individuals in the intervention community held in farmer groups' and not those in traditional or community engagements. This is because, firstly, the traditional and community organisation of both communities are similar due to their geographical proximity and both are predominantly of the Chewa tribe. As such, the number of available traditional and community leadership positions is more or less the same. Second, as discussed above, more households in the intervention community were involved in farmer groups than those in the counterfactual community. As such, there was a higher probability of an individual in the intervention community having a leadership position as compared to an individual in the counterfactual community, due to the former being more engaged in farmer groups.

4.3.2 Human capital characteristics

An analysis of the human capital characteristics demonstrates that there are more household heads in the intervention community with some level of formal education (84.2 %) than in the counterfactual community (74.5 %). This difference is statistically significant at the 5 % confidence level (Table 4.4). In terms of informal training, Table 4.4 further indicates that on average households in the intervention community had more training than households in the counterfactual community. Intervention community households had on average 2.64 and 1.92 trainings per year five years ago and in the 2007/2008 cropping season, respectively; while counterfactual community households had on average 0.93 and 0.90 trainings per year five years ago and in the 2007/2008 cropping season, respectively. The differences in informal training for the 2007/2008 cropping season and five years ago are statistically significant at the 10 % and 5 % confidence levels, respectively.

Table 4.4: Human capital characteristics

Human capital characteristics	Intervention community	Counterfactual community	t-value
Average number of training in 2008/2009	1.41	0.90	4.71
Average number of training in 2007/2008	1.92	0.90	10.56*
Average number of training five years ago	2.64	0.93	18.63**
% of HH head with some formal education	84.2	74.5	3.28**
% of spouses with some formal education	74.4	76.3	0.629
Average distance from extension office in km	7.5	20.0	810***
% of HH with contact with extension at least once a year	52.6	65.4	1.03*

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

For the 2008/2009 cropping season, the differences in trainings between households in the intervention community and counterfactual community are statistically non-significant. From these findings, it can be seen that differences in informal training between the two communities are less distinct and nearly non-existent in the most recent cropping seasons. This can be attributed to the fact that during the implementation of the ERI initiative, there was emphasis on community capacity building through informal training. However, since it phased out in the 2006/2007 cropping season, it is obvious from these results that existing public agricultural extension agents in the intervention community have not been able to maintain the levels of capacity building.

In contrast to these findings, more households in the counterfactual community (65.4 %) stated that they had frequent contact with a public extension agent as compared to households in the intervention (52.6 %). This difference was statistically significant at the 10 % confidence level. This was the case despite counterfactual community households being on average significantly further from the extension officers' houses and offices (20.0 km) than the intervention community households (7.5 km). Informal interviews with counterfactual community households revealed that the contact with the extension officer in the counterfactual community was more unplanned, and tended to occur on occasions in which the extension officer had to pass through several villages in the counterfactual community to get to the main tarmac road on his way to the city. Community members took such opportunities to gather information and relay their problems to the extension agents.

4.3.3 Farming characteristics

All households in the study area are in a maize-based farming system, in which maize is the predominant staple food crop cultivated in combination with different legumes and cash crops. Other food crops grown in the study area with maize included groundnuts (*Apios americana*), beans (*Phaseolus vulgaris*), sweet potatoes (*Ipomoea batatas*), soy beans (*Glycine max*) and cow peas (*Vigna unguiculata*). Tobacco was the only non-edible cash crop cultivated in the area, while other cash crops that are widely cultivated in other parts of the country, such as cotton and paprika, were not found. Apart from crop cultivation, it is the practice of households to rear livestock and the majority of sampled households (58.3 %) owned livestock. Livestock that was readily found in the study area included pigs, chickens, goats, oxen, donkeys, ducks, guinea fowls, turkey, sheep, dairy cattle, rabbits, and hamsters.

From Table 4.5, it can be seen that the average land holding sizes for households in the intervention community and counterfactual community were 1.72 hectares and 1.23 hectares respectively. These differences in land holding sizes, which are relatively similar to the average national household land holding size of 1.5 hectares (World Bank, 2009), were statistically non-significant. Despite this, further analysis shows that on average, households in the intervention had more separate pieces of land (3.1) and there were more households in the intervention owing a wetland for winter cultivation (94.1 %) as opposed to the counterfactual community, where households had fewer numbers of separate pieces of land (2.1) and fewer households owning a wetland (47.5 %). Both differences were highly statistically significant.

Differences in wetland[6] ownership are because the majority of households in the counterfactual community had sold their wetlands to semi-commercial urban farmers, who flock to the community seeking both wetland and arable upland for cultivation due to its proximity to a tarred road (about 4.5 km). This was not the case for the intervention community as the area was relatively far from a tarred road (approximately 29 km).

Table 4.5: Farming characteristics

Farming characteristics	Intervention community	Counterfactual community	t-value
Average number of separate farm plots	3.1	2.1	1.178***
Average land holding size (ha)	1.72	1.23	0.00
% of households with a wetland	94.1	47.5	673.76***
% of household owning livestock	85.0	45.0	23.84***
% of households hiring additional labour	39.6	19.8	41.09***
% of households receiving remittances	13.9	10.4	3.08
% of households who have access to credit	12.9	6.9	11.56*

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

Analysis of the differences in inorganic fertilizer application between the two communities showed that the majority of households in the intervention community applied inorganic fertilisers on their farms as compared to the households in the counterfactual community. As can be seen in Table 4.6, the differences between households using and not using inorganic fertilisers in the intervention community and the counterfactual community are highly statistically significant for the 2004/2005, 2005/2006, 2006/2007 and 2007/2008 agricultural seasons, but less significant for the 2008/2009 agricultural season. For the intervention community, at least 60 % of all households were using inorganic fertilisers on an annual basis since the 2004/2005 season. In the counterfactual community, it is only in the 2008/2009 agricultural season that the percentage of households using inorganic fertilisers exceeded 60 %. One main reason for the differences in the fertiliser use between the two communities is that the ERI approach encouraged farmers to reinvest in their farm enterprise as part of sustaining their agro-enterprises. Hence, these differences in fertiliser use levels could be an indication of farmers reinvesting in their farms as a result of changes in their decision-making patterns arising from participation in the ERI.

Sources of inorganic fertiliser between the two communities were, however, similar. The main sources from which inorganic fertiliser was obtained were from subsidised coupons, purchasing at the full market price, gifts from relatives, donations from Non-Governmental

[6] Wetlands; which are locally known as *dambos;* are any permanent or seasonal wet land area that are mainly found along rivers and steams and which are populated by herbaceous plants and vegetation. Wetlands are mainly cultivated in Malawi during the short dry winter season (May–July)

Organisations (NGOs), or from farmer groups. Fertiliser received from NGOs was mainly from the Malawi Rural Finance Company, which is a government-operated loan facility that provides fertiliser on a loan basis to rural smallholders.

Table 4.6: Fertiliser use patterns

Fertiliser use patterns	Intervention community	Counterfactual community	t-value
% of household using fertiliser in 2004/2005	59.0	29.2	12.37***
% of household using fertiliser in 2005/2006	68.0	35.1	2.33***
% of household using fertiliser in 2006/2007	72.0	41.1	24.08***
% of household using fertiliser in 2007/2008	80.0	54.0	103.32***
% of household using fertiliser in 2008/2009	83.0	65.8	52.65**

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

It can further be seen from Table 4.6 that in the intervention community and the counterfactual community, the percentage of households utilising inorganic fertiliser has increased each subsequent cropping season since the 2004/2005 cropping season. This can be attributed to the implementation of the fertiliser subsidy programme which made inorganic fertilisers more readily available and accessible in rural areas. This was the case despite only a minority of the respondents in either the intervention community or the counterfactual community having ever received subsidised fertiliser coupons. This is because although not all households received coupons for subsidised fertiliser, farmers often pooled money together to buy inorganic fertiliser at the subsidised price. As such, even farmers who had never received coupons for subsidised fertiliser had in fact accessed it.

Furthermore, the fertiliser subsidy programme had created a parallel market of inorganic fertiliser in the rural area, in that some rural entrepreneurs had started to purchase the coupons for subsidised fertiliser from rural beneficiaries who were not able to find the cash to purchase the subsidised fertiliser. Using the coupons bought, they would purchase the subsidised fertiliser and subsequently sell it in smaller portions to non-beneficiaries or beneficiaries who had not meet their fertiliser needs. The unit price of fertiliser on the parallel market was higher than the subsidised fertiliser price but much lower than the actual market price for the unsubsidised fertiliser. The parallel market was also one of the means through which inorganic fertiliser accessibility and availability had increased in the study area.

Apart from inorganic fertiliser, the results further indicated that the majority of households in either the intervention community (64.0 %) or the counterfactual community (62.4 %) did not incur any other input costs. This was the case, as farmers did not use herbicides or pesticides, and the majority stated that they often recycled their seed. Seed recycling is a common practice in Malawi, in spite of farmers being aware that it can potentially reduce agricultural productivity and efficiency (Lanteri & Quagliotti, 1997). Seed recycling is common practice due to three main reasons- cash constraints, frequent changes in government support policies and the distribution of free seed by both public and private institutions (Smale & Phiri, 1998). Cash constraints by producers prevent producers from purchasing sufficient seed to meet their needs. Changes in government support policies and the distribution of free seed hinders producers' abilities to develop and sustain independent strategies for sourcing seed in each new season.

4.4 IMPACT OF AIS DRIVEN RESEARCH

This section presents the results of the single differencing of livelihood outcomes carried out between intervention households and matched counterfactual households. It starts by presenting a validation of the ERI program participation model used to generate propensity scores for determining a valid counterfactual. This is then followed by empirical results of the impact evaluation.

4.4.1 Generation of propensity scores

The logistic regression model of ERI participation that was estimated to generate propensity scores was a good predictor of participation as demonstrated by the results of two alternative tests of goodness of model fit, the Hosmer and Lemeshow (H-L) statistic and the chi-square test (Table 4.7). The H-L goodness of fit test statistic was 10.310 and non-significant (p=0.244), indicating that the model was a good fit. A rule of thumb for accepting a logistic regression model is that the H-L statistic must be greater than 0.05 and should show non-significance (Hosmer & Lemeshow, 1989). Further, the model has a chi-square statistic of 23.747 which was statistically significant at the 1 % confidence level. This implied that all the predicators that were included in the model were capable of jointly predicting participation in the ERI initiative. As such the model is a good determinant of the factors influencing participation in the ERI initiative. This therefore implies that the model is capable of approximating the probability of a household's participation in the ERI initiative and it is capable of correctly generating propensity scores.

Therefore, using propensity scores for participation generated from the logistic regression model, households in the intervention community were matched to households in the counterfactual community on the basis of the proximity of propensity scores. All other households whose propensity scores for participation were far from the range of scores for the intervention households were dropped from the analysis.

Table 4.7: Parameter estimates of the logistic model of ERI participation

	Co-efficient	Significance	Odds ratio
Constant	-1.426	0.136	0.240
PRE_EXT_CONT	-0.157	0.019	0.855
SEX_HHEAD	0.571	0.118	1.771
HH_SIZE	0.092	0.087	1.092
PRE_PART_PREVIOUS[7]		0.003	
PRE_PART_PREVIOUS(1)	-0.559	0.310	0.572
PRE_PART_PREVIOUS(2)	1.486	0.068	4.419
Model Chi square	*23.747****	*Log-Likelihood*	*361.980*
H-L Chi square	*10.310 (p=0.244)*	*Nagelkerke R-square*	*1.205*
N=303			

By dropping all the counterfactual community households whose probability of participation was very far from the households in the intervention community, differences in livelihood outcomes were then compared between households that were more similar and therefore comparable. This can be seen by looking at the differences in the socio-economic characteristics of the matched households (Table 4.8), which are less distinct than differences between households in the entire sample, as discussed above in Tables 4.3, 4.4, 4.5, and 4.6. Some differences in the characteristics of households that were either statistically significant or very significant in the unmatched data set were no longer statistically significant for the matched data sets. Despite greater similarities between the two groups in the matched data set, some differences still exist in terms of the farming characteristics and fertiliser use patterns; which are to be expected, as rural households are not homogenous in nature.

[7] Previous participation is interpreted for the high previous participants and the intermediate/low previous participators separately. This presentation of categorical variables in logistic regression models is common and accepted practice. These results show that those that had high previous participation (pre_previous(2)) in development projects had higher probability of also participating in the ERI. The odds of the intermediate/low (pre_previous (1)) previous participators are shown to be lower.

Table 4.8: Socio-economic characteristics for matched households

	Intervention community	Counterfactual community	t-value
Household characteristics	n = 100	n = 100	
Average age of household head	40.93	38.93	0.393
Average age of spouse	34.85	31.32	4.69*
Average household size	5.02	4.68	0.186
% of households that are married	87.1	80.2	6.36
% of households in polygamous marriages	10.9	17.0	4.28
% of respondents with other occupation apart from farming	24.8	19.8	2.87
% of household heads that are members of farmer groups	26.7	27.7	1.100
% of spouses that are members of farmer groups	6.3	4.5	3.86
% of households that are members of more than one farmer group	5.9	4.0	1.683
% of household head with leadership position	40.6	25.10	23.43*
% of spouses with leadership position	18.4	14.9	0.957
Farming characteristics			
Average number of plots	3.1	2.2	0.217***
Average land holding size (ha)	1.72	1.23	0.218**
% of households with a wetland	94.1	56.4	304.1***
% of household owning livestock	85.0	42.6	15.33**
% of households hiring additional labour	39.6	23.8	22.37*
% of households receiving remittances	13.9	10.9	1.641
% of households who have access to credit	12.9	7.9	5.412
Human capital characteristics			
Average number of trainings in 2008/2009	1.41	1.22	0.791
Average number of trainings in 2007/2008	1.92	1.26	2.61
Average number of training five years ago	2.64	1.02	9.98*
% of household head with some formal education	84.2	75.8	3.12*
% of spouses with some formal education	74.4	80.5	1.27
Average distance from extension office/house in km	7.5	20	403.7***
% of respondents who have contact with extension agent at least once a year	52.6	81.2	21.78***
Fertiliser use patterns			
% of household using fertiliser in 2004/2005	59.0	31.7	6.97***
% of household using fertiliser in 2005/2006	68.0	36.6	4.065***
% of household using fertiliser in 2006/2007	72.0	45.5	21.00***
% of household using fertiliser in 2007/2008	80.0	58.4	42.60**
% of household using fertiliser in 2008/2009	83.0	69.3	21.95*

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

4.4.2 Differences in rural livelihood outcomes

This section presents the results of the single differencing of livelihood outcomes carried out between households in the intervention community and the matched households in the counterfactual community in order to determine the impact of the ERI intervention. This study employed the single differencing method as it involves the *ex-post* comparison of outcome variables between program participants and non-program participants. Single differencing is valid when program and non-program participants have more or less similar outcomes at the onset of a program and also in the absence of baseline data. The main limitation of the single differencing technique is that it does not effectively control for pre-intervention differences in outcome variables between the program and non-program communities. Therefore in cases in which the pre-intervention outcome variables of participations and non-participations differ greatly; single differencing can produce results that are not valid. The use of Propensity Score Matching (PSM) to establish a statistically valid comparator group however significantly reduces biases associated with pre-intervention differences in outcome variables between participating and non-participating groups (Ravallion, 2003).

4.4.2.1 Impact on production outcomes

The ERI intervention impacted upon many aspects of household production with statistically significant differences being observed for the outcomes pertaining to livestock production, upland crop production, value of maize production and asset ownership. Differences in maize yields were found to be unaffected by participation in the ERI intervention; implying that the ERI initiative did not have an impact on maize productivity. This finding is plausible because it is possible to increase production without necessarily increasing productivity (Beattie & Taylor, 1985). Hence in the study area it is possible that participating households managed to increase production of maize, not necessarily productivity, by increasing the area under maize cultivation.

An analysis of the value of all upland crops for the households in the study finds that the ERI intervention increased the value of crops produced for participating households by USD812.34 and USD627.10 for the 2007/2008 and 2008/2009 cropping seasons respectively (Table 4.9). The differences in the value of all upland crops were statistically significant at the 1 % and 5 % confidence levels for the 2007/2008 and 2008/2009 seasons respectively.

Table 4.9: Impact of the ERI intervention on production outcomes

Production outcomes (USD)	Intervention community		Counterfactual community		ERI program effect
	Mean	Standard deviation	Mean	Standard deviation	
Total value of livestock	445.03	1620.99	144.82	926.47	300.12*
Total value of upland crop production 2007/2008	1349.48	0.016	537.14	0.0114	812.34***
Total value of upland crop production 2008/2009	992.24	0.0179	365.14	0.0084	627.10**
Value of maize harvest 2007/2008	259.35	308.77	180.01	340.24	79.33
Value of maize harvest 2008/2009	506.76	0.013	219.66	490.80	287.09*
Maize yield 2007/2008 (Tons/hectare)	0.84	1.00	0.85	1.47	0.0055
Maize yield 2008/2009 (Tons/hectare)	1.17	2.61	0.88	1.33	0.287
Total value of assets	550.74	3008.51	159.65	581.58	391.00*

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

Furthermore, the results also show that in the 2008/2009 season, the value of maize produced by intervention community households was significantly higher than for households in the counterfactual community by USD287.09 at the 10 % level of confidence. This was, however, not the case for the 2007/2008 cropping season, as differences in maize production between the intervention community and counterfactual community households were not statistically significant. The difference in value of maize production in the 2008/2009 season cannot be attributed to higher maize prices for producers in the 2008/2009 season, which were at USD295.89 per ton as compared to prices in the 2007/2008 season of USD125.99 per ton; or to price differentials between the intervention community and counterfactual community. This is because an analysis of the farm gate prices for the different communities showed that all households in the study area received the same farm gate prices with no statistically significant differences being observed between maize prices in either community. Second, differences in maize production in the 2008/2009 season cannot be attributed to yield differences, as the results indicate that there were no statistically significant differences in the yields of maize between households in the intervention community and counterfactual community.

The significant differences in maize production between the intervention community and counterfactual community can therefore mainly be attributed to the fact that households in the intervention community cultivated more land than their counterparts in the counterfactual community. This is evidenced by statistically significant differences in the total land-holding sizes and the total number of separate farm plots that households in the intervention community owned and planted as compared to matched households in the counterfactual community. Households in the intervention community owed on average 1.72 hectares of land, while the matched counterfactual community households owed on average 1.23 hectares of land. In general households in both communities tend to cultivate the majority of the land that they own with approximately 95% of the land that is owned being under cultivation in either community in the 2008/09 cropping season. Households in the intervention community cultivated on average 3.1 separate pieces of farm plots; while matched counterfactual community households cultivated on average 2.2 separate pieces of farm plots. The differences in the land ownership and the separate pieces of cultivated farm plots between the intervention community and counterfactual community were statistically significant at the 5 % and 1 % confidence levels, respectively.

Analysis of differences in household assets and livestock ownership shows that the ERI initiative was significant in increasing the value of households' total assets and livestock ownership by USD391.00 and USD300.12, respectively. Both these differences were statistically significant at the 10 % level of confidence. Hence, households in the intervention community had higher valued assets than households in the counterfactual community. An analysis of the differences in livestock prices showed that there were very small differences between the market prices of the three major types of livestock traded in the study area; with the average price for the 2008/2009 season of chickens, pigs and goats not being statistically different between the two communities.

As such, it can be deduced that households in the intervention community had larger numbers of livestock as compared to households in the counterfactual community. This is confirmed by statistical analysis which showed that households in the intervention community owned an average of about 4 or more chickens and one extra pig and goat each as compared to households in the counterfactual community. The differences in the ownership of all three classes of livestock were highly statistically significant.

A major contributing factor to the larger livestock numbers in the intervention community, especially in terms of pig ownership, is that piggery was the agro-enterprise chosen to be developed under the ERI program. Because of this, participation in ERI entailed that households made more investments in the piggery through improved housing, feeding and hygiene; and improved their day-to-day management by keeping a record of all activities pertaining to the piggery. Participating households were trained in the construction of appropriate housing and feed formulation as well as in pest and disease control. In addition, farmer participatory research was put in place to test different feeding options and the cultivating of various types of feeds (Njuki, *et al.* 2007). These changes, together with greater market access arising from the establishment of a marketing committee in the community which was responsible for sourcing markets, led to the establishment of a stable market especially for piglets and this resulted in increased incomes. From informal interviews with participating households, it was revealed that this increased income, in combination with changes in the decision-making processes of participating households, enabled them to invest more in household assets as well as in other types of livestock.

4.4.2.2 Impact on household cash income

Analysis of the differences in household incomes indicates that the ERI intervention positively influenced cash incomes in both the 2007/2008 and 2008/2009 cropping seasons for the households who participated in the intervention. As can be seen in Table 4.10, households who participated in the ERI intervention had on average USD280.21 and USD340.54 more total income than their counterparts in the counterfactual community for the 2007/2008 and 2008/2009 cropping seasons respectively. The differences in household incomes were statistically significant at the 5 % confidence level.

Table 4.10: Impact of the ERI intervention on household cash incomes

Household cash incomes (USD)	Intervention community		Counterfactual community		ERI program effect
	Mean	Standard deviation	Mean	Standard deviation	
Total cash income from wetland 2008	27.24	87.27	23.67	66.98	3.56
Total cash income from wetland 2009	14.46	32.49	29.79	52.87	-15.39*
Total cash income for 2007/2008 season	511.49	0.0072	231.28	465.61	280.21**
Total cash income for 2008/2009 season	636.21	0.0088	299.56	655.75	340.54**
Total income from livestock sales 2008/2009	51.34	138.48	23.60	186.08	27.78

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

Increased cash incomes can be attributed to the ERI intervention focussing on assisting farmers to develop agro-enterprises in order to meet existing market opportunities as opposed to them marketing any surplus that they grew for subsistence. Hence, intervention communities conducted an analysis of existing market opportunities prior to the onset of the agricultural year in order to determine the type of agro-enterprises that would be most profitable. Through this analysis, the community identified piggery and dry bean cultivation as the most profitability agro-enterprises (Njuki, *et al.* 2007).

Furthermore the ERI intervention negatively affected incomes from the sale of crops cultivated in the wetland during the short dry season. Farmers in the intervention community had lower incomes from the wetlands as compared to households in the counterfactual community, with incomes in the intervention community being lower by USD15.39 in the 2009 winter season. This difference was statistically significant at the 10 % level of confidence. This was the case despite more households in the intervention community having wetlands as compared to households in the counterfactual community; with 94.1 % and 56.4 % of all sampled households in the intervention community and counterfactual community owning a wetland, respectively.

The winter season, which involves the cultivation of wetlands along rivers and streams, using residual moisture or irrigation, is a critical season for many rural producers in Malawi. For resource-poor farmers, cultivation during the winter season is mainly used for seed production. Hence, the majority of resource-poor smallholders who market a large portion of their harvest from the winter season are left without sufficient seed for the next rainy season; and are thus less prepared for the onset of the main cropping season. One key area of the ERI intervention was to build the capacity of households to better understand their farming systems and opportunities as well as threats to their livelihoods. Hence, the lower returns in sales from the wetland can be attributed to households in the intervention using the winter cultivation as an opportunity to retain seed for the main cropping season.

4.4.2.3 Impact on household training and group membership

An assessment of the ERI initiative's impact on membership of farmer groups and the number of trainings attended by a household was also carried out. Table 4.11 indicates that five years ago when ERI was in full implementation, households in the intervention community attended on average 1.62 more trainings than households in the counterfactual community. This difference was statistically significant at the 10 % level of confidence. These results are an indication that the ERI initiative provided participating communities with significantly more training opportunities than those that are provided by the local agricultural extension officers.

Table 4.11: Impact of the ERI intervention on trainings and group membership

Trainings and group membership	Intervention community		Counterfactual community		ERI program effect
	Mean	Standard deviation	Mean	Standard deviation	
Total number of farmer groups per HH	0.35	.0.865	0.49	0.074	-0.139
Average number of trainings five years ago	2.64	5.67	1.02	3.86	1.62*
Average number of trainings in 2007/2008	1.92	5.29	1.26	3.66	0.66
Average number of trainings in 2008/2009	1.14	4.43	1.22	3.55	-0.08

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

Further observations, however, showed that after the ERI initiative phased out in the 2006/2007 season, there were no statistically significant differences in the number of trainings attended by households in the intervention community and the counterfactual community for either the 2007/2008 or 2008/2009 agricultural seasons. This can be attributed to the phasing out of the ERI intervention that led to local agricultural extension officers reverting to pre-ERI training strategies in the intervention communities, which entailed less training. The results therefore indicate that the ERI initiative during its implementation had a positive impact, in that it increased the number of trainings that a household attended. This finding is logical as informal trainings and other capacity building activities were a major component of the ERI initiative. Hence, following its phasing out, the number of trainings for individuals in the participating communities and the counterfactual community were not statistically different.

Analysis of the differences in famer groups shows that the ERI initiative did not have a statistically significant impact on households' membership in farmer groups. This implies that households' participation in farmer groups between the two communities was not statistically different. This finding is surprising, as the ERI initiative worked towards establishing and strengthening farmer organisations, as it recognised that the most important success factor for increasing market access was well-established farmer organisations (Kaaria, *et al.* 2008).

4.4.2.4 Impact on fertiliser use patterns

The impact of the ERI intervention on fertiliser use patterns in the intervention community was assessed by analysing the differences in the number of 50 kg bags that farmers used per hectare of farm land. Inorganic fertilisers, in combination with hybrid seeds and good rainfall, play a crucial role in ensuring maize production and food security in Malawi. Hence, purchasing inorganic fertiliser demonstrates a household's decision-making patterns in terms of reinvestment in their farm enterprise. Table 4.12 shows that there were statistically significant differences between the amounts of inorganic fertiliser applied between the intervention community and counterfactual community households in the 2004/2005, 2005/2006 and 2006/2007 agricultural seasons at the 1%, 5% and 10% confidence levels respectively.

Table 4.12: Impact of the ERI intervention on fertiliser use patterns

Fertiliser use patterns (no. of 50 kg bags)	Intervention community		Counterfactual community		ERI program effect
	Mean	Standard deviation	Mean	Standard deviation	
2004/2005	1.24	1.85	0.567	1.55	0.679***
2005/2006	1.38	1.87	0.624	1.12	0.761**
2006/2007	1.50	1.88	0.858	1.38	0.644*
2007/2008	1.68	1.97	1.38	3.39	0.297
2008/2009	1.95	2.49	1.77	6.18	0.171

* Significant at 10 % level, ** Significant at 5 % level, *** Significant at 1 % level

Between the 2004/2005 and 2006/2007 cropping seasons, households in the intervention community applied on average nearly one extra 50 kg bag of inorganic fertiliser as compared to households in the counterfactual community. This difference can be attributed to the ERI intervention, as the increased market outcomes acted as incentives for households to reinvest in technology, such as inorganic fertiliser, in order to sustain their agro-enterprise. In general, the study finds that all households in both communities were applying an amount of fertiliser that was below the recommended rates for Lilongwe Agricultural Development Division (LADD) where the communities are located. The recommended fertiliser application rate for home consumption for LADD is the application of two bags of 23:21:0+4S and three bags of Urea which has 46 % nitrogen. For production for the market, the recommended application rate is one bag each of 23:21:0+4S and Urea (Benson, 1999). Table 4.12 further shows that it is only households in the intervention community that were close to reaching the recommendation for the market production, with a mean of 1.92 (50 kg) bags of inorganic fertiliser being applied per hectare for the 2008/2009 year.

Further observation shows that in more recent years, the differences in the amounts of inorganic fertiliser applied are less distinct. In the 2007/2008 and 2008/2009 cropping seasons, the differences are non-significant between the intervention community and counterfactual community. This can be attributed to the increased availability of fertiliser due to the implementation of the full fertiliser subsidy programme in the country. The implementation of the fertiliser subsidy programme increased the availability and accessibility of inorganic fertiliser throughout the rural areas of the country, hence increasing the opportunity for all farmers to access inorganic fertiliser. This was the case despite informal interviews with farmers revealing that initially when the fertiliser subsidy programme had started in the 2005/2006 season, farmers tended to sell their coupons for cash. They would then use part of the cash received to purchase a small amount of fertiliser (e.g. one 20 kg bucket) and to meet other household needs. In the intervention community, it was found that in subsequent years as farmers become more organised under the ERI programme, they refrained from the practice of selling their subsidised coupons. Hence, the increased fertiliser use in the intervention community can further be attributed to not only the implementation of the fertiliser subsidy programme but also to improved and better decision making on the part of farmers as a result of the ERI intervention.

In conclusion, it has been demonstrated that rural livelihood outcomes pertaining to crop and livestock production, income, asset ownership and fertiliser use are significantly impacted upon by AIS driven research interventions. In-depth analysis, however, demonstrates that although participating households have more robust livelihoods during the intervention, phasing out of the research programme reduces the effects on livelihoods.

4.5 HOUSEHOLD TYPOLOGIES AND INCOME PORTFOLIOS

The second objective of this study aims to demonstrate the impact on livelihood outcomes of macro-economic policy shocks and policy sequencing, depending on participation in agricultural research interventions that are driven by innovation systems concepts, socio-economic characteristics and resource endowments. In order to assess the effects of differences in resource endowments and socio-economic characteristics in the sampled households, household typologies were developed and income portfolios for the different household categories were analysed. These are presented in the following two sections.

4.5.1 Household typologies

Household typologies offer a useful tool for stratifying households into homogenous units with minimal internal differences which provide an opportunity for developing appropriate disaggregated policy recommendations. Variables used for developing the typologies were based on the knowledge that causes heterogeneity between households. These included value of assets (USD), farm sizes (acres), number of farm plots and household sizes (Table 4.13).

Table 4.13: Summary of the cluster solution

Cluster	Intervention community			
	Mean value			
	Value of assets USD)	Farm size (acres)	Number of farm plots	Household size
1	92.14	3.2	2.4	5.1
2	318.62	5.6	4.2	4.7
3	1605.06	5.6	2.6	6.4
	Counterfactual community			
1	976.41	11.96	5.0	4.0
2	119.55	4.34	2.8	5.9
3	41.78	2.06	1.6	3.72

Clustering was done separately for households in the two communities to allow for differences in livelihood outcomes that may arise from the ERI program. Using the agglomerative hierarchical clustering technique, three distinct clusters were formed in both the intervention community and the counterfactual community. Divisive non-hierarchical clustering was employed as an alternative approach in order to validate the cluster solutions

To determine the significance of the differences between the mean values of the variables used for clustering, a one-way Analysis of Variance (ANOVA) was carried out. The results of the ANOVA for the intervention community and counterfactual community are presented in Tables 4.14 and 4.15 respectively. From Table 4.14, it can be seen that in the intervention community, statistically significant differences exist between the three clusters that have been identified. These differences exist for the farm size, the value of assets and the number of farm plots which are all highly statistically significant. The mean household sizes were, however, found to be non-significant between the three clusters in the intervention community. This implies that the household sizes in the intervention community are generally similar and, as such, household size was not an important variable in describing the differences between households in the intervention community.

Table 4.14: ANOVA results for the intervention community clusters

Variable		Intervention community				
		Sum of squares	Degrees of freedom	Mean square	F-ratio	ρ
Value of assets (USD)	Between groups	2.11	2	1.06	105.41	0.000
	Within groups	9.82	98	1.00		
	Total	3.09	100			
Farm size (acres)	Between groups	149.70	2	74.85	15.29	0.000
	Within groups	479.51	98	4.89		
	Total	629.22	100			
Number of farm plots	Between groups	79.53	2	39.76	87.51	0.000
	Within groups	44.53	98	0.454		
	Total	124.06	100			
Household size	Between groups	12.64	2	6.32	0.92	0.403
	Within groups	675	98	6.88		
	Total	687.64	100			

For the counterfactual community, Table 4.15 shows that there are high statistically significant differences between the mean values for all the variables used to obtain a cluster solution. Hence, in the counterfactual community, the three clusters formed differ significantly in terms of the value of assets, the farm size, the number of farm plots and the household size. The results of an ANOVA generally indicate that differences exist between at least one pair of the clusters in terms of the mean values for the variables used for clustering. This is the case for the clusters in both communities.

The ANOVA results do not, however, provide information about the pair of clusters that are similar or dissimilar from one another. *Post hoc* tests are hence used to determine the actual differences between pairs. In this study, the Tukey's Honestly Significant Difference (HSD) test was used for such pairwise comparisons. *Post hoc* tests are only carried out on variables whose F-statistic in the ANOVA was statistically significant. This implies that for the intervention community, the household size was dropped from the *post hoc* test as had an F-statistic that was non-significant. Appendix 1 (Tables A4.1 and A4.2) present the results of the Tukey's HSD test for the intervention community and counterfactual community, respectively.

Table 4.15: ANOVA results for counterfactual community clusters

Variable	Counterfactual community					
		Sum of squares	Degrees of freedom	Mean square	F-ratio	ρ
Value of assets (USD)	Between groups	5.303	2	2.65	140.336	0.000
	Within groups	3.760	199	1.88		
	Total	9.062	201			
Farm size (acres)	Between groups	483.676	2	241.84	59.695	0.000
	Within groups	806.188	199	4.051		
	Total	1289.864	201			
Number of farm plots	Between groups	93.920	2	46.66	60.974	0.000
	Within groups	152.284	199	0.765		
	Total	245.604	201			
Household size	Between groups	229.563	2	114.78	32.057	0.000
	Within groups	712.536	199	3.581		
	Total	942.099	201			

The *post hoc* test for the intervention community shows that in terms of the value of assets, all the three clusters in the intervention community are statistically different from each other, with cluster one having lower monetary value of assets than either cluster two or three; and cluster three having higher monetary value of assets than cluster two. In terms of farm size, the results indicate that there is a statistically significant difference between the farm size of cluster two and cluster one; but cluster two is not significantly different from cluster three in terms of land ownership. Furthermore, the results indicate that households in cluster two had on average more separate farm plots than either cluster one or three; and that there is no statistically significant difference between clusters one and three in terms of ownership of separate farm plots.

For the counterfactual community, the results of the Tukey's HSD *post hoc* test show that cluster one differs significantly from both cluster two and three in that it has significantly higher means for the values of assets, the farm holding size and the number of separate farm plots. Cluster two has lower-valued assets, fewer numbers of farm plots and low land holdings as compared to cluster three. In terms of household size, the results show that cluster two differs significantly from cluster three in that households in cluster two had larger household sizes than those in cluster three. There are, however, no statistically significant differences between cluster one and three or clusters one and two in terms of the household sizes.

Validation of the identified clusters was done by applying an alternative clustering method to the agglomerative hierarchical clustering technique. Divisive non-hierarchical clustering was hence used as the alternative clustering technique. ANOVA was carried out to ascertain the existence of differences between the clusters formed and the Tukey's HSD test was applied *post hoc* to determine actual differences between specific pairs. Results of the divisive non-hierarchical clustering (Appendix 1) are consistent with the results of the agglomerative hierarchical clustering technique. Although small differences exist in terms of the cluster sizes, the results of the two clustering algorithms are generally similar.

Final profiling of the clusters was based on those variables that had a significant F-statistic from the ANOVA as well as on differences observed on additional variables. For this study, the total household income for the 2008/2009 cropping season, fertiliser use in the 2008/2009 cropping season and the value of upland maize harvest for the 2008/2009 agricultural season were used as the additional profiling variables. These additional variables are those that are able to indicate the potential for differences in livelihood outcomes between the clusters. Validation focuses on variables included in the clustering, while profiling of clusters focuses on variables not included in the cluster solution (Hair, *et al.* 1995:454). ANOVA results of the three clusters in both the intervention and counterfactual communities for the additional profile variables are given in Tables 4.16 and 4.17, respectively.

ANOVA results indicate that differences exist between the three identified clusters in both communities. For the intervention community (Table 4.16), differences exist for the three clusters in terms of the value of the 2008/2009 maize harvest, fertiliser use and total income for the 2008/2009 agricultural season. For the counterfactual community (Table 4.17), the three clusters have differences existing for only the 2008/2009 total income and the value of maize harvested in the 2008/2009 season, but not for the inorganic fertiliser use for the same season.

Table 4.16: ANOVA results for intervention community (profile variables)

Variable (2008/2009)		Intervention community				
		Sum of squares	Degrees of freedom	Mean square	F-ratio	ρ
Total income (USD)	Between groups	2.908	2	1.454	11.524	0.000
	Within groups	1.236	98	1.262		
	Total	1.527	100			
Fertiliser use (no. of 50 kg bags)	Between groups	56.802	2	28.401	4.955	0.009
	Within groups	561.708	98	5.732		
	Total	618.511	100			
Value of maize harvest (USD)	Between groups	5.252	2	2.626	8.468	0.000
	Within groups	3.039	98	3.101		
	Total	3.565	100			

Actual differences between pairs of clusters for the additional profile variables were obtained by conducting the Tukey's HSD test. For the counterfactual community, inorganic fertiliser use was dropped from the *post hoc* test as the F-statistic in the ANOVA was non-significant. Results of the Tukey's HSD *post hoc* tests (Appendix 1) show that cluster three in the intervention community has positive significant differences from either cluster one or two with higher mean incomes, higher value of maize harvest, and largest number of 50 kg bags of inorganic fertilisers applied for the 2008/2009 cropping season. Clusters one and two do not have any statistically significant differences between them in terms of fertiliser use and the value of maize harvest for the 2008/2009 agricultural season. However, there is a statistically significant difference between income for the 2008/2009 season with households in cluster two having higher incomes than those in cluster one.

Table 4.17: ANOVA results for counterfactual community (profile variables)

Variable (2008/2009)		Counterfactual community				
		Sum of squares	Degrees of freedom	Mean square	F-ratio	ρ
Total income (USD)	Between groups	4.719	2	2.359	76.720	0.000
	Within groups	6.120	199	3.075		
	Total	1.084	201			
Fertiliser use (no. of 50 kg bags)	Between groups	39.595	2	19797	0.989	0.374
	Within groups	3984.75	199	20.024		
	Total	4024.35	201			
Value of maize harvest (USD)	Between groups	2.363	2	1.818	80.195	0.000
	Within groups	2.917	198	1.473		
	Total	5.879	200			

For the counterfactual community, cluster one has the highest level of total income and value of maize harvest for the 2008/2009 cropping season as compared to the other clusters, and these differences were statistically significant. Cluster three, on the other hand, had the lowest levels of either total income and value of maize harvest in the 2008/2009 season. These findings for the counterfactual community are consistent with the findings from both the agglomerative hierarchical and divisive non-hierarchical clustering techniques. This clearly distinguished three clusters, with cluster one having the highest mean values for all variables under consideration and cluster three having the lowest.

Based on the cluster analysis solution as well as the validation results, the study finds that in the counterfactual community, there are three very distinct household typologies; while in the intervention community, there are three overlapping household types. In the counterfactual community, the three distinct household types are the low resourced, medium resourced and large resourced households. Although the first two categories have similarities, there are very distinct differences in terms of land ownership. Table 4.18 provides a summary of each of the household typologies from the counterfactual community.

Table 4.18: Description of household typologies in the counterfactual community

Households characteristics	Household resource group		
	Low resourced (62%)	**Medium resourced (37%)**	**Large resourced (1%)**
Asset value	Assets valued between USD0 to USD338	Assets valued between USD0 to USD600	Assets valued between USD750 to USD1000
Farm size	Size range: 0.162–2.23 ha About 1–3 pieces of land	Size range: 0.162–5.06 ha About 1–6 pieces of land	Size range 2.18–9.31 ha About 2–8 pieces of land
Fertiliser use (2008/2009)	Applied between 0 and 6.1 50 kg bags	Applied between 0 and 6.18 50 kg bags	Applied between 2.3 and 6.8 50 kg bags
Maize harvest	Valued in between USD1.16 to USD650	Valued in between USD11 to USD2800	Valued in between USD905 to USD2600
Annual cash income	Ranged from no income to USD1860	Ranged from no income to USD600	Ranged from USD1186 to USD5600
Educational attainment	Spouses: 40.8 % no formal education Household heads: 46.4 % attended but did not finish primary school Participation in extension trainings (2008/2009): 21.6 %	Spouses: 43.2 % attended but not finished primary school Household head:45.9 % attended but did not finish primary school Participation in extension trainings (2008/2009): 29.7 %	Spouses: 66.7 % no formal education Household heads: 66.7 % no formal education Participation in extension trainings (2008/2009): 0 %
Age range	19 to 78	18 to 87	30 to 63

*All monetary values are for the 2008/2009cropping season

Table 4.19 provides a summary description of each of these household types for the intervention community. In the intervention community, there was an overlap in the household typologies and the three categories of households can be categorised into low resourced households, large resourced households with low cash income, and medium resourced households with high cash income.

Table 4.19: Description of household typologies in intervention community

Households characteristics	Household resource group		
	Low resourced (56%)	**Medium resourced with high cash income (5%)**	**Large resourced with low income (39%)**
Asset value	Assets valued between USD0 to USD405	Assets valued between USD1198 to USD2044	Assets valued between USD15 to USD1400
Farm size	Size range: 0.40–2.63 ha About 1–3 pieces of land	Size range: 1.32–2.63 ha About 2–4 pieces of land	Size range: 0.73–8.09 ha About 3–6 pieces of land
Fertiliser use (2008/2009)	Applied between 0 and 16 50 kg bags	Applied between 2.28 and 8.24 50 kg bags	Applied between 0 and 6.97 50 kg bags
Maize harvest	Valued in between USD4.64 to USD6900	Valued in between USD600 to USD10446	Valued in between USD35 to USD5000
Annual cash income	Ranged from no income to USD2900	Ranged from USD500 to USD5000	Ranged from USD63 to USD4700
Educational attainment	Spouses: 40.4 % no formal education Household heads: 38.6 % attended but did not finish primary school Participation in extension trainings (2008/2009): 28.1 %	Spouses: 60 %attended but not finished primary school Household heads: 4 % attended but not finished primary school Participation in extension trainings (2008/2009): 40 %	Spouses: 33.3 % attended but not finished primary school Household head: 28.2 % attended but did not finish primary school Participation in extension trainings (2008/2009): 20.5 %
Age range	23 to 79	30 to 63	24 to 68

*All monetary values are for the 2008/2009cropping season

In conclusion the typology development exercise has demonstrated that households in the study area are differentiated in terms of not only physical resource endowments such as land ownership but also based on household staple food production, income earnings and input usage. These differences are not only found between different communities in the study area but also between households within the same community. The implications of these findings are that policy changes that affect the labour market, maize input and output markets as well as pricing and marketing of inorganic fertiliser will result in inter and intra community differences. Furthermore these findings imply that there is diversity in living standards amongst the poor in rural Malawi. This therefore entails that households respond differently to risks that arise either from nature factors or from man-made factors such as market reforms and policy changes. The simulation analysis in Chapter 6 tests this hypothesis and provides insight into how the use of innovation systems in agricultural research and development contributes to household decision making and reaction to policy changes that transmit through the market.

4.5.2 Household income portfolios

Using the household typologies that were identified, an analysis of each type of household's income portfolio was carried out. In this study, household income is not synonymous with cash income. Rather, it includes cash income earned from various employments; non-cash income earned from the sale of labour and other on and off-farm employment; income earned from the marketing of agricultural crops and livestock produce; and the imputed value of all crops harvested which are retained for home consumption. Imputed values are calculated on prevailing retail prices as that is the price that households would pay if they needed to purchase it. The ten different sources of household income that were identified are described in Table 4.20.

The largest share of household cash income in many poor rural communities in Sub-Saharan Africa comes from crop sales (Ellis, 2006). Through the pricing of staple food commodities such as maize which are regulated by government, rural households find that the incomes that they earn in marketing staple food crops; and the value of the crops that they produce and retain for consumption link them to the market economy. The linkages created by staple food production and sales is, however, not very robust, as staple food products provide less income and require little or no inputs (Davis, *et al.* 2002).

Table 4.20: Sources of income

Source of income	Description
All upland crops except maize	Value of all upland crops harvested in the main rainy season with the exception of maize that were sold at market prices; and the imputed value of crops kept for home consumption.
Salaried employment	Summation of income earned from both part and full-time salaried employment that was non-farm in nature.
On-farm seasonal employment	Summation of income from on-farm seasonal employment which during the main cropping season is mainly labour employment.
Off-farm seasonal employment	Income from semi-skilled and skilled work, such as carpentry, brick making/burning, brick laying and house building, which takes place usually just prior to the rainy season.
Non-agro based enterprise	Income from income-generating activities that were not agro-based. Common non-agro based commercial enterprises included operation of a general grocer and the selling of second-hand clothing.
Agro-based enterprise	Income from marketing own produce from own farm or crops brought from other farmers or areas. Also includes income from sale of processed goods, such as cooked and baked food stuff.
Livestock	Value of all livestock sold during a cropping season.
Wetland crops	Total income from all crops cultivated during the short dry winter season using either irrigation or residual moisture.
Remittances	All income not earned but received from relatives or through other channels.
Maize income	Income from maize included the computed value of maize sold, maize exchanged for other commodities and maize kept for household consumption.

As a result of this, rural households have little backward and forward production linkages to the rest of the economy (Davis, *et al.* 2002). However, in non-industrial communities such as the area for this study, income from staple food production provides the only valid linkages to the market economy. Aggregate household income was therefore estimated as an arithmetic function which summed the real and imputed income earnings from the ten different income sources. Estimates for income were computed for the household as one as opposed to it being disaggregated by household members due to the lack of disaggregated data.

An analysis of the income portfolios for the 2008/2009 cropping season for the different household types shows that there were differences in terms of the share of income from different sources for within and between households in the intervention community and the counterfactual community (Table 4.21). In the intervention community, for the low and large resourced households, maize contributed 49 % and 30 % of the total income, with all other upland crops contributing 42 % and 55 % of total income, respectively. For the medium resourced households, the main source of income was other upland crops (excluding maize) which contributed about 92 % of the total income. Maize contributed approximately 6 % of the total income.

Table 4.21: Income portfolio compositions (%)

Source of income*	Household typology by community		
	Low resourced	Medium resourced	Large resourced
	Intervention community		
Maize	49	6	30
All other upland crops	42	92	55
Livestock	2	1	11
All other sources	7	1	4
	Counterfactual community		
Maize	42	28	3
All other upland crops	33	56	64
Livestock	1	1	30
All other sources	7	6	1
Employment	17	9	2

*The values include both the real market earnings and the imputed value of non-marketed crop and livestock goods

In general, households in the intervention community had three main sources of income. These are income from the sale of maize; income from all other upland crops (excluding maize); and income from other sources, which constituted the income from both off and on-farm seasonal employment, other salaried employment, income from agro-based and non-agro-based commercial enterprises, and income from crops cultivated in the wetland and remittances. The last category of income (other sources) generally contributed to the smallest proportion of total income in all the three household categories. Income from livestock marketing was a small contributor towards total income in the low (2 %) and medium (1 %) resourced households, with only households in the large resourced category earning about 11 % of their total income from livestock.

In the counterfactual community, the low and medium resourced households received the largest share of their total income from crop production. For the low-resourced households, maize and all other upland crops constituted the largest and second largest share of income at 42 % and 33 %, respectively. For the medium resourced households, all other upland crops and maize production constituted the largest and second largest contributor at 56 % and 28 %, respectively. Employment which included on and off-farm labour employment and any full and part time employment, was the third largest contributor to total income for both the low and medium resource-endowed households, with contributions of 17 % and 9 % to total household income, respectively. Employment played a bigger role in the income portfolio of households in the counterfactual community than in the intervention community. This can be attributed to the proximity of the counterfactual community to a tarmac road. In addition, the proximity to the tarmac road makes the counterfactual community an attractive area for urban semi-commercial farmers who purchase land (both upland and wetland) from the locals. These semi-commercial farmers also provide employment to households in the counterfactual community, as they hire both men and women throughout the cropping season.

Apart from these major sources of income in the counterfactual community, both the low and medium resourced households also received a considerable amount of their incomes from other sources. These sources included incomes from remittances, non-agricultural commercial enterprises, agro-based commercial enterprises, and income from crops cultivated in the wetland. All other sources contributed approximately 7 % and 6 % in the low and medium resourced households, respectively. Income from livestock sales contributed about 1 % of total income in both the low and medium resourced household categories. For the large resourced households in the counterfactual community, all upland crops except maize contributed to approximately 64 % to the total income; followed by livestock, which contributed 30 % of the total income. This household type is the only one in this study in which livestock income contributed a significantly large share of total income. Apart from all other upland crops and livestock, other income sources for the large resourced households included maize (3 %), employment (2 %) and all other sources (1 %). Employment comprised any salaried employment, off and on-farm employment, while other sources of income included income from the sale of wetland crops, remittances, and income from both agro and non-agro-based commercial enterprises.

4.6 CHAPTER SUMMARY

It has been demonstrated that statistically significant differences exist between households in the intervention community and counterfactual community; which can be attributed to the ERI initiative. Positive impacts of ERI have been demonstrated for different livelihood outcomes, including production, income generation, fertiliser use patterns, as well as trainings and membership of farmer groups for households in the communities in which it was implemented. Hence, AIS driven research interventions have the potential to impact upon and change the livelihood outcomes of rural households within the maize-based farming system in Malawi by creating greater opportunities for linking the communities to markets. These findings provide the proof for the first hypothesis.

In addition, identified household typologies showed that low-resourced households have less diversified income portfolios; and concentrate more on earning income from casual wage employment with high dependence on subsistence food production as compared to better-off households. Income from maize was, however, a key contributor to household incomes for all typologies, with low-resource households having a larger share of their income emanating from maize as compared to better-off households. The implications of these findings are that macro-economic policy shocks that transmit through maize prices have the potential to impact differently upon the incomes and therefore livelihood outcomes of different households in the two communities. This hypothesis is tested using simulation analysis in Chapter 6. However, prior to this, Chapter 5 provides a basis for understanding maize price formation in Malawi maize markets and will develop a full partial equilibrium maize model which will be used to test the above-mentioned hypothesis.