An empirical analysis of poverty, inequality and the labour market in Malawi

by

Anderson Sawira Gondwe

Dissertation presented for the degree of Doctor of Philosophy (Economics) in the Faculty of Economic and Management Sciences at Stellenbosch University

Supervisor: Prof. Servaas van der Berg

December 2016

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2016

____

(Anderson Gondwe)

Abstract

This thesis is a consolidation of three related studies on Malawi. The first study contains spatial and temporal comparisons of poverty and inequality in Malawi using two non-monetary dimensions, namely an asset index and child nutritional status. Through stochastic dominance tests, the study establishes that poverty and inequality are unambiguously higher in rural areas, which contain 85% of the population, in the Southern region and among households headed by females. Results indicate that poverty has significantly declined over time and that the gains from growth have been pro-poor. We show that welfare does not vary much across regions and areas with respect to child nutritional status but there are large differences in asset poverty. Stunting is a bigger problem among children under the age of five than body wasting and being underweight. Econometric analysis shows that asset ownership is positively associated with household size, the age of household head and education attainment. Age dependency ratio and incidence of sickness are negatively associated with asset ownership. Multivariate analysis of child nutrition reveals that malnutrition first worsens before improving at some critical age. This is consistent with possible recovery found in some of the studies that track children over time. Also in accordance with some literature, we find that boys have weaker nutritional status than girls.

The second study looks at the role of education in poverty reduction identified through the labour market. This study contributes to research on returns to education by including self-employment activities and non-farm business enterprises. Unlike previous studies, this study uses panel data which has many advantages, as acknowledged in the literature. We find large and positive returns to education in Malawi suggesting that education is a good investment. The returns increase with the levels of education. Interestingly, females have higher returns to education than males with similar skills. Since the Malawian labour market is not homogeneous, our analysis distinguishes between the formal and informal employment sectors. Furthermore, studying Malawi's informal sector is important as it accounts for 78% of total employment. Our results show that education externalities exist and play an important role in non-farm enterprises. The findings are robust to sample selection and treatment of outliers. We further show that dealing with inconsistencies in the data helps improves the quality and reliability of the results.

The third study applies spatial panel data econometric techniques to the study of migration and employment in Malawi. The study shows that the magnitudes of coefficients drop after taking into account spatial dependencies. This confirms that studies that fail to take into account the spatial effects tend to overstate the results. By matching geographical codes that are consistent over time, it is now feasible to integrate census data with other data for similar spatial analysis. The study further evaluates the impact of land reform policy on spatial migration and employment using a difference-in-difference

estimation strategy. Results show that the policy has had significant effects on migration and employment patterns in Malawi.

Opsomming

Hierdie tesis is a konsolidasie van drie verwante studies oor Malawi. Die eerste studie ondersoek armoede en ongelykheid in Malawi oor tyd en ruimte heen deur twee nie-monetêre dimensies, naamlik 'n bate-indeks en die voedingstatus van kinders, te gebruik. Deur middel van stogastiese dominansie-toetse word ondubbelsinnig getoon dat armoede en ongelykheid hoër is in landelike gebiede, wat 85% van die bevolking huisves, in die Suidelike streek en onder huishoudings met vroue as hoof van die huis. Resultate toon dat armoede beduidend afgeneem het en dat groei tot voordeel van die armes strek. Ons resultate toon weinig verskille in welsyn tussen streke en gebiede met betrekking tot die voeding status van kinders, maar groot verskille in bate-armoede . Vertraagde groei is 'n groter probleem by kinders onder die ouderdom van vyf jaar as kwyning en ondergewig. Ekonometriese ontleding toon dat bate-besit positief verband hou met die grootte van die huishouding en die ouderdom en opvoedingsvlak van die hoof van die huishouding . Die ouderdom-afhanklikheidslas en die voorkoms van siekte hou negatief verband met bate-besit. Regressie-analise wys dat wanvoeding onder kinders eers met ouderdom toeneem voordat dit by hoër ouderdomme afneem, wat konsekwent is met die moontlikheid van herstel soos party studies wat kinders oor 'n tydperk volg bevind. Ook, in ooreenstemming met party studies, word bevind dat die voedingstatus van dogters beter is as dié van seuns.

Die tweede studie bestudeer die rol van onderwys in die vermindering van armoede in die arbeidsmark. Deur die insluiting van selfwerksaamheidsaktiwiteite en nie-landbou sakeondernemings dra die studie by tot navorsing oor die voordele van opvoeding in Malawi. Anders as in vorige studies, gebruik hierdie studie paneeldata, wat baie voordele inhou, soos in die literatuur bevestig. Ons vind groot en positiewe opbrengste op onderwys, wat daarop dui dat dit 'n goeie belegging is. Opbrengste neem toe met vlakke van onderwys. Interessant genoeg, ervaar vroue hoër opbrengste op belegging in onderwys as mans met dieselfde vaardighede. Aangesien die arbeidsmark in Malawi nie homogeen is nie, tref ons analise ‘n onderskeid tussen die formele en informele indiensnemingsektore. Dit belangrik om Malawi se informele sektor in ag te neem, aangesien dit 78% van die totale indiensneming uitmaak. Ons resultate wys dat daar eksternaliteite van onderwys bestaan wat 'n belangrike rol speel in nie-landbou ondernemings. Ons resultate is robuust virsteekproefseleksie en die hantering van uitskieters. Die uitstryk van data-onreëlmatighede dra tot 'n verbetering in die kwaliteit en betroubaarheid van die resultate by.

Die derde studie pas ruimtelike paneeldata ekonometriese tegnieke toe op migrasie en indiensneming in Malawi. Die grootte van koëffisiënte neem af as ruimtelike afhanklikhede in ag geneem word. Dit bevestig dat studies wat nalaat om ruimtelike aspekte in berekening te bring geneig is om effekte te oorskat. Deur konsekwente geografiese kodes oor tyd te verbind is dit nou moontlik om sensusdata met ander data te integreer vir verdere ruimtelike analise. Die studie evalueer ook die uitwerking van die

grondhervormingsbeleid op ruimtelike migrasie en indiensneming deur die gebruik van 'n verskil-in-verskille metodeevalueer. Die resultate dui daarop dat hierdie beleid 'n beduidende uitwerking op migrasie en werkloosheid in Malawi het.

Acknowledgements

Firstly, I would like to thank *Professor Servaas van der Berg* from whom I have received tremendous feedback and guidance throughout my PhD study. I will forever remember him for his positive contribution to my life and for opening my mind to the big picture of research. I know *Servaas* as a very humble, patient and wise man with outstanding fatherly care. Considering my non-academic background, the fact that I am completing my studies within three years speaks volumes about how *Servaas* is able to identify and develop talent in people.

Secondly, I acknowledge the contribution from Research on Social Economic Policy (RESEP) which is led by *Servaas*. RESEP has been important to my progress and timely completion of the studies. Apart from the networking, the organisation has been the source of the much needed additional funding for my studies and academic conferences. The RESEP team also boasts an excellent pool of researchers from which I have greatly benefited. It has provided a platform for the transfer of new skills and refining my work through the weekly departmental seminars and training workshops. I particularly thank *Dr Dieter von Fintel* for his insights that have also helped shape my study.

Thirdly, I also say my heart-felt thanks to The Stellenbosch Institute for Advanced Study (STIAS) for the scholarship funding provided through The Graduate School of Economic and Management Sciences (GEMS). The GEMS cohort system, which recruits PhD researchers from different parts of the world, has enabled me to develop important networks with colleagues from three different continents, namely Africa, Asia and Europe. In this regard, special mention goes to *Dr Jaco Franken*, the manager of the graduate school programme and fellow PhD students in the programme.

Lastly, I thank my family and friends who have provided encouragement to me during my physical absence from them.

I have made it to the glory of the LORD and Jesus Christ.

Dedication

I dedicate this work to my family and particularly *Mary Gondwe*. My wish is that there may never cease to be people who attain PhD education throughout our family generations.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Introduction

This Chapter provides some background information on Malawi to help the reader contextualise the study before turning to the topics addressed in this thesis. In Section 1.2, we look at Malawi's geography and history. Section 1.3 discusses the economy. The problem statement is discussed in Section 1.4. The background information on the national data sources is given in Section 1.5. Finally, Section 1.6 gives the thesis structure.

1.2 Geography and history

Malawi is a landlocked country located in Southern Africa. It is bordered by Zambia to the north-west, Tanzania to the northeast and Mozambique to the west, south and east as shown in Figure 1.1. The country is a long strip of about 901 km and ranges between 80 and 161 km in width[1]. The total surface area is about 118,484 km^2 of which 94,276 km^2 is made up of land. The remaining area is largely made up of Lake Malawi, Africa's third-largest fresh-water lake, about 475 km long (National Statistical Office & ICF Macro, 2011).

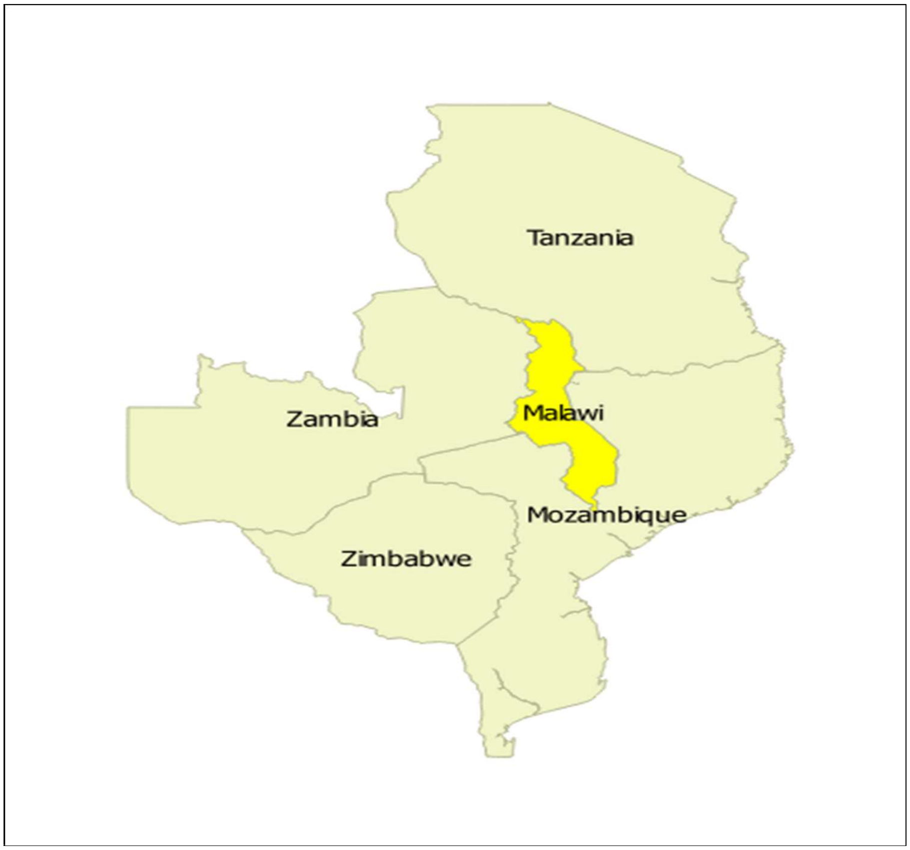
Administratively, Malawi is divided into three regions, namely the Northern, Central and Southern regions. Regions are also divided into districts and there are 28 districts in total: six, nine and 13 districts in the North, Centre and South, respectively. Each of the 28 districts is further subdivided into traditional authorities (TAs), ruled by senior chiefs. The smallest units of administration are villages typically governed by village headmen or women. Only a small proportion (about 15%) of Malawi's population resides in urban areas (National Statistical Office & ICF Macro, 2011). The geographical division of Malawi into regions, districts and traditional authorities provides an important dimension for decomposition analysis as we will see later.

Malawi became a protectorate of Britain in 1891. From 1953 to 1963, Malawi (formerly called Nyasaland) was part of the Federation of Rhodesia and Nyasaland together with Zambia (formerly Northern Rhodesia) and Zimbabwe (formerly Southern Rhodesia). Malawi became independent in 1964 and gained the status of a republic in 1966 (National Statistical Office & ICF Macro, 2011). Although politically independent for 52 years, the country is still highly dependent on foreign aid; the United Kingdom, the European Commission, the Global Fund and the World Bank continue to make up the four largest donors (Organisation for Economic Co-operation and Development, 2008). Multiparty democracy was abolished in 1966 but later reintroduced in 1993 after a national referendum. During the

[1] In Chapter 4 we compute distance matrices for spatial analysis.

first year of a multiparty democracy in 1994, primary education in government schools became free and this resulted in a large increase in school enrolment (Kadzamira & Rose, 2015).

Figure 1.1: Political map of Malawi



Source: Own construction from country shapefiles

1.3 Economy

The Malawian economy is largely dependent on agriculture, which made up about 30% of the gross domestic product (GDP) in 2015 and continues to directly benefit more than 75% of the population. The agricultural sector in Malawi consists of the smallholder sector mainly for subsistence production and the estate sector for exportation. The main food crops are maize, rice and cassava. In 2014, tobacco generated about 64% of foreign exchange. Other important export crops in Malawi are tea and sugar, accounting for about 9% and 8% of total export value, respectively (Reserve Bank of Malawi, 2015). However, more recently, tobacco, which is the main cash crop for export, has come under threat because of the world-wide anti-smoking campaign.

According to Food and Agriculture Organisation (FAO) AQUASTAT (2015), less than 1% of cultivated areas is under irrigation[2]. Since much of Malawi's agriculture is dependent on rains, food security and household incomes are threatened by flooding and droughts. Irrigation development, if combined with advancements in good cropping systems and inputs, has the potential of improving farm incomes and food security for the majority of the population who are involved in subsistence agriculture. The Green Belt Irrigation is one of the key priority areas identified in the national development agenda with the aim of facilitating growth and development in the coming years by taking advantage of the country's abundant water resources. Potentially, 1 million hectares can be irrigated in Malawi. The current project is expected to expand the amount of land under irrigation from 90,000 to 400,000 hectares (Government of Malawi, 2007, 2012).

The development policy agenda for Malawi is focussed on poverty reduction and is summarised in the Malawi Growth and Development Strategy (MGDS), a five-year strategy. The first programme (MGDS I) was launched in July 2007 and ran through to 2011. The second programme, the Malawi Growth and Development Strategy II (MGDS II), was being implemented from 2011 and expired on 30 June 2016. Some of the elements of MGDSII have been rolled over by the government. MGDS I and MGDS II share the same themes but the latter has six themes instead of five. The themes or thematic areas are Sustainable Economic Growth, Social Development, Social Support and Disaster Risk Management, Infrastructure Development, Improved Governance and Cross-Cutting Issues. MGDS II includes a cross-cutting theme that deals with gender imbalances, capacity development, and research and development, HIV and AIDS, nutrition, environment, climate change, population and science and technology (Government of Malawi, 2007, 2012).

Malawi's national strategies are developed by the Government of Malawi in consultation with key stakeholders, particularly the International Monetary Fund (IMF) and the World Bank, from whom the country continues to receive technical support. In the 1980s and 1990s, Malawi embarked on economic and structural reforms under the Structural Adjustment Policies (SAPS) of the World Bank and IMF. The country was also granted debt relief under the 1996 Heavily Indebted Poor Countries (HIPC) initiative which the IMF and World Bank implemented for a number of developing countries. Despite these reforms and other targeted interventions, Malawi is still one of the least developed countries in the world and its economy is still undiversified (Government of Malawi, 2012; Organisation for Economic Co-operation and Development, 2008).

[2] Available at http://www.fao.org/nr/water/aquastat/main/index.stm.

According to the Human Development Report (2015), the Human Development Index (HDI) value for Malawi in 2014 was 0.445 and this puts the country on a ranking of 173 out of 188 countries and regions recognised by the United Nations (UN). Table1.1 provides a review of Malawi's progress between 1980 and 2014 in each of the three indicators that make up the HDI. The three dimensions of the HDI are a long and healthy life (measured by life expectancy), access to knowledge (measured by expected and mean years of schooling) and a decent standard of living (measured by Gross National Income (GNI))[3].

Table 1.1: Malawi's HDI trends based on consistent methodology and data

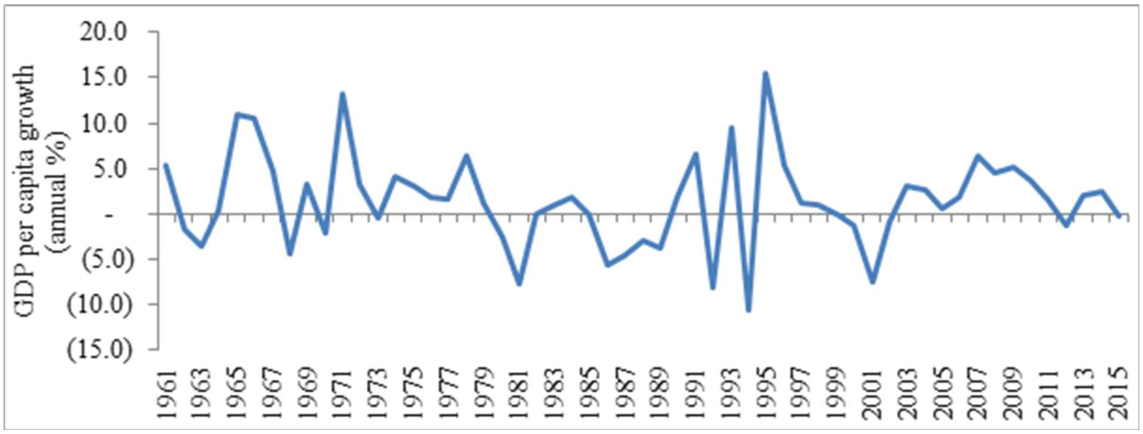
Year	Life expectancy at birth	Expected years of schooling	Mean years of schooling	GNI per capita (2011 PPP$)	HDI Value
1980	44.8	4.8	1.8	705	0.278
1985	45.1	4.6	2.1	643	0.278
1990	43.8	5.3	2.5	612	0.284
1995	43.5	11.1	2.7	556	0.334
2000	44.1	10.3	3.1	613	0.340
2005	48.3	9.6	3.4	601	0.355
2010	56.9	10.6	4.3	722	0.420
2011	58.6	10.8	4.3	732	0.429
2012	60.1	10.8	4.3	717	0.433
2013	61.5	10.8	4.3	726	0.439
2014	62.8	10.8	4.3	747	0.445

Source: Human Development Report (2015)

The table shows that from 1980 to 2014, life expectancy improved by 18 years, expected years of schooling by 6 years, mean years of schooling by 2.5 years and GNI per capita by about 6%. Comparing HDI changes over time using previously published reports would be erroneous due to the revisions and updates that take place from time to time. However, the comparisons provided in the table are based on consistent indicators and methodology developed by the United Nations Development Programme (UNDP) for the purpose of analysis over time. Therefore, the figures in the tables show real changes in values and Malawi's actual progress over time (Human Development, 2015). In Figure 1.2 we show that the annual GDP per capita growth for Malawi between 1961 and 2015 has been both volatile and dismal in some of the years.

[3] The average number of years of education is among individuals aged 25 years and older while the expected years of schooling are the total number of years of schooling a child of school-entry age is expected to receive assuming that the prevailing patterns of age-specific enrolment rates stay the same during the child's life.

Figure 1.2: Malawi's GDP per capita annual growth rates (1961-2015)



Source: Own computation from World Bank country data[4]

Population growth, alongside nutrition, is another critical area identified under the MGDS II's theme of Social Development. Due to high levels of fertility rates, averaging 5.05 children per woman in 2008, Malawi's population is rising rapidly. During the last population and housing census in 2008, the population was estimated at 13 million and projected at 17 million in 2015 (National Statistical Office, 2008). Malawi is now ranked at number 62 out of 196 countries by population with 149 people per km^2. This expanding population has implications for the labour market as well as the economy's ability to support jobs. Furthermore, in the MGDS II, labour and employment are identified as some of the sub-themes for sustainable economic development. Within this framework, the focus is placed on the creation of employment with a view to poverty reduction, incorporating gender specific issues in all labour initiatives and interventions, reducing practices of discrimination in the labour market and improved statistics on the labour market (Government of Malawi, 2012).

1.4 Problem statement

Our study is placed within the context of the critical issues facing Malawi as outlined in the national development plan, namely the MGDS. The government identifies malnutrition as a crisis particularly for the rural areas where most of the children suffer from high levels of stunting, wasting and underweight. Agriculture naturally comes on the national development agenda because it is the major source of employment for the Malawian population, as earlier stated. It is for this reason that the government continues to invest in agriculture with the aim of improving food security which has implications for nutritional status for the population. Land is identified as one of the critical resources for agricultural development and the Government of Malawi recently embarked on the land reform policy with the aim of improving production and incomes for the rural population which makes up 85% of the population (World Bank, 2012). A related project is the Integrated Rural Development (IRD)

[4] Available at http://www.worldbank.org/en/country/malawi.

programme which seeks to promote economic development in rural areas through the establishment of rural growth centres and provision of social services. Other critical areas in the national development plan include labour and employment, addressing gender imbalances and improving education (Government of Malawi, 2012).

It is against this background that our study focuses on poverty, child-nutritional status, education, employment and migration. The overarching theme in the thesis is poverty reduction and economic development. This study's contribution is towards an improved empirical understanding of the aforementioned economic phenomena. It, therefore, can inform and have implications for policy and development strategy.

1.5 National data sources

This study makes use of multiple data sources, all of which complement each other in the understanding of the issues discussed in the thesis. The main agency for data collection in Malawi is the National Statistical Office (NSO), responsible inter alia for conducting nationally representative surveys in the country. Malawi has an Integrated Household Survey programme consisting of national censuses, the Integrated Household Surveys (IHS) and Demographic Health Surveys (DHS). We identify these data sets as the most suitable for this thesis.

Conducted every ten years since 1966, the national censuses provide a unique source of data for understanding long-term patterns of economic phenomena in Malawi. Furthermore, they consist of information at small geographical areas which is important for understanding issues and planning at low levels of administration. Information collected in the censuses includes literacy, education, migration and employment, among other population characteristics (National Statistical Office, 2008).

The IHSs are conducted every five years and collect information on consumption expenditure, education, time use and labour, agriculture, health and child anthropometry. The first IHS was carried out in 1990 and was called the Household Expenditure and Small Scale Economic Activities Survey (HESSEA). Three rounds of integrated household surveys have been conducted after HESSEA, namely IHS1 conducted in 1997/8, IHS2 conducted in 2004/5 and recently IHS3 conducted between March 2010 and March 2011 (National Statistical Office, 2012). In between the surveys are the Welfare Monitoring Survey (WMS) normally conducted every year with the aim of tracking the living conditions of people, identification of the vulnerable population groups and collecting indicators for monitoring the attainment of national goals to which Malawi has committed itself, such as the MGDS and the

Millennium Development Goals (MDGs)[5]. So far, seven rounds of WMSs have been conducted in Malawi, namely for 2005, 2006, 2007, 2008, 2009, 2011 and 2014 (National Statistical Office, 2015).

An alternative data source for the understanding of labour markets issues is the 2013 Malawi Labour Force Survey (MLFS), which was conducted to provide a situational analysis of employment and unemployment in Malawi. The previous labour force survey was conducted in 1983 but was not publicly made available (National Statistical Office, 2012). Although it complements data from the IHSs and WMSs, we do not use data from the labour force survey as it is stand-alone and, therefore, inadequate for comparisons over time.

The DHSs are conducted every four years with main emphasis on health and nutrition, which is an area of focus of study in this thesis. The first Malawi DHS was conducted in 1992 (National Statistical Office & ICF Macro, 2011). Although the integrated household surveys also collect information on assets, health and child-anthropometry, the DHSs provide a better source because of the detailed extent to which these issues are dealt with in the questionnaire. In the context of Malawi, Verduzo-Gallo, Ecker, and Pauw (2014) point out some serious inconsistencies and data quality issues in child-nutrition estimates obtained using IHS data sets compared to anthropometric records from other nationally representative data sets. Specifically, they find that while the 2010 DHS and the 2009 National Micronutrient Survey (NMS) yield national child stunting rates of between 47% and 49%, estimates from the 2010/11 IHS3 suggest a prevalence rate of only about 30%. Similarly, estimates of child stunting levels based on the IHS2 data in 2004/05 are about 9 percentage points below the incidence rates based on the 2004/05 DHS and the 2006 Multiple Indicator Cluster Survey (MICS).

1.6 Thesis structure

This thesis is a consolidation of three related studies on Malawi. The introductory discussion provided in the previous sections of this Chapter covers the issues related to all the three studies. In order to allow for a detailed analysis, each of the studies forms a separate chapter. Similarly, the theoretical underpinnings of the studies are also discussed separately in each of the main chapters. In Chapter 2, we provide some theoretical considerations on poverty and inequality measurement, stochastic dominance analysis and measurement of pro-poor growth. In Chapter 3, two main competing groups of theories for explaining labour market outcomes are discussed, namely the traditional neoclassical model of labour supply and the segmented labour market hypothesis. Examples of the theories discussed are the human capital theory, Roy's (1951) two sector model and the Harris-Todaro (1970) model of migration, among others. Furthermore, we discuss the different theories or explanations as to what constitutes or gives rise

[5] MDGs have now replaced with Sustainable Development Goals (SDGs) since 2015.

to informal labour markets. In Chapter 4, we discuss the theoretical perspectives of migration, broadly grouped under either the disequilibrium or equilibrium perspectives. The specific theories discussed in the chapter are the gravity models of migration, the human capital theory and the 'spatial job-search models'.

Chapter 2 contains spatial and temporal comparisons of multidimensional poverty and inequality in Malawi based on two non-monetary dimensions of welfare, namely an asset index and child nutritional status. Data for this study are drawn from the DHSs. Through this paper, we attempt to present Malawi's profile of poverty and inequality, including the progress made over time. We also show the extent to which the observed changes over time have been pro-poor. Child nutritional status is identified as a problem requiring attention in the country's national development agenda (MGDS). The first part of the chapter deals with the derivation of our two dimensions of poverty. The second part conducts robust comparisons of economic welfare and an econometric analysis of underlying possible reasons behind the observed changes. The third and final part analyses pro-poor changes in poverty over time.

In Chapter 3, we look at the role of education in poverty reduction in Malawi by using data from the Malawi Integrated Household Panel Survey (IHPS). The linkage between education and poverty reduction is identified through the labour market. The argument made here is twofold, namely that education improves an individual's chances of getting employment and that education positively impacts on earnings. This partly justifies why governments invest in education, which is acknowledged to have positive externalities on households and communities (e.g. Basu & Foster, 1998). In the context of Malawi, primary education was universally made free in 1994, which is before the MDGs, while university education is either directly subsidised or students are granted study loans. The overall objective of this chapter is to estimate returns to education. Our analysis distinguishes between wage employment and self-employment activities (household enterprises), which make up a large percentage of total employment in Malawi. With respect to self-employment, the returns to education are calculated at the household level using the maximum level of education in the household. Prior to the analysis, we also conduct some consistency checks in the data to ensure data quality and meaningful comparability over time.

The theme of employment continues through Chapter 4, where emphasis is now placed on gender related issues and the role of spatial effects in employment. The data used are from the national censuses. Segregation of results by gender is important because women form a large percentage of the labour force in Malawi, with the majority of them engaged in the agricultural sector. Specifically, Chapter 4 applies spatial panel data econometric techniques to the study of migration and employment in Malawi. It is widely recognised in the literature that both geography (space) and time are important to the understanding of economic phenomena. However, only few studies incorporate spatiotemporal analysis and within the context of Malawi, this study is the first-time attempt. First, we first match geographical codes so that they are consistent over time. Once this is done, we analyse long-term patterns of migration and employment in Malawi. In 2004, the Government of Malawi introduced land reform with the aim of increasing the incomes of poor rural families in four Southern region districts of the country. This policy was aimed at poverty reduction through increase of incomes and improvement in food security for the participating families. To be specific, willing individuals purchased agricultural land from willing sellers and resettled in the new areas. Therefore, the second part of this study is dedicated to the analysis of the effects that the land reform policy had on migration and employment. The findings from this study are important because of the future joint plans by the World Bank and Malawi Government to scale up the project to the rest of Malawi.

Finally, Chapter 5 concludes the thesis. The chapter provides a discussion on how the thesis addresses the research questions developed in each of the studies. We also look at the significance of the thesis in terms of the contributions made to research, implications of the research, its limitations and suggestions for future study.

Chapter 2

Measuring poverty and pro-poor growth in Malawi

2.1 Introduction

Poverty and inequality remain big concerns in Malawi, a very poor country. Deprivation exists in a number of dimensions such as education, consumption, child nutritional status and assets. Based on household per capita consumption estimates from the Third Household Integrated Survey (IHS3), about 51% of households in Malawi are poor. In addition, the Gini coefficient shows that inequality has increased over the past five survey years from 0.390 in 2005 to 0.452 in 2011. Estimates based on the 2010 Malawi Demographic Health Survey (MDHS) indicate that the incidence of child stunting in Malawi stands at 47% (National Statistical Office & ICF Macro, 2011).

Following the works of Sen (1985, 1987), a number of approaches have been developed to measure multidimensional poverty. A multidimensional view of poverty considers more than one aspect of deprivation. Conventionally and for a long time, poverty has been looked at in terms of either income or consumption. However, this view of poverty has been criticised for ignoring other important dimensions of well-being such as health, education, empowerment and freedom of association. Based on the literature, one can group the existing approaches to the measurement of multidimensional poverty into three alternatives.

The first approach aggregates a number of dimensions of poverty such as life expectancy, literacy and Gross Domestic Product (GDP) into a single one-dimensional index. Examples include the Human Development Index (HDI) and the Multidimensional Poverty Index (MPI). Ravallion (2011) raises questions as to whether such single-one dimensional indices (which he refers to as "mashup" indices) are sufficient for poverty measurements as opposed to developing a set of poverty indicators that are relevant within a particular setting. Specifically, Ravallion (2011) criticises the composite indices for not being so useful for sound development policy making because they essentially 'collapse' important dimensions into a single index which is difficult to interpret. Another criticism of the multidimensional indices raised by Ravallion (2011) is how weights are applied in the construction of multidimensional indices. Specifically, in as much as it is recognised that poverty is multidimensional, weights need not be determined by the poverty analyst but should rather be consistent with the choices made by the poor people. The second approach considers two or more dimensions of poverty such as income, education, health, etc. but analyses each dimension independently, without taking into account the possible correlations which may exist between the dimensions (e.g., Sahn & Stifel, 2003; Mussa, 2013). The third approach also considers two or more dimensions but unlike the second approach takes into account the potential interrelationships among dimensions (e.g., Gondwe, 2011; Duclos, Sahn, & Younger, 2006; Batana, 2008; Batana & Duclos, 2008). In this approach, a poverty line is set for each of the dimensions and then a decision is made as to whether an individual is to be considered poor if deprived in just one, some or all of the chosen indicators. In the literature, poverty has been found to be higher in distributions with higher correlations between the measures of well-being than those with lower correlations. Therefore, it is possible that univariate and multivariate analyses of poverty produce different rankings of poverty between distributions (Bourguignon & Chakravarty, 2003).

The third approach, therefore, looks at poverty measures that make for possible substitutions and complementarities between the levels of dimensions. Assuming two dimensions, for the substitutability assumption, we expect that the more someone has of one dimension of poverty, the less is overall poverty deemed to be reduced if their value of the other dimension is increased. On the other hand, for the complementarity assumption, increasing one dimension would reduce overall poverty. For example, transferring education from the poorly nourished to the better nourished would reduce overall poverty because better-nourished children learn better (Bourguignon & Chakravarty, 2003).

Most previous studies on poverty in Malawi concentrated on unidimensional poverty analysis (e.g., Murkherjee & Benson, 2003; Bokosi, 2006). Mussa (2013) considered three dimensions of poverty and inequality in Malawi, namely household per capita consumption, education and health. However, the study looked at the three dimensions one-at-a-time (or independently) without taking into account the correlations that exist between the dimensions of well-being. Gondwe (2011) did account for the possible correlations that exist between the dimensions of poverty. Two dimensions were used, household per capita consumption and education.

This study conducts spatial comparisons of multidimensional poverty and inequality using two non-monetary dimensions, namely an asset index and child nutritional status. We look at the two dimensions separately. It is the first time attempt to apply the asset index approach to the measurement of poverty in Malawi and uses a more recent DHS data set compared to the 2004 DHS data used by Alkire and Santos (2010). Also, the study conducts pro-poor growth analysis in the selected dimensions of living standards over two decades, from 1992 to 2010. As pointed out in Grosse et al. (2008), pro-poor growth analysis has recently become important to researchers and policy makers particularly with respect to monitoring progress towards the attainment of MDGs (now SDGs). However, the current efforts of pro-poor growth analysis have largely focussed on monetary dimensions of poverty thereby ignoring the non-monetary aspects of well-being. This study tries to reduce this shortcoming in the literature of the current pro-poor growth analysis by using two non-monetary dimensions of poverty, namely assets and nutritional status which are also central to the attainment of SDGs.

Household income, consumption expenditures and assets are the three main indicators of economic status that exist in the literature. The use of assets has gained popularity in recent years (Filmer & Pritchett, 1998; Sahn & Stifel, 2000; Booysen, Van der Berg, Burger, Von Maltiz, & Du Rand, 2008). Using data from Demographic and Health Surveys (DHS), an index is computed from a number of asset variables and this forms the basis for ranking households by their long-run socio-economic status. Furthermore, assets as a measure of economic status have been found to have more advantages than both income and expenditure. We discuss four main advantages. Firstly, unlike assets, both income and household expenditures are associated with measurement problems. For example, many respondents hide their incomes and only provide income figures in ranges. Income and consumption are also associated with seasonality and, therefore, unreliable as long term measures of status. Secondly, unearned income such as interest on loans, gambling, etc. is not reported. Income on home production and self-employment activities is usually excluded just as expenditure on non-routine goods and services. Thirdly, it is usually the income or consumption expenditure of the respondent (in most cases household head) that is recorded as opposed to the rest of the household members. Fourthly, data collected on income and expenditure is usually over the past month, week or day thereby raising questions as to what period of time should be covered (Rutstein & Johnson, 2004).

Based on the foregoing discussion, wealth is not only said to represent a more permanent status than income and expenditure but also more easily measured with only a single respondent required in most cases. In addition, the collection of asset information requires fewer questions than in income and expenditure surveys (Rutstein & Johnson, 2004).

The study achieves five objectives. First, it presents spatial poverty and inequality comparisons in assets and child nutritional status across population groups (areas, regions and sex of household head) in Malawi. Related to the first objective, we conduct poverty and inequality decompositions to see the relative contributions by the respective population groups or distributions. Second, it establishes a robust ranking of poverty and inequality across the groups that are compared. Third, it identifies the factors associated with asset poverty and child nutritional status in Malawi. Fourth, it tracks the incidences of asset poverty and child malnutrition in Malawi over the past two decades using a series of cross-sectional data sets. Finally, it establishes if the observed changes in living standards and child nutritional status over time have been pro-poor, absolutely and relatively speaking. Relative pro-poor changes in welfare have implications for inequality since poor people benefit more from the changes than the rich.

2.2 Theoretical considerations in poverty measurement

Three conditions are necessary for poverty measurement, namely a set of welfare indicators, poverty line and poverty measure (World Bank, 2004). The first condition is the choice of the welfare indicator which can be grouped into two, namely monetary (e.g. consumption or income) and non-monetary dimensions (e.g. assets or child-nutritional status). There is debate in the literature regarding which is a better indicator of welfare. The second condition is the choice of the poverty which can be looked at as the threshold separating the poor from the non-poor with the former falling below it. There exist two definitions of the poverty line. On the one hand, we have the absolute poverty line which is set for a particular group without reference to other members in the population. This poverty line is determined with respect to the basic needs needed for a living by a household or individual. On the other hand, we have the relative poverty line set with reference to the population, say at 60% of the average percapita consumption. The choice of which poverty line to use depends on the population we are studying. For poor countries, an absolute poverty line seems appropriate since emphasis is to ensure that the basic needs of the population are met. However, for richer countries that have met the basic needs a relative poverty line would make sense.

Having chosen the measure of welfare and poverty line, the third step is the choice of the poverty measure to use (Haughton & Khandker, 2009). Several poverty measures are available in the literature such as the Watts index and the Sen-Shorrocks-Thon index, among others. A good poverty measure is supposed to satisfy some basic axioms to be considered reliable (see for example Sen, 1976; Kakwani, 1980; Foster, Greer, & Thorbecke, 1984). In this study, we use the Foster-Greer-Thorbecke (FGT) measures because of the decomposability property which they possess in addition to other favourable characteristics. We consider three FGT indices, namely the headcount index, poverty gap index and the squared poverty gap or poverty severity index (Foster, Greer, & Thorbecke, 1984). The FGT measures are given as:

$$P(z,\alpha)=\frac{1}{N}\sum_{i=1}^{N}\left(\frac{z-y_i}{z}\right)^{\alpha} I(y_i<z) \tag{2.1}$$

Where: y_i, z, N, α are the welfare indicator, poverty line, population size and measure of poverty aversion, respectively. $I(.)$ is an indicator function that takes on a value of 1 if the expression is true and 0 otherwise. When $\alpha=0$, the result is the poverty headcount index which is a measure of the proportion of the population that is poor. For $\alpha=1$ we have the poverty gap index which indicates the extent to which individuals on average fall below the poverty line and expresses it as a percentage of the poverty line. Finally, when $\alpha=2$, we have the squared gap index which averages the squares of the poverty gaps relative to the poverty line.

Although the headcount index (P_0) is easy to understand and measure, it does not indicate how poor the poor are. Unlike the headcount index, the poverty gap index (P_1) measures the extent to which individuals fall below the poverty line (the poverty gaps) as a proportion of the poverty line. The main limitation of the poverty gap index is that it does not reflect changes in inequality among the poor. By averaging the squares of the poverty gaps relative to the poverty line, the squared poverty gap index (also called the poverty severity index, P_2) is able to show the changes in inequality among the poor.

2.3 Inequality measurement

Unlike poverty analysis which only focuses on the poor individuals or households, inequality is defined over the entire population and takes into account both the rich and the poor. Here we only consider the Gini and Theil indices measures due to their desirable properties. The other measures of inequality discussed in the literature include Decile Dispersion Ratio, Atkinson's inequality measures and Coefficient of Variation (see Haughton & Khandker, 2009; Duclos & Araar, 2009, for discussion).

The Theil indices are advantageous because they are additive across different population subgroups and enable us to see between and with group inequalities. On the other hand, the Gini is easy to understand and has a desirable graphical representation. It is for this reason that the Gini is preferred in most studies. The Gini coefficient varies between 0 (representing equal distribution) and 1 (representing a complete inequality). On the other hand, Theil index values vary between zero and infinity, which reflect complete equality and inequality, respectively. Graphically, the Gini coefficient is calculated as the area above the curve but below the line of perfect equality divided by the total area below the line of perfect equality. Apart from measuring the level of inequality, the Lorenz curve is also be used to test for inequality dominance between two distributions.

The Gini coefficient is calculated by the following formula:

$$G = 1 - \sum_{i=1}^{N} (x_i - x_{i-1})(y_i + y_{i-1}) \qquad (2.2)$$

Where: G refers to the Gini coefficient; x_i is the cumulative proportion of the population (represented on the x-axis) and y_i be the cumulative proportion of the welfare indicator (in our case child-nutritional status and asset index).

If there are N equal intervals on the x-axis, equation (2.2) collapses to:

$$Gini = 1 - \frac{1}{N} \sum_{i=1}^{N} (y_i + y_{i-1}) \qquad (2.3)$$

The Gini satisfies four main properties and these are mean independence, population size independence and the Pigou-Dalton Transfer sensitivity. However, the Gini does not satisfy two important characteristics, namely decomposability by population groups, dimensions or sources and statistical testability over time although this is less problematic now due to the fact that confidence intervals can typically be obtained through the use of bootstrap techniques. The Theil index measures satisfy all of the six properties.

The Theil indices are part of a larger family of measures referred to as the Generalised Entropy (GE) class of indices. The general specification of the GE measures is given as:

$$GE(\theta) = \frac{1}{\theta(\theta - 1)}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\bar{y}}\right)^{\theta} - 1\right] \tag{2.4}$$

Where: y is the selected welfare indicator or dimension and $\bar{y}$ is its average or mean. The parameter θ gives the weight given to distances between values of a given indicator at different parts of the distribution, and can take any real value. The GE index is more sensitive to changes in the lower tail of the distribution for lower values of θ, and for higher values, GE is more sensitive to changes occurring at the upper tail. When $\theta = 0$, we have the Theil-L index, also called mean the mean log deviation measure, and when $\theta = 1$, the result is the Theil-T index.

2.4 Stochastic dominance analysis

Dominance tests are necessary because poverty or inequality ranking can be reversed by different choices of poverty lines, measures, aggregation procedures and samples. Stochastic dominance analysis seeks to achieve non-ambiguous ranking in terms of welfare and inequality between any two distributions (Araar, 2006; Davidson & Duclos, 2000).

First, we discuss poverty dominance. Assuming two distributions, A and B, for our dimensions of poverty, namely asset index and child nutritional status, F_A and F_B will be the cumulative density functions (CDFs). Distribution B is said to dominate distribution A stochastically at first order if, for any argument y, $F_A(y) \geq F_B(y)$. In terms of poverty, this means that there is (weakly) more poverty in distribution A than there is in B. Higher orders of stochastic dominance are obtained through repeated integrals of the CDF of each distribution (Davidson & Duclos, 2000). Generally, we have:

$$D^1(y) = F(y), \quad D^{s+1}(y) = \int_0^y D^s(z) \; dz, \quad for \quad s = 1,2,3,... \tag{2.5}$$

Where:

D^1 is the CDF of the distribution under study;

$D^2(y)$ is the integral of D^1 from 0 to y;

$D^3(y)$ is the integral of D^2 from 0 to y, and so on.

By definition, distribution B dominates A at order s if $D_A^s(y) \geq D_B^s(y)$ for all arguments $y \in [0, z_{\max}]$. The lower limit of 0 represents the lowest value of the welfare indicator while $z_{\max}$ is the maximum acceptable poverty line for each welfare indicator. First-order dominance implies dominance at all higher orders (Davidson & Duclos, 2000). Where first order dominance is not established, we proceed to higher levels but stop at third order dominance as is the practice in the literature (e.g., Mussa, 2013).

Lorenz curves are the widely used approach to testing stochastic dominance in inequality (Araar, 2006). A given distribution is said to Lorenz dominate another distribution if the Lorenz curve of the first distribution lies everywhere above that of the latter. We then say that there is less inequality in the distribution with the higher curve than in that with a lower curve. Simply put, inequality is higher in A than in B if $L_B(p)$ is everywhere above $L_A(p)$. Distribution B dominates distribution A in inequality, with the second order, if

$$L_A(p) > L_B(p) \quad \forall p \in [0,1] \tag{2.6}$$

Where p is the percentile. The Lorenz curve for the percentile p can be defined as follows:

$$L(p) = \frac{\int_0^p Q(q)dq}{\int_0^1 Q(q)dq} = \frac{1}{\mu}\int_0^p Q(q)dq \tag{2.7}$$

$L(p)$ is the cumulative proportion of the welfare indicator (asset index or child-nutritional status) held by a cumulative percentage p of the population, when individuals are ordered in increasing asset or child-nutritional values. The integral $\int_0^p Q(q)dq$ gives the sum of the values of the welfare indicator of the bottom p proportion (the poorest $100p\%$) of the population. $\int_0^1 Q(q)dq$ gives the sum the welfare indicator values of all (Duclos & Araar, 2006).

The inequality dominance tests used in this study are based on Araar's (2006) theoretical developments. Specifically, generalised Lorenz dominance tests are used, and these turn out to be the same thing as second-order stochastic poverty dominance (Araar & Duclos, 2013).

2.5 Poverty and inequality decomposition

As indicated in Sections 2.2 and 2.3, the FGT and GE indices are decomposable by population groups. In this study, we follow decompositions based on Araar and Duclos (2013). The decomposition of the FGT index enables us to determine the absolute or relative contribution of each group such as area, region or etc. It takes the following form:

$$\hat{P}(z;\alpha)=\sum_{g=1}^{G}\hat{\phi}(g)\hat{P}(z;\alpha;g) \tag{2.8}$$

Where: G refers to the total number of population groups; $\hat{P}(z;\alpha;g)$ and $\hat{\phi}(g)$ are the estimated FGT index and population share of subgroup g; $\hat{\phi}(g)\hat{P}(z;\alpha;g)$ and $\frac{\hat{\phi}(g)\hat{P}(z;\alpha;g)}{\hat{P}(z;\alpha)}$ are the estimated absolute and relative contributions to total poverty by subgroup g.

GE decomposition takes the following form:

$$\hat{I}(\theta)=\sum_{k=1}^{K}\hat{\phi}(k)\left(\frac{\hat{\mu}(k)}{\hat{\mu}}\right)^{\theta}\bullet\hat{I}(k;\theta)+\hat{\bar{I}}(\theta) \tag{2.9}$$

Where: K refers to the total number of population groups; $\hat{\phi}(k)$ is population share of subgroup k; $\hat{\mu}(k)$ is the mean of the selected indicator subgroup k; $\hat{I}(k;\theta)$ is the inequality within subgroup k; $\hat{\bar{I}}(\theta)$ is population inequality if each individual in subgroup k is given the mean for the poverty indicator of subgroup k, $\hat{\mu}(k)$.

2.6 Pro-poor growth analysis

In the literature, outcomes of pro-poor growth between any two given periods are analysed by calculating the growth rate (g) and five different pro-poor indices (Duclos & Verdier-Chouchane, 2010). The first three of these indices are measures of absolute pro-poorness and they are: the Ravallion and Chen (2003) index, the Kwakwani and Pernia (2000) index and the PEGR index. The other two indices namely, the Ravallion and Chen (2003) index minus (g) and the PEGR index minus (g) are indices of relative pro-poor growth.

There exist two different approaches to the definition of pro-poor growth, namely a relative and an absolute approach. Growth is defined as pro-poor in the absolute sense if it reduces absolute poverty.

Using the relative approach, growth is pro-poor if reduces inequality and relative poverty. In this sense, the poor proportionately benefit more from growth than the non-poor.

If the growth rate and the Ravallion and Chen (2003), the Kwakwani and Pernia (2000) and the Poverty Equivalent Growth Rate (PEGR) indices are positive, there is absolute pro-poor growth from one period to another. When g is positive and the Kwakwani and Pernia (2000) is negative or when g is negative and the Kwakwani and Pernia (2000) index is positive, then the distributive change has increased absolute poverty. Growth is said to be anti-poor when this is the case. When the Ravallion and Chen (2003) minus g and the PEGR minus g are positive, the distributive change is considered to be relatively pro-poor. A similar conclusion is arrived at if the Kakwani and Pernia (2000)'s index is larger than 1. In this case, growth among the poor is higher than average growth. The poor have, therefore, been favourably affected by the change.

In order to understand Ravallion and Chen (2003)'s growth incidence curves, we, first of all, explain what a "quantile" is. Suppose there are n incomes in a given distribution ranked from the lowest to the highest. A quantile of a given population is given by the income level that is found at a particular rank in that distribution. The rank of the level of income y_i will be given by i/n. Growth incidence in the population can be understood by comparing *quantile* curves before and after a change in a distribution has taken place. Let the pre-change distribution be given by y_A and the post-change distribution be given by y_B, each of equal size n. We can build quantile curves for each of these distributions; these are given by the incomes y_i^A and y_i^B found at different ranks i/n. We can then assess the incidence of growth at any particular rank i/n by comparing the quantile curves at the point i/n. The absolute value change is given by $y_i^B - y_i^A$. The proportional change is given by $\dfrac{y_i^B - y_i^A}{y_i^A}$.

The Ravallion and Chen (2003) growth incidence curve is a plot of the proportional change against all possible values of ranks i/n. The incidence curve shows the rates of growth for various ranks in the distribution. Absolute pro-poorness of growth is obtained when the absolute value change is everywhere positive for the range of ranks over which the initially poor individuals or households are located. Relative pro-poorness of growth is obtained when the growth incidence curve is everywhere above the proportional change in the mean income.

The Kakwani and Pernia (2000) index compares the actual poverty outcome of a distributive change to the outcome that would have been observed if the change had been distribution-neutral (Kakwani & Pernia, 2000). Two main distribution-neutrality criteria are provided. The first one assumes that everyone's income has changed by the same absolute amount while the second one considers that everyone's income has changed by the same proportional amount. There exist several views on what that proportion should be but the most common one is the proportional change in average income.

Suppose P^A and P^B be the actual pre- and post-change poverty levels, and P^{BN} let be post change poverty under distribution neutrality. Then, $\dfrac{P^A - P^B}{P^A - P^{BN}}$ is the ratio of the actual change in poverty to the change that would have been observed under distribution neutrality. Several poverty indices can be chosen for P . In the main text, P is specified as the headcount ratio. Several scenarios of distribution neutrality can also serve to specify BN. Let for instance $A = y_1^A, y_2^A, ..., y_n^A$ and $B = y_1^B, y_2^B, ..., y_n^B$. Kakwani and Pernia (2000)'s index uses the following definition for BN:

$$BN = \left(\frac{\mu_B}{\mu_A} y_1^A, \frac{\mu_B}{\mu_A} y_2^A, \frac{\mu_B}{\mu_A} y_n^A \right) \tag{2.10}$$

It says that a change is distribution neutral if incomes change in proportion to the proportional change in average income. This index thus gives the ratio of the observed change in poverty to the change that would have been observed under constant inequality.

The Poverty Equivalent Growth Rate (PEGR) index, also called the Kakwani, Khandker and Son (2003) index, assesses the pro-poorness of growth by calculating "poverty equivalent growth rates". PEGR is the growth rate that would have resulted in the same observed level of poverty change if the distribution of income shares had not changed. PEGR can be thought of as the counterfactual distribution of income

$BN = ((1 + PEGR)y_1^A, (1 + PEGR)y_2^A, ..., (1 + PEGR)y_n^A)$ as giving the same final level of poverty as the one that is actually observed. When the growth rate PEGR is applied to all of the initial income y^A, poverty thus equals poverty with the distribution of y^B.We, therefore, have $P_B = P_{BN}$ and thus that $P_A - P_B = P_A - P_{BN}$

If PEGR is greater than 0, the distributive change is judged to be absolutely pro-poor by this approach. This is the case if and only if $P_A - P_B > 0$.

Let $g = \frac{\mu_B - \mu_A}{\mu_A}$ and it is thus the actual rate of growth in average income. If income shares remain constant in the movement from A to B, then we must have that $y_i^B = (1+g)y_i^A$ for all i. Since $P_A - P_{BN}$, with constant income shares, it must be that $PEGR = g$. The poverty equivalent growth rate is, therefore, just the usual growth rate if inequality has remained unchanged. Movements in inequality will, however, create a divergence between the poverty equivalent growth rate and the usual growth rate. The greater the adverse effects of inequality on the poor, the greater the value of P_B, and therefore the lower the value of $PEGR$.

The difference, $PEGR - g$, can, therefore, help assess whether the distributive change has affected the income shares of the poor. If $PEGR - g$ is negative, growth among the poor is lower than average growth, and the income shares of the poor have therefore been adversely affected by the change. The converse is true when $PEGR - g$ is positive.

Although pro-poor growth analysis based on non-monetary dimensions has been shown to yield important results, there exist a number of potential problems of extending pro-poor growth analysis to non-monetary dimensions of welfare. Grosse et al. (2008) provide a good discussion of some of these potential limitations. In the context of this study, the main challenge relates to comparing relative changes in the asset index score or HAZ score on a linear scale. For example, in our data sets, HAZ scores range from -6 to 6 which clearly includes negative, 0 and positive values. This raises the question of how to compare a relative change from a score of -6 in 1992 to -5 in 2010 with an improvement from 5 to 6 over the same period. We overcome this challenge by transforming the z-scores into percentiles in the distribution. Similarly, we convert the asset index scores so that they are all positive by adding to the asset index a value slightly higher than the absolute value of the most negative number. The conversions and transformations of the HAZ and asset index scores are further discussed in Section 2.7.

2.7 Data

Malawi has participated in four nationally representative Demographic Health Surveys (DHSs). These surveys have provided up-to-date information on living conditions and health programmes in Malawi. The surveys were conducted by the National Statistical Office (NSO) and the Community Health Sciences Unit (CHSU) during different times of the calendar year. Traditionally, DHSs consist of data sets relating to the household, men, women and children. This study makes use of the household and children's data sets only. Child nutritional status is particularly important because it affects their growth and development, which has a direct link to their future health status as adult men and women. The household data set contains information on assets which are found to be more reliable than consumption expenditure and income as a measure of the long-term living standards for households, as we explain shortly. Table 2.1 provides a summary of the DHS data sets used in the study. Detailed tables showing the distribution of households and children by the different subgroups are respectively presented in Tables A1 and A2 in the appendices.

Table 2.1: Summary of Malawi DHS data sets used

Year	Survey period	No. of households	No. of children (0-59 months)
1992	September-November 1992	5,323	3,353
2000	July-November 2000	14,213	9,753
2004	October 2004 -January 2005	13,664	8,707
2010	June-November 2010	24,825	4,801

Source: Own computation from MDHS data

From the data sets, we derive two non-monetary measures of welfare, namely household asset index and child nutritional status in terms of anthropometric indices of children. The number of children indicated in the table excludes children whose age, height and weight measurements are missing. Poverty and inequality measurement is based on the 2010 data sets only. On the other hand, pro-poor growth analysis is done using all the DHS survey periods from 1992 to 2010.

As earlier discussed, in the literature the use of assets as a measure of economic status has been found to be advantageous over household income and consumption. Wealth[6] is not only said to represent a more permanent status than income and expenditure but is also more easily measured with only a single respondent required in most cases (Rutstein & Johnson, 2004).

Since wealth cannot be directly observed, asset variables have to be identified to proxy wealth. There exists no best approach for selecting which asset variables to use (Montgomery et al., 2000). Although the choice of asset variables has varied by author, generally all assets and utility services reflecting the economic status of a household need to be included. This broader criterion, as opposed to selecting a few assets, is preferred because using a large number of variables potentially prevents a situation where households are concentrated on certain index scores (Rutstein & Johnson, 2004). In adopting the broad criterion, the study uses 31 asset variables and access to utilities including household assets, means of transport, source of lighting, ownership of livestock, agricultural land, having separate room for kitchen, rooms for sleeping, source of drinking water, type of toilet facility, type of material for the main floor, main wall material, main roof material and type of cooking fuel.

[6] In the literature, wealth and asset index are loosely used synonymously because the assets owned by a household represent its asset wealth.

The derivation of the asset index requires that the indicator variables for asset ownership[7] be captured or transformed into binary form, e.g. 1 for "yes" if a household owns a given asset and 0 for "no" if a household does not own the asset. It is, however, not necessary to transform variables that are already categorical such as the source of drinking water, type of toilet facility, floor, wall, roof material and source of cooking fuel. Stata 13.1, the software used, recognises the categorisation automatically even in cases where there are more than two categories. After categorisation in the manner explained, the *mca* command was applied to the asset variables. The index is weighted by household size and sample weight. Principal components analysis (PCA) and factor analysis (FA) were used as robustness checks.

Results from MCA showed that assets and living conditions associated with good higher economic status contribute negatively to the index while those associated with low economic status contribute positively to the index. For example, ownership of a durable asset (e.g. radio, television, refrigerator, etc.) and having piped water, flush toilet, carpet floor and good source of lighting and cooking fuel, among others, contribute negatively to the asset index. On the other hand, not owning a durable asset and living in poor conditions contribute positively to the asset index. The negative signs imply that the index generated from MCA is not an asset index but rather a poverty index. This implies that if quintiles were to be created from the index, the poorest households would be in the fifth quintile and the richest households in the first quintile. This complicates the interpretation of the index and the suggested solution in the literature is to generate an asset index by multiplying the poverty index by (-1). When the transformation is done, the richest households would now be in the fifth quintile and the poorest households in the first quintile, giving the normal interpretation of an asset index. This approach is motivated by Greenacre (2007) and adopted in the other studies (e.g. Da Maia, 2012). In this case, the poverty and asset indices are negative opposites of each other.

Asset and poverty indices typically contain negative values which are unsuitable for poverty and inequality measurement. Following the literature, we adjusted the asset index by adding to it a value slightly higher than the absolute value of the most negative number (e.g., Da Maia, 2012). In our case, since -1.056288 was the minimum value, we added 1.056289 to the asset index. This results in a shift of the distribution of the asset index to the right. The resultant asset index, which we call the adjusted asset index, is what is used for poverty and inequality analysis. Some selected summary statistics for the poverty and asset indices are provided in Table 2.2.

[7] In this study, asset ownership is used to mean ownership of private assets as well as access to public services.

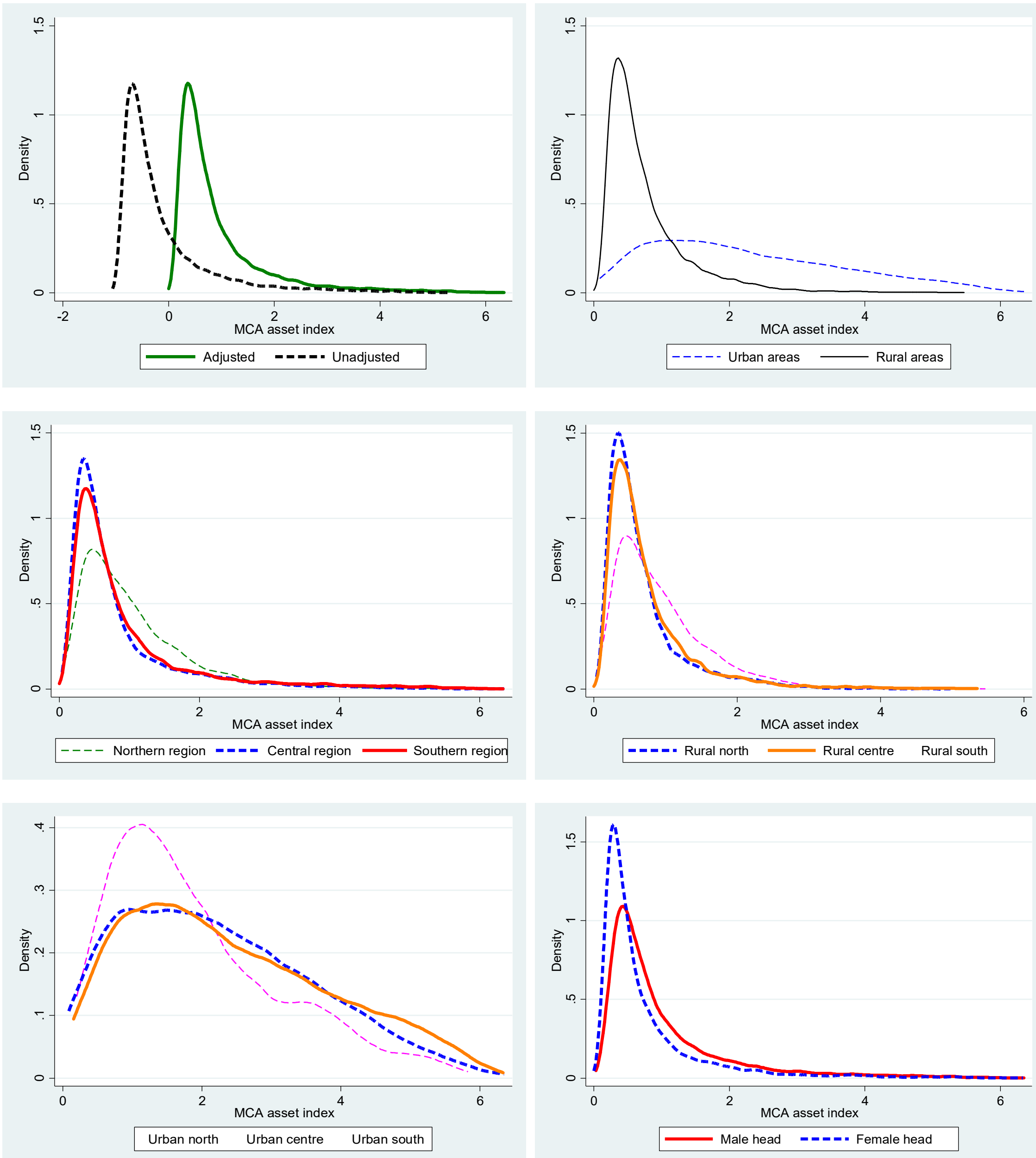
Table 2.2: Descriptive statistics for the poverty and asset indices

Description	Poverty index	Asset index	Adjusted asset index
Percentiles			
1%	-3.65	-0.92	0.14
5%	-1.84	-0.85	0.20
10%	-1.00	-0.80	0.26
25%	-0.08	-0.67	0.38
50%	0.43	-0.43	0.63
75%	0.67	0.08	1.13
90%	0.80	1.00	2.06
95%	0.85	1.84	2.90
99%	0.92	3.65	4.71
Statistic			
Minimum	-5.29	-1.06	0.00
Maximum	1.06	5.29	6.34
Mean	0.11	-0.11	0.94
Standard deviation	0.91	0.91	0.91
Variance	0.82	0.82	0.82
Skewness	-2.35	2.35	2.35
Kurtosis	9.36	9.36	9.36
Observations	24,825	24,825	24,825

Source: Own computation from MDHS 2010

In Figure 2.1, we show the adjusted and unadjusted asset indices in addition to some kernel density plots across population sub-groups. The results show that urban areas have a higher average of asset index scores compared to rural areas. Amongst regions, the Northern region has the highest asset score values seconded by the Central region and finally the Southern region. A similar trend is observed amongst Rural north, Rural centre and Rural south. On the other hand, Urban centre and Urban south seem to be doing better than Urban north.

Figure 2.1: Adjusted and unadjusted asset indices by population subgroups



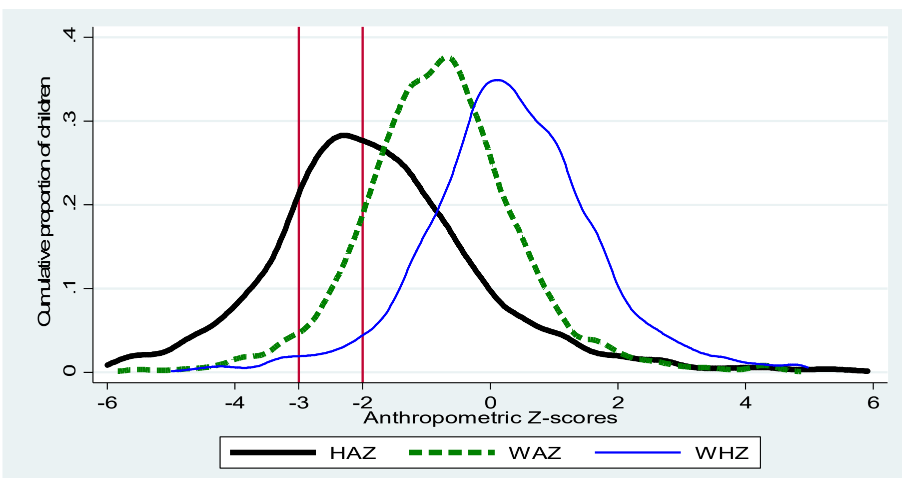
Source: Own computation from MDHS 2010

With respect to our second poverty measure, i.e. child- nutritional status, we calculate three anthropometric indices or z-scores, namely height-for-age (HAZ), weight-for-age (WAZ) and weight-for-height (WHZ) for children aged between 0 and 59 months. We discuss each of these three measures separately in the next few paragraphs.

Firstly, HAZ measures stunted growth and reflects cumulative linear growth and failure to receive adequate nutrition over a long period. HAZ, therefore, indicates the long-term effects of malnutrition in a population. Secondly, WHZ is a measure body mass in relation to body height or length and reflects the current nutritional status. It can be used to describe body wasting which represents the failure to receive adequate nutrition in the period immediately preceding the survey. Wasting may result from inadequate food intake or recent episodes of illness causing loss of weight and the onset of malnutrition. Finally, WAZ (a composite index of HAZ and WHZ) gives the overall malnutrition level and takes into account both acute and chronic malnutrition. It measures the state of body weight.

The z-scores are calculated using the new 2006 World Health Organisation (WHO) child growth standards based on the Multicentre Growth Reference Study done on a 'healthy' sample size of 8,440 children drawn from six countries across the world. The analysis is done using the *zscore* 06 module[8]. The z-scores express normal and abnormal departures of an individual child's height or weight from the average height or weight of comparable children of the same sex and age in the standard reference population. According to WHO (2006), z-scores below -2 standard deviations (SD) and -3 SD from the median of the reference population indicate malnourishment and extreme malnourishment, respectively. Normal children have z-scores of greater than or equal to -2 SD (>=-2SD). Figure 2.2 provides kernel density estimates for HAZ, WAZ and WHZ. Summary descriptive statistics are provided in Table 2.3.

Figure 2.2: Distribution of anthropometric Z-scores for HAZ, WAZ and WHZ



Source: Own computation from MDHS 2010

Levels of child malnutrition seem to depend on the choice of measure. HAZ shows the highest levels of child malnutrition, whereas WAZ and especially WHZ appear to reflect lower levels. Child malnutrition rates for the sample (i.e. scores more than 2 standard deviations below the mean for the standard

---

[8] zscore06: Stata command for the calculation of anthropometric z-scores using the 2006 WHO child growth standards; http://www.ifpri.org/staffprofile/jefleroy.

reference population) stand at 46%, 14% and 4% as based on HAZ, WAZ and WHZ, respectively. Extreme child malnutrition (more than 3 standard deviations below the reference population mean) stands at 19%, 3% and 2% based on HAZ, WAZ and WHZ, respectively.

Table 2.3: Child malnutrition rates by population groups

Description	HAZ		WAZ		WHZ	
	<-2SD	< -3SD	<-2SD	< -3SD	<-2SD	< -3SD
Age						
0-23	38.60%	18.20%	13.50%	3.70%	6.40%	2.60%
24-59	50.70%	19.80%	13.80%	3.30%	2.30%	0.70%
Sex						
Male	49.00%	22.20%	14.80%	3.10%	4.40%	1.80%
Female	42.20%	16.20%	12.60%	3.80%	3.80%	1.30%
Areas						
Urban	39.40%	15.30%	11.40%	3.20%	2.40%	0.60%
Rural	46.60%	19.80%	14.10%	3.50%	4.40%	1.70%
Region						
Northern	42.50%	17.90%	12.70%	2.50%	2.70%	0.50%
Central	45.40%	18.50%	14.10%	4.00%	4.40%	1.80%
Southern	46.40%	20.00%	13.40%	3.20%	4.10%	1.50%
Residence						
Rural North	42.50%	18.30%	13.80%	2.50%	3.00%	0.60%
Rural Centre	46.30%	19.20%	14.30%	4.00%	4.60%	2.00%
Rural South	48.10%	20.80%	13.90%	3.30%	4.50%	1.70%
Urban North	41.70%	14.90%	3.40%	2.50%	0.80%	0.00%
Urban Centre	40.50%	14.40%	13.20%	3.60%	2.80%	0.60%
Urban South	38.10%	16.20%	11.10%	2.80%	2.20%	0.60%
Total	45.50%	19.10%	13.70%	3.50%	4.10%	1.50%

Source: Own computation from MDHS 2010

The level of malnutrition is higher amongst children aged above 24 months compared to those below 24 months. We follow the WHO age-group comparison although a more detailed analysis can be done across smaller age groups. Malnutrition levels are also lower amongst girls compared to boys. Urban residents have a lower incidence of child malnutrition compared to rural residents. Amongst regions, child malnutrition is the highest in the Southern region followed by the Central region. Similar observations are made across areas and regions combined.

However, it is worth noting that the gap in the levels of child mal-nutrition across regions, areas and regions are smaller compared to those obtained by age, sex and area; we get very similar incidence levels of child mal-nutrition amongst the three regions, rural areas (Rural north, Rural Centre and Rural South) and urban areas (Urban north, Urban Centre and Urban South). This seems to suggest that Malawi is uniform in terms of the incidence of child malnutrition. On the other hand, we get a different picture with assets where the incidence levels seem much different not only between areas but also across regions and by sex of the household head.

2.8 Poverty lines

Based on household expenditure data, 50.7% of households in Malawi live below the poverty line. The asset poverty line is set at an equivalent of 50.7% since it is assumed that the appropriate asset poverty line should place the same proportion of households in poverty. The absolute poverty line for the asset index is, therefore, the value at the 50.7th percentile and this turns out to be 0.63399.

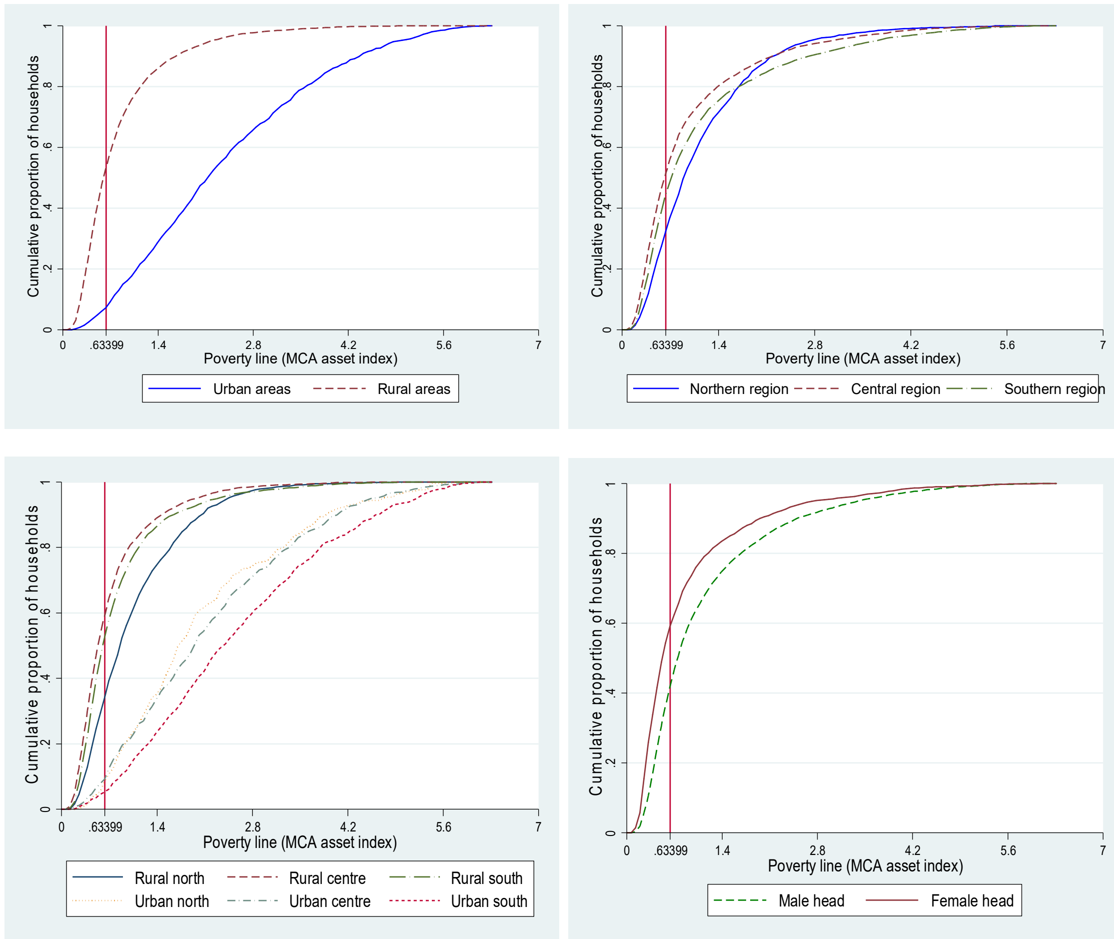
With respect to child nutritional status, we convert each of the three anthropometric z-scores into percentiles. This approach has also been previously used by Mussa (2010). The conversion involves calculating the area under the standard normal curve to the left of the z-score. The area under the curve adds up to unity and the mean (z-score of 0) splits the area into two halves of 0.5 each. The conversion is monotonic and does not affect the ranking of the children. Therefore, for each point of the SDs, there is a corresponding percentile or cumulative probability which is fixed along the x-axis. For example, if we look up in the standard normal distribution tables, a z-score of -2 gives 0.0228 as the area to the left of -2 or simply the 2.3rd percentile. Similarly, -3 corresponds to 0.0013 or the 0.13th percentile.

In this study, we are interested in malnutrition (<-2 SD) as opposed to extreme malnutrition (<-3 SD) and, therefore, use 2.3 as our poverty line. A percentile gives the value of a variable below which a certain percentage of observations (or population) falls. In our case, as we can see from Figure 2.2, almost half of the data points for HAZ lie to the left of -2. We also note that all our three anthropometric measures (HAZ, WAZ and WHZ) follow the standard normal distribution pattern.

2.9 Cumulative density curves

Cumulative density curves (CDCs) indicate how poverty incidence varies with the level of poverty lines but are also used to test for dominance between two distributions. We show the CDCs for the asset index (when $\alpha = 0$) in Figure 2.3.A distribution whose curve lies above the other reflects a higher level of poverty. The figures generally show that there is poverty dominance in the poverty relevant range. The CDCs only cross each other at very high asset levels where it is difficult to conclude that poverty is higher in one population subgroup than the other. However, it is only the poverty relevant range that we are interested in for practical purposes. The figures also show that the gap between the CDCs is largest between urban and rural areas; small differences exist between regions and by sex of household head.

Figure 2.3: MCA asset index cumulative density curves by population groups



Source: Own computation from MDHS 2010

2.10 FGT poverty estimates

Table 2.4 shows our poverty estimates for the three FGT classes, namely the headcount index $(\alpha = 0)$, the average poverty gap index $(\alpha = 1)$ and the poverty severity index $(\alpha = 2)$, respectively. Household observations are weighted by sampling weights and household size. Sampling weights are used so that the chosen households are representative of all households in Malawi. On the other hand, household size takes into account the effect of size on welfare. Thus, a poor household with more members is given a higher weight in the analysis than a similarly poor household with fewer members. For the children's data set, we only apply sampling weights since household size is not applicable for individual level data. Without the use of weights, our results would be either overestimated or underestimated.

The difference between the results in Table 2.4 and Figure 2.3 is that the latter is calculated for the whole range of the poverty lines while the former is calculated at a specific and chosen level of the poverty line. For example with the poverty line set at as asset index of 0.63399, the table shows that 46.0% of the Malawian population is asset poor for $\alpha = 0$. For the same measure (headcount index), asset poverty is higher in rural areas (53.3%) compared to urban areas (7.2%). The Central region has the highest levels of asset poverty amongst the three regions with the incidence of household poverty at 51.3%. A similar observation is made for rural centre and urban centre where 59.4% and 9.2% of the households are living below the poverty line, respectively. Asset poverty is also higher in female-headed households than male headed households.

Child malnutrition estimates are dependent on the type of measure used except for rural-urban population group where rural areas are found to be the poorer than urban areas for all the three indicators, namely HAZ, WAZ and WHZ. At a national level $(\alpha = 0)$, 47.1% of the children are malnourished with respect to HAZ compared to 13.2% and 4.2% for WAZ and WHZ, respectively.

Table 2.4: Poverty headcount, average poverty gap and poverty severity estimates

Description	HAZ			WAZ			WHZ			Asset index		
	α=0	α=1	α=2	α=0	α=1	α=2	α=0	α=1	α=2	α=0	α=1	α=2
Area												
Urban	40.8%	31.0%	26.4%	10.5%	6.7%	5.1%	2.5%	1.8%	1.5%	7.2%	2.4%	1.1%
Rural	48.3%	38.0%	33.2%	13.7%	8.7%	7.0%	4.6%	3.3%	2.8%	53.3%	20.3%	10.1%
Region												
Northern	44.8%	34.6%	30.0%	11.8%	6.8%	5.2%	2.9%	1.5%	1.0%	32.4%	11.0%	5.1%
Central	47.2%	37.2%	32.3%	13.5%	9.0%	7.2%	4.5%	3.4%	2.9%	51.3%	20.6%	10.6%
Southern	47.6%	37.3%	32.6%	13.2%	8.3%	6.5%	4.3%	3.0%	2.6%	44.5%	16.2%	7.7%
Residence												
Rural north	44.9%	35.0%	30.5%	12.9%	7.3%	5.6%	3.1%	1.6%	1.1%	34.5%	11.8%	5.5%
Rural centre	48.1%	38.0%	33.1%	13.8%	9.1%	7.3%	4.8%	3.6%	3.2%	59.4%	24.0%	12.3%
Rural south	49.3%	38.9%	34.1%	13.8%	8.7%	6.9%	4.7%	3.3%	2.8%	52.9%	19.3%	9.2%
Urban north	43.8%	32.0%	26.0%	2.9%	2.3%	2.0%	0.9%	0.6%	0.5%	8.6%	2.0%	0.7%
Urban centre	42.1%	32.4%	27.7%	12.3%	8.4%	6.6%	2.9%	2.1%	1.6%	9.2%	3.1%	1.5%
Urban south	39.2%	29.5%	25.3%	10.0%	5.9%	4.2%	2.3%	1.8%	1.6%	5.3%	1.8%	0.8%
Sex												
Male head	47.1%	36.8%	32.1%	12.9%	8.4%	6.7%	4.3%	3.0%	2.6%	41.7%	14.8%	6.9%
Female head	47.6%	38.5%	34.0%	17.0%	9.3%	6.8%	4.1%	3.2%	2.7%	59.1%	25.8%	13.9%
Malawi	47.1%	37.0%	32.2%	13.2%	8.4%	6.7%	4.2%	3.0%	2.6%	46.0%	17.5%	8.6%

Source: Own computation from MDHS 2010

It is, however, important to state that the national poverty estimates based on HAZ and asset index are similar in magnitude suggesting a similar national poverty profile for Malawi. We also note that the differences in the incidence of poverty between groups (e.g. rural and urban areas) seem to be larger for the asset index than for the three anthropometric measures. Consequently, the results tell us that levels of malnutrition are similar in magnitude across many population subgroups in Malawi. On the other hand, the incidence of asset poverty varies much by population groups and a similar pattern is obtained based on household consumption expenditure studies such as Mussa (2013). Although we have not presented the CDCs of HAZ, WAZ and WHZ, we find them to be much closer to one another than is the case for the asset index in Figure 2.3.

Poverty ranking by population group is the same for all the three indices, $\alpha = 0$, $\alpha = 1$ and $\alpha = 2$. However, this result only holds at the specifically chosen level of poverty line and not all poverty lines. As discussed, it is difficult to conclusively identify which population groups have higher poverty levels because the CDCs cross at different levels. The estimates depend on the choice of the poverty measure used as is the case with child nutritional status. This problem is resolved through stochastic dominance analysis dealt with in Section 2.11 which follows.

2.11 Poverty dominance analysis

Our dominance tests are done up to the third order, as is the tradition in the literature (e.g., Mussa, 2013). Naturally, where first order dominance exists, there is no need to test for second or third order dominance, or where second order dominance exists, there is no need to test for third order dominance, as it would automatically hold. Table 2.5 shows the results. Poverty dominance is measured and interpreted in terms of welfare or standard of living. A distribution that dominates the other has lower levels of poverty than the one that is dominated.

Table 2.5: Poverty stochastic dominance test results for population subgroups

Population group pair	HAZ	WAZ	WHZ	Asset
Area				
Urban v. Rural	U>R: order 2	ND	U>R: order 2	U>R: order 1
Region				
Northern v. Central	ND	ND	N>C: order 2	ND
Northern v. Southern	ND	ND	N>S: order 1	N>S: order 3
Central v. Southern	ND	ND	ND	ND
Residence				
Rural north v. Rural centre	ND	ND	RN>RC: order 3	ND
Rural north v. Rural south	ND	ND	RN>RS: order 1	RN>RS: order 2
Rural centre v. Rural south	ND	ND	ND	ND
Urban north v. Urban centre	UN>UC: order 2	ND	UN>UC: order 2	ND
Urban north v. Urban south	ND	ND	UN>US: order 1	ND
Urban centre v. Urban south	ND	ND	ND	UC>US: order 2
Sex				
Male head v. female head	ND	ND	ND	MH>FH

Source: Own computation from MDHS 2010

Notes: *U=Urban areas, R=Rural areas, ND= no dominance up to third order, UN=Urban north, UC=Urban centre, US=Urban south, N=Northern region, C=Central region, S=Southern region, RN=Rural north, RC=Rural centre, RS=Rural south, MH=male-headed household, FH=female-headed household*

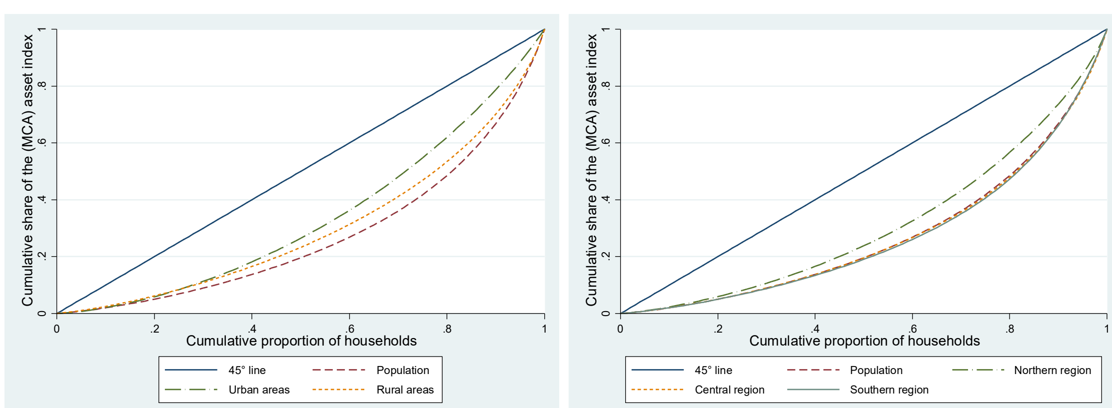
As with the FGT estimates, dominance test results for child nutritional status depend on the measure used. Where dominance is established, the order of dominance is not reversed. For example, all measures establish that urban areas are better in terms of welfare than rural areas and not the opposite. Using HAZ, urban areas dominate rural areas at second order dominance. Likewise, urban north dominates urban centre. No dominance is established for the rest of the subgroups. When WAZ is used, no dominance is established at all for all the pairs of subgroups. Finally, when WHZ is used, non-dominance is found for the Central and Southern regions, Rural centre and Rural south and Urban centre and Urban south. No dominance is established between male and female-headed households- the same finding for HAZ and WAZ. It is, however, important to note that in most of these cases, non-dominance occurs outside the poverty relevant ranges of the poverty line as illustrated in the CDCs.

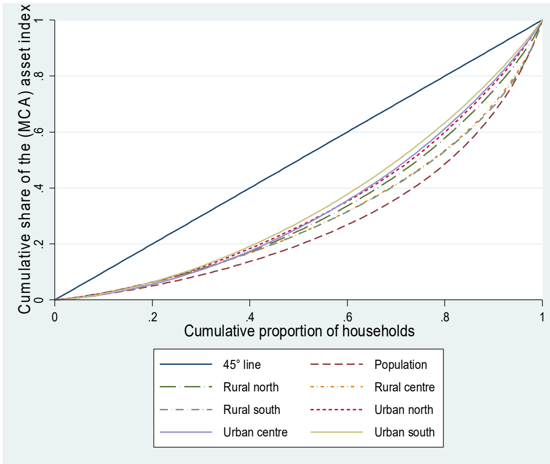
Using the asset index, urban areas dominate rural areas at first order. The Northern region dominates the Southern region. When areas and regions are further broken down, dominance is only established for two pairs, namely rural north dominates rural south and urban centre dominates urban south. Male headed households dominate those headed by females.

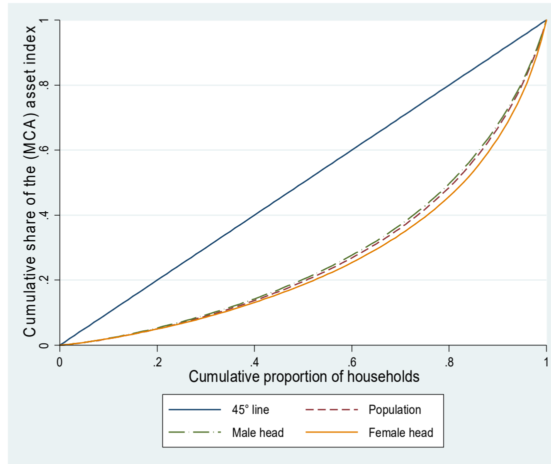
2.12 Gini and GE inequality estimates

In Figure 2.4 are Lorenz curves for Malawi's population groups based on the asset index. The Lorenz curve maps the cumulative share of the asset index, on the vertical axis against the population distribution on the horizontal axis. The 45-degree line indicates perfect equality, a case in which each household has the same share of assets. The numerical inequality estimates for both the asset index and child malnutrition are presented in Table 2.6 that follows.

Figure 2.4: MCA asset index Lorenz curves by population subgroups







Source: Own computation from MDHS 2010

Table 2.6: Inequality estimates across population subgroups

Description	Gini				Theil L (theta=0)				Theil T (theta=1)			
	HAZ	WAZ	WHZ	Asset	HAZ	WAZ	WHZ	Asset	HAZ	WAZ	WHZ	Asset
Area												
Urban	0.683	0.461	0.281	0.326	2.165	0.756	0.279	0.205	0.852	0.364	0.149	0.170
Rural	0.744	0.521	0.313	0.397	2.563	0.889	0.429	0.261	1.058	0.459	0.186	0.267
Region												
Northern region	0.733	0.502	0.290	0.376	2.416	0.787	0.292	0.244	1.018	0.424	0.159	0.230
Central region	0.734	0.513	0.306	0.457	2.547	0.918	0.430	0.353	1.021	0.447	0.181	0.356
Southern region	0.736	0.513	0.314	0.462	2.481	0.840	0.408	0.362	1.029	0.445	0.186	0.361
Residence												
Rural north	0.739	0.512	0.300	0.361	2.487	0.794	0.310	0.225	1.040	0.441	0.168	0.210
Rural centre	0.742	0.519	0.309	0.394	2.608	0.943	0.450	0.257	1.051	0.457	0.183	0.265
Rural south	0.747	0.523	0.321	0.397	2.532	0.851	0.437	0.259	1.070	0.462	0.194	0.272
Urban north	0.680	0.413	0.197	0.337	1.811	0.691	0.116	*0.200*	0.845	0.282	0.075	0.180
Urban centre	0.681	0.475	0.292	0.338	2.189	0.768	0.316	0.220	0.850	0.390	0.166	0.182
Urban south	0.678	0.454	0.281	0.308	2.182	0.752	0.268	0.184	0.842	0.353	0.145	0.152
Sex												
Male head	0.734	0.512	0.308	0.441	2.502	0.872	0.402	0.327	1.022	0.443	0.180	0.325
Female head	0.741	0.519	0.313	0.478	2.528	0.856	0.453	0.385	1.048	0.459	0.186	0.400
Malawi	0.735	0.513	0.308	0.453	2.505	0.871	0.406	0.348	1.024	0.444	0.181	0.346

Source: Own computation from MDHS 2010

The results show that inequality is higher in rural areas compared to urban areas and this is consistent for all the three inequality measures used, namely Gini, Theil L and Theil T. Nevertheless, amongst the three measures of child nutrition, HAZ generally yields the highest levels of inequality. Just as with poverty measurement, levels of inequalities depend on the measure used. Consequently, it is difficult to conclude without ambiguity which population groups have higher levels of inequality. We, therefore, turn to inequality dominance testing as a solution. The inequality dominance test results are presented in Section 2.13 and act as robustness checks.

2.13 Inequality dominance analysis

Results of the Lorenz dominance tests are presented in Table 2.7. Using HAZ, inequality dominance is only established for the urban and rural area pair. Since the Lorenz curve for urban areas lies above that of rural areas, inequality is said to be lower in urban areas compared to rural areas. This finding is confirmed with WHZ as well as the asset index. No dominance is established using WAZ.

Table 2.7: Generalised Lorenz dominance test results across population subgroups

Population group pair	HAZ	WAZ	WHZ	Asset
Area				
Urban v. Rural	U>R: order 2	ND	U>R: order 2	U>R: order 2
Region				
Northern v. Central	ND	ND	N<C	ND
Northern v. Southern	ND	ND	N<S	ND
Central v. Southern	ND	ND	ND	ND
Residence				
Rural north v. Rural centre	ND	ND	ND	ND
Rural north v. Rural south	ND	ND	RN<RS: order 2	RN<RS: order 2
Rural centre v. Rural south	ND	ND	ND	ND
Urban north v. Urban centre	ND	ND	UN<UC: order 2	ND
Urban north v. Urban south	ND	ND	UN<US: order 2	ND
Urban centre v. Urban south	ND	ND	ND	UC<US: order 2
Sex				
Male head v. female head	ND	ND	ND	MH<FH: order 2

Source: Own computation from MDHS 2010

Notes: *U=Urban areas, R=Rural areas, ND= no dominance, UN=Urban north, UC=Urban centre, US=Urban south, N=Northern region, C=Central region, S=Southern region, RN=Rural north, RC=Rural centre, RS=Rural south, MH=male-headed household, FH=female-headed household*

Amongst the three regions, no inequality dominance is established using HAZ, WAZ and asset index. For WHZ, inequality is found to be the highest in the Northern region compared to both the Central and Southern regions. With respect to the sex of the household head, dominance is established for the asset index only, in which case inequality is found to be higher among male-headed households. WHZ and asset index establish dominance between some areas such as Rural north, Rural south, Urban north, Urban centre and Urban south.

2.14 Poverty decomposition

Table 2.8 shows results of the FGT decompositions across rural-urban areas, regions, areas and sex of the household head for the three anthropometric indicators and the asset index. We provide the estimated population share of each subgroup as well as the estimated relative contribution of each subgroup to total poverty. Three FGT measures are used, namely the poverty headcount, average poverty gap and poverty severity indices. The population shares and relative contributions to poverty sum up to unity. The objective here is to establish if the most populated areas are also the ones with the highest incidence of poverty.

Our decomposition results for the asset index indicate that most poor people in Malawi are also in the rural areas-where the population is large. For example, the table shows that 84% of the households are based in rural areas which contribute 97.5% to the poverty headcount when the asset index is used as the measure of welfare. The contributions are slightly higher with the poverty gap and severity indices at 97.8% and 98%, respectively. This does not only indicate that the incidence of poverty is above the national average in rural areas but also that the poorer parts of the population are more concentrated in rural areas. Amongst the three regions, much of the poverty is contributed by the Central region followed by the Northern region. A similar observation is made within rural and urban areas, e.g. Rural north and Rural centre. Households headed by males contribute more to poverty than those headed by females.

Using the nutritional measures, rural areas are also found to contribute more to poverty than urban areas but the difference between the rural population share and the contribution of rural areas to these measures of nutritional poverty is small. This is surprising since it appears as if nutritional status is not much better in urban than in rural areas, despite the fact that asset holdings in urban areas are definitely considerably greater and asset poverty less severe. With respect to the three regions of the country, it is the Central region that contributes the highest to poverty in both absolutes and relative terms. Rural centre and urban centre also contribute more to poverty than the other population subgroups. Finally, just as is the case with the asset index, male-headed households contribute more to poverty than those headed by females.

2.15 Subgroup inequality decomposition

Table 2.9 presents generalised entropy inequality estimates decomposed to indicate within (vertical) and between (horizontal) subgroup inequalities. Irrespective of the measures used, inequalities are largely driven by within population subgroups as opposed to between population subgroups. Between sub-group inequality is only a greater magnitude for the urban-rural areas comparison using the Theil L or Theil T measure for asset inequality, when it contributes somewhere between one-quarter and one-third to overall asset inequality. All other between group contributions to inequality are extremely small, reflecting the fact that locational differences in nutritional status appear to be non-systematic.

Table 2.8: FGT poverty sub-group decomposition

Description	Population share	HAZ			WAZ			WHZ			Asset index		
		α=0	α=1	α=2	α=0	α=1	α=2	α=0	α=1	α=2	α=0	α=1	α=2
Area													
Urban	15.0%	13.0%	12.6%	12.4%	12.6%	12.9%	12.6%	8.6%	8.8%	8.4%	2.5%	2.2%	2.0%
Rural	85.0%	87.0%	87.4%	87.6%	87.4%	87.1%	87.4%	91.4%	91.2%	91.6%	97.5%	97.8%	98.0%
Region													
Northern	10.7%	10.0%	9.8%	9.8%	10.0%	9.4%	9.4%	7.1%	5.1%	4.3%	8.4%	7.5%	7.0%
Central	46.4%	46.3%	46.5%	46.4%	47.8%	48.9%	49.1%	49.6%	52.3%	53.0%	48.4%	51.1%	53.0%
Southern	42.9%	43.7%	43.7%	43.8%	42.3%	41.7%	41.5%	43.3%	42.7%	42.7%	43.3%	41.4%	40.0%
Residence													
Rural north	9.6%	9.0%	8.9%	8.9%	9.7%	9.1%	9.0%	6.9%	4.8%	4.1%	8.2%	7.4%	6.9%
Rural centre	39.6%	40.2%	40.5%	40.5%	41.3%	42.2%	42.7%	45.1%	47.8%	49.1%	47.0%	49.9%	51.8%
Rural south	35.8%	37.7%	38.0%	38.2%	36.4%	35.8%	35.8%	39.4%	38.5%	38.4%	42.3%	40.6%	39.2%
Urban north	1.1%	1.0%	1.0%	0.9%	0.3%	0.4%	0.4%	0.2%	0.2%	0.2%	0.2%	0.1%	0.1%
Urban centre	6.7%	6.0%	5.9%	5.8%	6.4%	6.6%	6.4%	4.5%	4.5%	3.9%	1.4%	1.3%	1.2%
Urban south	7.2%	6.0%	5.7%	5.6%	5.9%	5.9%	5.7%	3.9%	4.1%	4.3%	0.9%	0.8%	0.7%
Sex													
Male head	91.9%	91.9%	91.6%	91.5%	90.0%	91.4%	91.9%	91.6%	90.6%	90.4%	68.5%	63.9%	60.6%
Female head	8.1%	8.1%	8.4%	8.5%	10.0%	8.6%	8.1%	8.4%	9.4%	9.6%	31.5%	36.1%	39.4%
Malawi	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

Source: Own computation from MDHS 2010

Table 2.9: GE inequality decomposition for asset index and child nutrition

Description	Sub-group	Areas		Regions		Within areas/regions		Household head	
GE index	Type	Absolute	Relative	Absolute	Relative	Absolute	Relative	Absolute	Relative
Asset Theil L	Within	0.254	72.2%	0.349	99.0%	0.248	70.3%	0.344	97.5%
	Between	0.098	27.8%	0.004	1.0%	0.105	29.7%	0.009	2.5%
	Total	0.353	100.0%	0.353	100.0%	0.353	100.0%	0.353	100.0%
Asset Theil T	Within	0.236	66.9%	0.348	99.0%	0.230	65.3%	0.344	97.6%
	Between	0.116	33.1%	0.004	1.0%	0.122	34.7%	0.008	2.4%
	Total	0.352	100.0%	0.352	100.0%	0.352	100.0%	0.352	100.0%
HAZ Theil L	Within	2.503	99.9%	2.505	100.0%	2.501	99.9%	2.505	100.0%
	Between	0.002	0.1%	0.000	0.0%	0.003	0.1%	0.000	0.0%
	Total	2.505	100.0%	2.505	100.0%	2.505	100.0%	2.505	100.0%
HAZ Theil T	Within	1.022	99.8%	1.024	100.0%	1.020	99.7%	1.024	100.0%
	Between	0.002	0.2%	0.000	0.0%	0.003	0.3%	0.000	0.0%
	Total	1.024	100.0%	1.024	100.0%	1.024	100.0%	1.024	100.0%
WAZ Theil L	Within	0.869	99.7%	0.871	99.9%	0.868	99.6%	0.871	100.0%
	Between	0.002	0.3%	0.001	0.1%	0.004	0.4%	0.000	0.0%
	Total	0.871	100.0%	0.871	100.0%	0.871	100.0%	0.871	100.0%
WAZ Theil T	Within	0.442	99.5%	0.444	99.9%	0.441	99.2%	0.444	100.0%
	Between	0.002	0.5%	0.001	0.1%	0.004	0.8%	0.000	0.0%
	Total	0.444	100.0%	0.444	100.0%	0.444	100.0%	0.444	100.0%
WHZ Theil L	Within	0.406	100.0%	0.406	100.0%	0.406	99.9%	0.406	100.0%
	Between	0.000	0.1%	0.000	0.0%	0.000	0.1%	0.000	0.0%
	Total	0.406	100.0%	0.406	100.0%	0.406	100.0%	0.406	100.0%
WHZ Theil T	Within	0.180	99.9%	0.180	99.9%	0.180	99.7%	0.180	100.0%
	Between	0.000	0.2%	0.000	0.0%	0.000	0.2%	0.000	0.0%
	Total	0.181	100.0%	0.180	100.0%	0.180	100.0%	0.180	100.0%

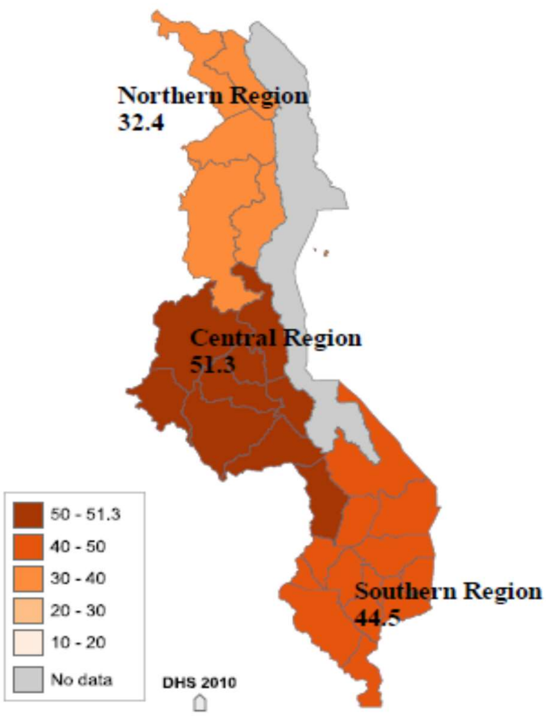
Source: Own computation from MDHS 2010

2.16 Spatial distribution of poverty and inequality

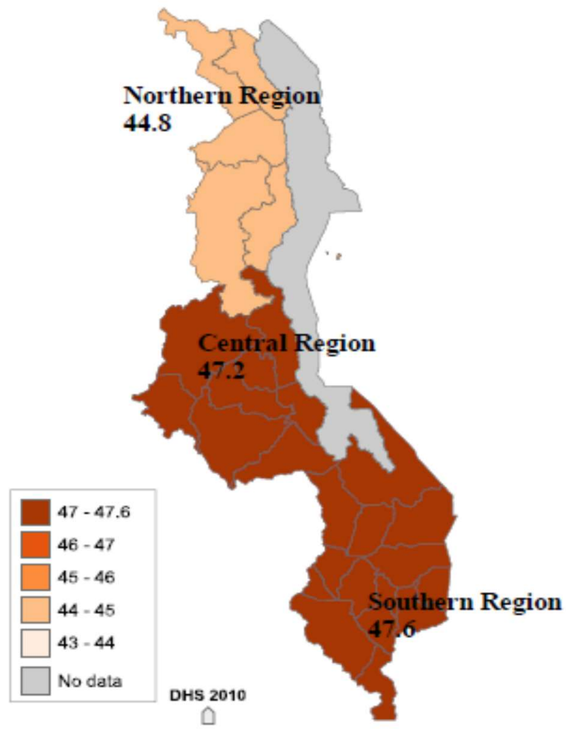
The maps are drawn to provide a visual representation of the FGT headcount and Gini inequality estimates presented in Table 2.4 and Table 2.6. They are drawn for Malawi's three regions (Northern, Central and Southern) using StatPlanet, an interactive data visualisation and mapping software[9]. The colours on the maps vary positively with the levels of poverty and inequality. Therefore, darker and lighter colours represent areas with higher and lower levels of poverty and inequality, respectively. Figure 2.5 and Figure 2.6 provide the spatial distribution of poverty and inequality using our four selected indicators, namely the asset index, HAZ, WAZ and WHZ. The mapping shows little variation in terms of both poverty and inequality estimates across Malawi's regions.

Figure 2.5: Spatial distribution of poverty using asset index and child-nutritional status

Poverty headcount (asset index)

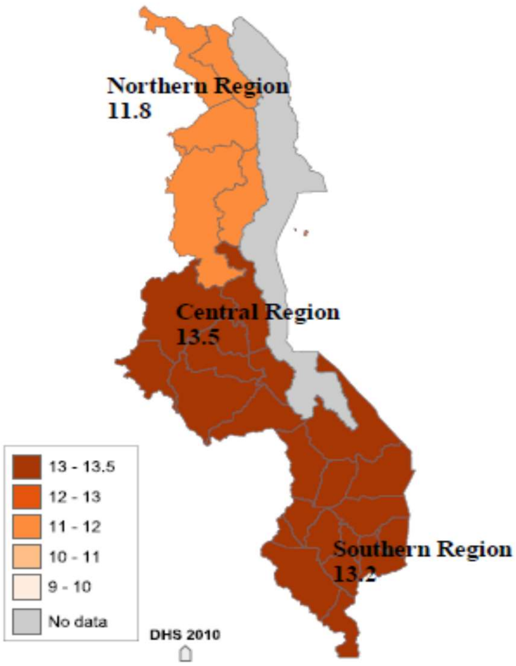


Poverty headcount (HAZ)

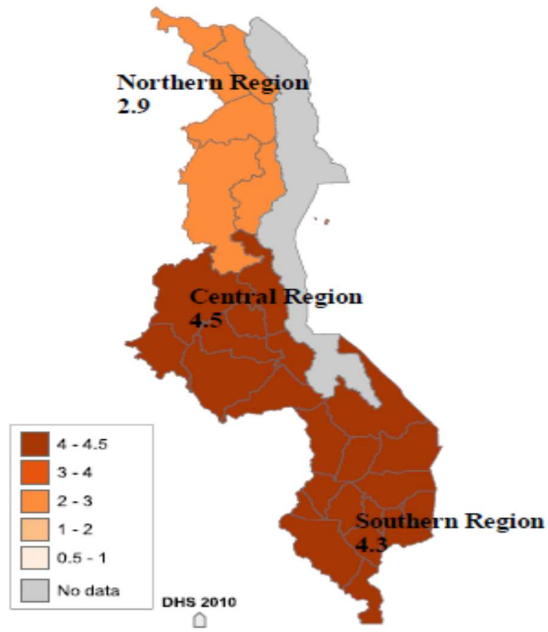


[9] We multiply our estimates by 100 for easy customisation of the map colours in StatPlanet.

Poverty headcount (WAZ)



Poverty headcount (WHZ)



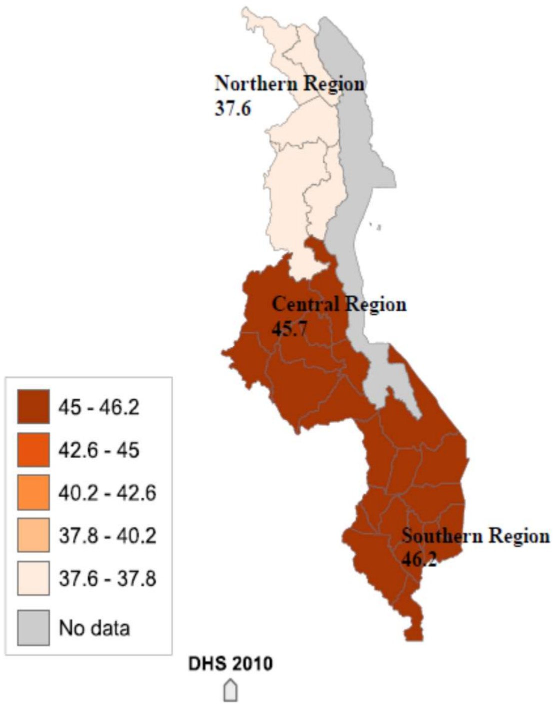
Source: Own computation from MDHS 2010

The maps also confirm that regardless of the indicator used, the poverty incidence is the lowest in the Northern region. The ranking between the Central and Southern regions is dependent on the indicator used. Based on the asset index, WAZ and WHZ, the Central region has the highest incidence of poverty. On the other hand, the Southern region has the highest incidence of poverty when HAZ is used as the indicator.

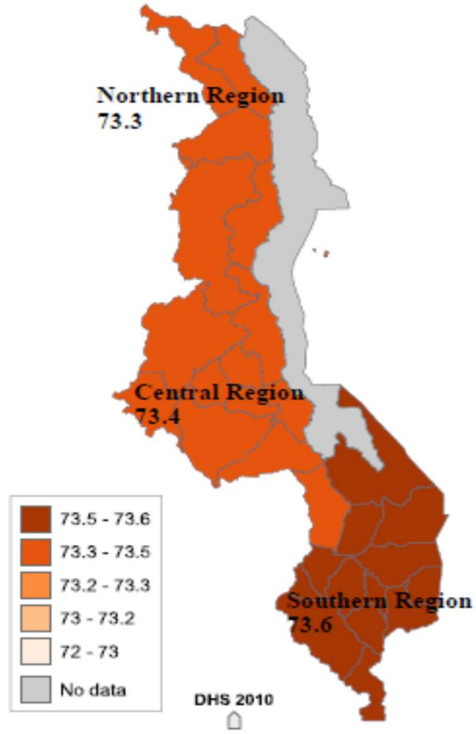
With respect to Gini mapping, it is also shown that inequality is lowest in the Northern region for all the measures used. The Central and Southern regions fairly rank the same for all the measures used apart from HAZ where inequality is highest in the Southern region. This ranking is similar to the one found in the poverty headcount when HAZ is used.

Figure 2.6: Spatial distribution of inequality for asset index and child-nutritional status

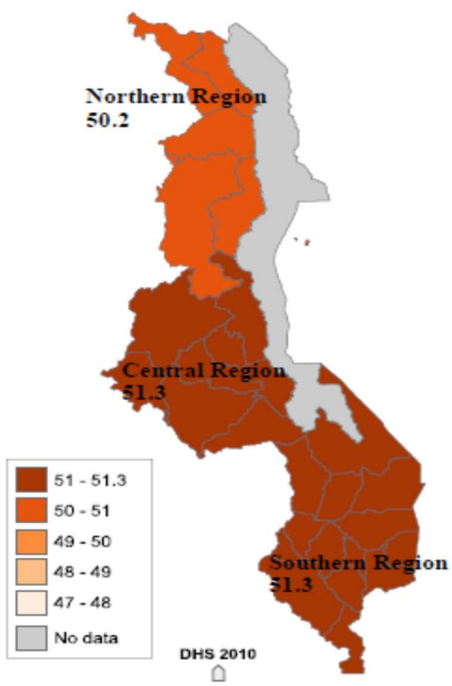
Gini inequality (asset index)



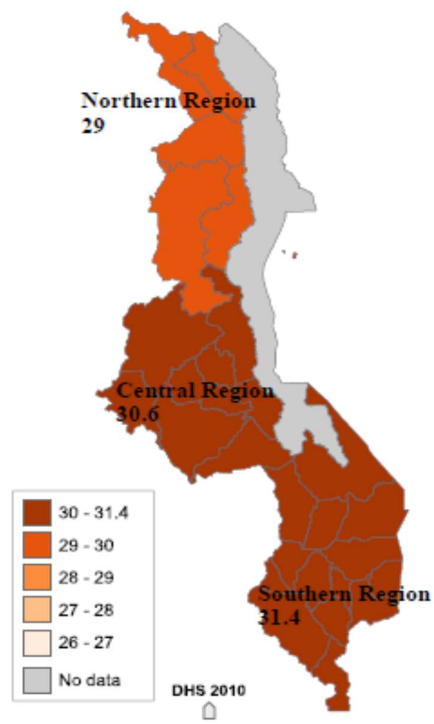
Gini inequality (HAZ)



Gini inequality (WAZ)



Gini inequality (WHZ)



Source: Own computation from MDHS 2010

2.17 Factors affecting asset poverty

So far, we have only looked at the poverty profile for Malawi without explaining the factors associated with poverty. As acknowledged in the literature, a satisfactory explanation of why households are poor is important if we are to be able to deal with the roots of poverty (Haughton & Khandker, 2009). This chapter, therefore, addresses the question of what factors are associated with asset poverty in Malawi using the 2010 MDHS data. Our dependent variable is the asset index, reflecting long-term economic status. The choice of the explanatory variables is informed by the literature.

The correlates include household size, age dependency ratio, age of the household head, sex of the household head (male=1, female=2), incidence of sickness in the household (no=0, yes=1) and the levels of education in the household. We also include dummies for area (urban=1, rural=2) and region (1=Northern, 2=Central and 3=Southern region) to control for the location in which the households reside. The educational status of household members has four categories, namely 0=no education and preschool, 1=primary education, 2=secondary education and 3=post-secondary education. Shocks to the household are represented by the incidence of sickness in the household, captured as (no=0, yes=1).Table 2.10 shows descriptive statistics of the variables used in our analysis of asset poverty.

Table 2.10: Summary descriptive statistics for the asset model

Variables	Obs	Mean/Prop.	Std. Dev.
Asset index	24,825	0.95	0.903
Age of head	24,798	43.34	16.368
Household size	24,825	4.79	2.295
Dependency ratio	24,825	49.2%	0.245
Female head	24,825	28.4%	0.451
Sickness in the household	19,947	2.5%	0.157
Household's education			
Primary	24,724	61.0%	0.488
Secondary	24,724	17.3%	0.378
Higher	24,724	2.6%	0.160
Rural area	24,825	88.3%	0.322
Region			
Central	24,825	33.7%	0.473
Southern	24,825	48.8%	0.500

Source: Own computation from MDHS 2010

The proportions relate to categorical variables. The asset index has a mean value of about 0.95 and standard deviation of 0.903. On average, the household head is aged about 43 years with standard deviation of 16.368. Based on the sample, the average household size in Malawi is about 4.79 persons. Based on an alternative data set, the Malawi IHS3 data, the average household size is 4.6 members. The age dependency ratio stands at 49% and is calculated as the number of children aged 0 to 14 years and the number of persons aged 65 years divided by the household size[10]. About 28% of the households are headed by females. With respect to the incidence of sickness, only about 2.5% of the households reported not to have been very sick for 3 or more months in the year prior to the survey. With regards to the levels of education, most households (about 61%) have primary education as the highest level of qualification. This is followed by no education or preschool education at 19%, secondary education at 17.3% and finally post-secondary education at 2.6%. With regard to area of residence, about 88% of the people reside in households which are rural compared to 12% which are urban. The Southern region has the highest number of households with 49%, followed by the Central region with 34% and finally the Northern region with 17%.

Table 2.11 presents the OLS regression results of the asset index. The results show that about 55% of the variance in the asset index is explained by the explanatory variables. All the variables have the expected signs and statistically significant at conventional levels. There is also satisfactory performance for our control variables.

We checked for the robustness of the model based on an alternative definition of the age dependency ratio and use of sickness of the father and mother in place of sickness for any of the household members. Our preferred model, which is presented, was chosen because it gives the largest size of R-squared and age of the household head had the expected positive sign.

The coefficient for household size is positive and statistically significant. It indicates that a unit increase in household size is associated with an increase of 0.071 standard deviations in the asset index while holding all the other factors constant. Put differently, larger households are associated with more assets. The coefficient for age dependency ratio is also significant coefficient but negative indicating that the greater the proportion of economically inactive members in the household, the poorer is that household in terms of asset ownership. Households that are headed by older people are associated with better long-term economic status by about 0.005 standard deviations. The regression results indicate that the there is no statistically significant difference in the asset ownership between male and female headed households.

[10] Strictly speaking, age dependency ratio is given as the ratio of dependents (people younger than 15 or older than 64) to the working-age population (those aged 15-64). However, this strict definition generates missing values since some of the households do not have people in either of the three age categories.

With respect to education, it is found that higher levels of education are positively associated with asset ownership in Malawi when compared to the base category (no education). It is worth noting with great interest that the size of the education coefficient increases with the level of education. Urban areas are associated with higher levels of asset ownership than rural areas by about 1.051 standard deviations. Regional dummies have mixed performance with the Central region being statistically different from the Northern region which is the base. Controlling for the 27 districts in the data does not add much to the model so we do not report in the tables.

Table 2.11: OLS regression results for asset poverty

Description	Asset index	SE
Household size	0.071***	(0.005)
Age dependency ratio	-0.505***	(0.047)
Age of household head	0.005***	(0.001)
Sex of household head	0.038**	(0.015)
Household member is sick	-0.053	(0.034)
Household education		
Primary	0.236***	(0.015)
Secondary	0.864***	(0.037)
Higher	2.390***	(0.093)
Rural area	-1.051***	(0.067)
Region		
Central	-0.141***	(0.040)
Southern	0.003	(0.039)
Constant	2.283***	(0.130)
R-squared	0.552	
Observations	19,900	
Prob >F	0.000	
F statistic	190.350	

Notes: *, **, *** denote significance at 10%, 5% and 1% levels

2.18 Child nutritional status in Malawi

We look at factors associated with child nutritional status in Malawi using the 2010 MDHS children data set. A previous study by Ngalawa & Chirwa (2008) was based on the 1997-1998 IHS data set. As earlier discussed, DHS surveys provide a better source of data on anthropometric measurements when compared to IHS data (see Verduzco-Gallo et al., 2014). Furthermore, the present analysis is based on the most recent data. Our dependent variables in the study are HAZ, WAZ and WHZ as indicators of stunting, underweight and body wasting, respectively. Table 2.12 shows descriptive statistics of the variables used in our three nutrition models. The proportions relate to all categorical variables just as before.

Table 2.12: Descriptive statistics for the nutritional models

Variable	Obs	Mean	Std. Dev.
HAZ	4,653	-1.76	1.662
WAZ	4,783	-0.79	1.215
WHZ	4,609	0.31	1.35
Age in months	4,801	28.91	16.903
Square of age in moths	4,801	11.22	10.334
Weight at birth	4,801	5.64	3.271
Parental age difference	4,411	5.52	4.692
Birth order number	4,801	3.69	2.321
Female child	4,801	50.60%	0.50
Rural area	4,801	90.00%	0.299
Child is twin	4,801	3.20%	0.176
Female household head	4,801	7.90%	0.269
Asset index	4,801	-0.097	0.899
Square of asset index	4,801	0.817	3.235
Mother's education			
Incomplete primary	4,801	60.60%	0.489
Complete primary	4,801	9.40%	0.292
Incomplete secondary	4,801	9.30%	0.291
Complete secondary	4,801	3.40%	0.182
Higher	4,801	0.50%	0.068
Father's education			
Incomplete primary	4,717	56.30%	0.496
Complete primary	4,717	8.30%	0.275
Incomplete secondary	4,717	23.90%	0.426
Higher	4,717	1.60%	0.127
Region			
Central	4,801	37.10%	0.483
Southern	4,801	45.50%	0.498

Source: Own computation from MDHS 2010

Explanatory variables include child characteristics such as sex (male=1, female=2), age in months, age squared included to account for the possible non-linearity between age and nutritional status, the weight at birth and the status of being a twin (no=0, yes=1) and absolute birth order. We control for area of residence (urban=1, rural=2), sex of the household head (male head=1, female head=2), mothers' and fathers' education (0/5). The levels of education for the father and mother are represented by dummies representing six categories, namely no education, incomplete primary, complete primary, incomplete secondary, complete secondary and higher education(post-secondary education). Father's education has five categories only. There are no fathers with complete secondary education in the data set.

The age difference between the father and mother to captures the bargaining position of the mother in the household. According to the bargaining literature on household decisions, bargaining status could influence the resources that the mother may receive for herself as well as for her child, possibly leading to adverse nutrition consequences (Linnemayr, Alderman & Ka, 2008). Finally, we include the asset index to capture the economic status of the household to which the child belongs. In the literature, economic status has been found to be a strong determinant of the nutritional status of the children (e.g., Dancer, Rammohan & Smith, 2008).

The table reveals that amongst the three measures, WHZ has the highest mean followed by WAZ. The average age of the children is about 29 months. On average, the children were born with a weight of around 5.64kg which is actually high and in line with our statistic suggesting that being underweight is the least of the malnutrition problems in Malawi. The difference in age between fathers and mothers is about 5.52 years. The average birth order of the children in our study is 3.69. Rural areas constitute a very large proportion of children- about 90% of the total. The incidence of a twin is very low at 3.2% of the total number of children. Only about 8% of the children come from households which are headed by females. The majority of the mothers have incomplete primary education (60.6%) compared to 56.3% for fathers. About 46% and 37% of the children reside in the Southern region Central region, respectively.

2.19 Multivariate analysis of child nutrition

Table 2.13 presents results from the OLS regression analysis of HAZ, WAZ and WHZ[11]. The model statistics show that the R-squared stands at 10%, 9% and 11% for HAZ, WAZ and WHZ, respectively. Though the F-statistics indicate the hypothesis that all slope coefficients are equal to zero, it is important to note that the WHZ model does not perform well. The results do not improve even if we regress by the recommended WHO age categories of <24 months and >=24 months (see Table A3 in the appendix). This finding is in line with an earlier study by Chirwa and Ngalawa (2008) which suggests that children in Malawi seem to be ‘fatter for their age’. There could be other factors such as poor feeding practices or genetic factors that affect their nutritional status or indeed a concern for consistent measurement error.

Child characteristics such as age, sex and twin status are statistically significant for HAZ and WAZ but not WHZ. Birth order is significant for the WAZ and WHZ models only. The relationship between the age of the children and nutritional status is negative and statistically significant for HAZ and WAZ only. The square of age is included to account for possible non-linearity that exists between child nutrition and age. The coefficient of age-squared is positive for HAZ and WAZ and statistically significant at

[11] Area of residence, the weight of the child at birth, parental age difference and sex of the household head are not significant and, therefore, not reported in the results.

1%. This implies that the relationship between age and nutritional status of a child is U-shaped (convexity) with respect to HAZ and WHZ. When WHZ is used, the relationship is inverted U-shaped (concavity) but insignificant. The convexity or concavity indicates that malnutrition in under-five children worsens or improves with age but this is only up to some critical age beyond which a child's nutrition status improves or worsens with age.

Table 2.13: OLS regression results for child nutritional status

	HAZ		WAZ		WHZ	
Variables	Coeff	SE	Coeff	SE	Coeff	SE
Age in months	-0.094***	(0.008)	-0.045***	(0.007)	0.002	(0.007)
Square of age	0.128***	(0.012)	0.052***	(0.010)	-0.001	(0.010)
Female child	0.280***	(0.057)	0.102*	(0.048)	-0.072	(0.049)
Child is twin	-0.930***	(0.169)	-0.901***	(0.160)	-0.286*	(0.145)
Birth order number	0.025	(0.016)	0.029*	(0.014)	0.038**	(0.015)
Mother's education						
Incomplete primary	0.082	(0.099)	0.154*	(0.076)	0.124	(0.083)
Complete primary	0.110	(0.138)	0.057	(0.100)	0.112	(0.118)
Incomplete secondary	0.186	(0.143)	0.302**	(0.114)	0.165	(0.116)
Complete secondary	0.060	(0.192)	0.243	(0.172)	0.369*	(0.166)
Post-secondary	0.439	(0.473)	0.236	(0.384)	-0.049	(0.428)
Father's education						
Incomplete primary	-0.095	(0.124)	-0.191*	(0.082)	-0.048	(0.094)
Complete primary	-0.178	(0.168)	-0.208	(0.111)	0.083	(0.130)
Incomplete secondary	-0.006	(0.139)	-0.202*	(0.101)	-0.092	(0.108)
Post-secondary	-0.055	(0.235)	-0.169	(0.159)	-0.007	(0.211)
Asset index	0.183**	(0.063)	0.195***	(0.052)	-0.002	(0.055)
Square of asset index	-0.011	(0.016)	-0.017	(0.011)	0.008	(0.014)
Rural area	0.127	(0.108)	0.052	(0.094)	-0.105	(0.095)
Region						
Central	-0.031	(0.111)	0.009	(0.075)	-0.102	(0.075)
Southern	-0.006	(0.110)	-0.081	(0.072)	-0.174*	(0.072)
Constant	-1.031***	(0.261)	-0.259	(0.197)	0.368	(0.213)
R-squared	0.101		0.088		0.011	
Prob > F	0.000		0.000		0.010	
Observations	4,574		4,700		4,531	

Notes: *, **, *** denote significance at 10%, 5% and 1% levels

The critical ages are 37 months for HAZ, 44 months for WAZ and 54 months for WHZ. The significance and signs of coefficients for both of age and age squared are also confirmed in Chirwa and Ngalawa (2006) who find 30, 34 and 35 months as critical ages for HAZ, WAZ and WHZ. There exists literature which links recovery from stunting to cognitive outcomes. For example, a panel data study by Casale and Desmond (2015) in South Africa finds that children who were stunted at age 2 (in years) but recovered by age 5 performed better than children who remained stunted over the study period. They also find that children who recover from stunting by age 5 perform significantly worse on cognitive tests than children who do not experience early malnutrition. Moreover, they find that recovery from stunting is not uncommon among children- a similar finding in this study. However, the most important thing is the timing of recovery as well as the need to provide the required nutritional inputs at an early stage if the goal is to help improve children's cognitive performance. Evidence in this research suggests it took longer by between 7 and 19 months for children to 'recover' in 2010 compared to what Chirwa and Ngalawa (2006) found using the 1997/8 data[12].

The coefficient of a variable capturing the status of being one of a twin is negative in all models but only statistically significant for HAZ and WAZ. Mussa (2014) uses HAZ as a measure of long-term nutritional status and finds opposite signs of coefficients for age, age-squared and twin status of a child. With respect to age and age-squared, Mussa (2014) finds concavity in which case the nutritional status of a child improves with age but begins to worsen after some time. This is the same finding in this study for the WHZ model although not statistically significant. We have found that the coefficient for birth order is also positive and statistically significant for WAZ and WHZ. Mussa (2014) finds a significant relationship for rural areas only which seems to confirm the empirical finding that the effect of child-order is cultural and must be interpreted within a given cultural context. With respect to parental age-difference, we find no significant relation for all the three models. On the other hand, Mussa (2014) finds a negative relationship implying that households with older fathers compared to mothers perform worse in terms of child nutritional outcomes.

The level of education of the mother matters more than that of the father in our models for child nutritional status. For both the fathers and mothers, no education at all is used as the base category. All levels of mothers' education except post-secondary education play a significant role in the WAZ model. In the HAZ model it is only incomplete secondary that matters while for WHZ, incomplete primary and complete secondary school levels play a role. The education level of the father is only significant for incomplete primary education and only for the WAZ model. However, this happens at the lowest levels of education and has a negative coefficient implying that lower levels of education are associated with poor nutritional status. Mussa (2014), who also uses "no education" as the base category, finds negative coefficients for all levels of education but only statistically significant for some levels of educations. Chirwa and Ngalawa (2006) find mixed results, a case also found in this study. Lower levels of education have negative coefficients while higher levels have positive coefficients. Chirwa and Ngalawa (2006) find that both the mothers' and father's matter.

---

[12] Since we do not have panel data, we cannot conclude this is recovery but simply a finding that is consistent with possible patterns of recovery.

Higher asset index scores are associated with better levels of child nutrition levels for two models only, HAZ and WAZ. Mussa (2014) uses five categories of the asset index of the households, namely poorest, poor, middle, richer and richest. Using the poorest as the base category, he finds that household wealth seems to matter more in improving height-for-age z-scores in rural areas than in urban areas. For example, a child born into the wealthiest quintile in rural areas has a height-for-age z-score that is 0.31 standard deviations better than that of a child from the poorest wealth quintile. We also used asset quintiles (poorest as the base category) and found significant results (HAZ and WAZ only) for all quintiles except the middle quintile. Of great similarity is the finding that a child from the richest quintile is better by 0.33 and 0.32 standard deviations for HAZ and WAZ, respectively. The square of the asset index displays convexity for HAZ and WAZ and concavity for WHZ. The turning points are asset index scores of 5.66 (richest quintile), 7.11 (richest quintile) and -0.45(middle quintile) for HAZ, WAZ and WHZ, respectively.

2.20 Asset index and pro-poor growth analysis

Over time, DHS surveys have been upgraded to include more information and also adjusted for technological and demographic changes that have taken place. Our comparisons are, therefore, only based on 9 asset variables which are common in all data sets. Table 2.14 shows that the average MCA asset index scores have increased over time from 0.67 in 1992 to 1.04 in 2010, respectively. The spread as measured by standard deviation has declined suggesting a change in the distribution of asset index scores over time.

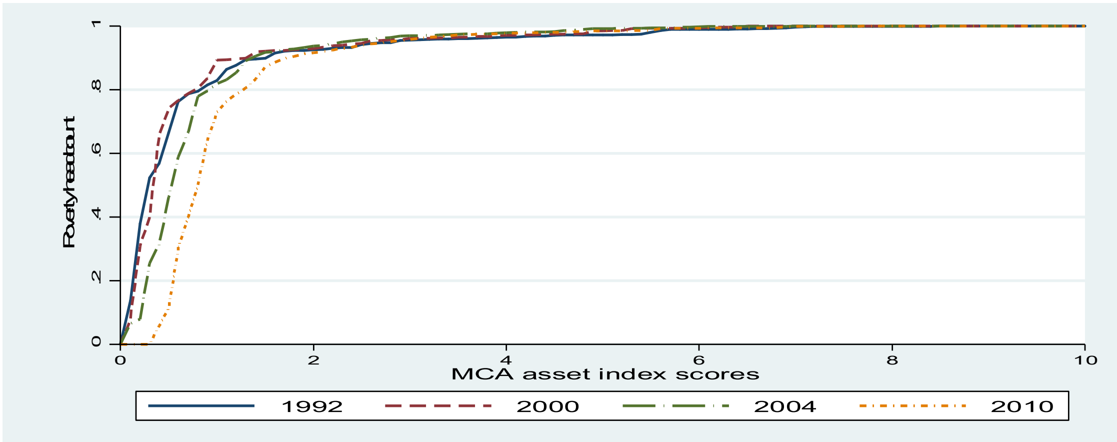
Table 2.14: Descriptive statistics for asset index scores

Survey year	Observations	Average	Standard deviation
1992	5,323	0.67	1.14
2000	14,213	0.62	0.99
2004	13,664	0.76	0.86
2010	24,825	1.04	0.92
Total	58,025	0.84	0.97

Source: Own computation from MDHS 1992, 2000, 2004 and 2010

Figure 2.7 shows the poverty incidence curves drawn using the asset index scores for 1992, 2000, 2004 and 2010. Higher curves indicate higher levels of poverty. The curves indicate that at lower levels of the MCA asset index scores, asset poverty is lowest in 2010, followed by 2004, 2000 and finally 1992. For higher levels of the poverty line, the curves cross making it difficult for us to conclude in which year poverty is higher.

Figure 2.7: Asset poverty incidence curves by survey year



Source: Own computation from MDHS data

We also calculate differences in the incidence of asset poverty since 1992 with the poverty line set at the 50.7th percentile (equal to an asset score of 0.807) as the cut-off point below which households are considered poor. The null difference is that there exists no difference in household asset ownership between any two chosen periods against the alternative hypothesis that there exist differences.

Table 2.15 shows that there has been an improvement in living standards in Malawi as shown by the reduction in poverty incidence from 79.5% in 1992 to 50.0% in 2010. Poverty incidence increased by 1 percentage point from 79.5% in 1992 to 80.5% in 2000 but the increase was not statistically significant. During this period, Malawi generally experienced a down turn in economic activity although we do not place a causal link here. This worsening of long term economic conditions was, however, offset by the improvements made over the three survey period from 2000 and 2010. The greatest improvement was registered between 2004 and 2010 during which was a period of economic prosperity for Malawi.in the country driven by strong growth in the agricultural sector[13].

[13] As noted earlier, the agricultural sector is one of the most important economic sectors in Malawi and it contributes at least 30% to national GDP.

Table 2.15: Differences in poverty headcount indices for household asset ownership

Description	Estimate	Std. Err.	T	P>|t|	[95% Conf. interval]		Pov. Line
1992	0.795	0.006	143.76	0.000	0.784	0.806	0.807
2000	0.805	0.003	242.46	0.000	0.799	0.812	0.807
Difference	0.010	0.006	1.56	0.118	-0.003	0.023	---
2000	0.805	0.003	242.46	0.000	0.799	0.812	0.807
2004	0.785	0.004	223.15	0.000	0.778	0.792	0.807
Difference	-0.021	0.005	-4.26	0.000	-0.030	-0.011	---
2004	0.785	0.004	223.15	0.000	0.778	0.792	0.807
2010	0.500	0.007	70.05	0.000	0.486	0.514	0.807
Difference	-0.284	0.008	-35.73	0.000	-0.300	-0.269	---
1992	0.795	0.006	143.76	0.000	0.784	0.806	0.807
2010	0.500	0.007	70.05	0.000	0.486	0.514	0.807
Difference	-0.295	0.009	-32.65	0.000	-0.313	-0.277	---

Source: Own computation from MDHS 1992, 2000, 2004 and 2010

Table 2.16 provides the growth rate in the asset index and estimates from the five different pro-poor indices discussed in Section 2.6. The results indicate that there has been both absolute and relative pro-poor growth in asset ownership in Malawi over the entire period of study from 1992 and 2010.

Table 2.16: Indices of pro-poorness in child nutritional status between 1992 and 2010

Pro-poor indices	1992-2000	2000-2004	2004-2010	1992-2010
Growth rate(g)	-0.076	0.217	0.445	0.626
Ravallion and Chen (2003) index	0.017	0.654	0.545	1.217
Kakwani and Pernia (2000) index	0.565	0.788	1.098	2.240
PEGR index	-0.043	0.171	0.489	1.402
Ravallion and Chen (2003) index – g	0.093	0.437	0.100	0.591
PEGR index - g	0.033	-0.046	0.044	0.776

Source: Own computation from MDHS 1992, 2000, 2004 and 2010

Malawi experienced a negative growth rate of 0.08% between 1992 and 2000. The negative growth suggests that not only did poverty increase but the poor households were also negatively affected in relative terms. Between 2000 and 2004, we find evidence of absolute but not relative pro-poor growth. However, for the period between 2004 and 2010, we are able to conclude absolute but not relative pro-poor growth because the PEGR index – g is negative.

In Table 2.17 we show the mean access of assets by area of residence and region based on the pooled data set. The table helps us to determine the driving factors behind the observed rural-urban and regional differences in the levels of asset poverty in Malawi. Furthermore, this analysis is important because as pointed out in Section 2.7, adjusting the asset index shifts the distribution to the right which may be a limitation for inequality comparisons over time because the mean has changed (see, e.g., Wittenberg & Leibbrandt (2015). Therefore, we use this as robustness checks for our pro-poor growth analysis results.

With respect to private assets, large differences in ownership are noted for “Bicycle”, “Electricity”, “Car/truck” and “Motorcycle” between urban and rural areas. Most of the households from the Northern region own “Paraffin lamps”. Ownership of toilet facilities is low in Malawi; even in urban areas, mean access is at a dismal 14%. Traditional pit latrines remain the dominant form of toilet facilities. Decent floor material (tile and cement) and excellent water sources are more accessible in urban areas.

Table 2.17: Pooled mean access of assets by area and region, all periods

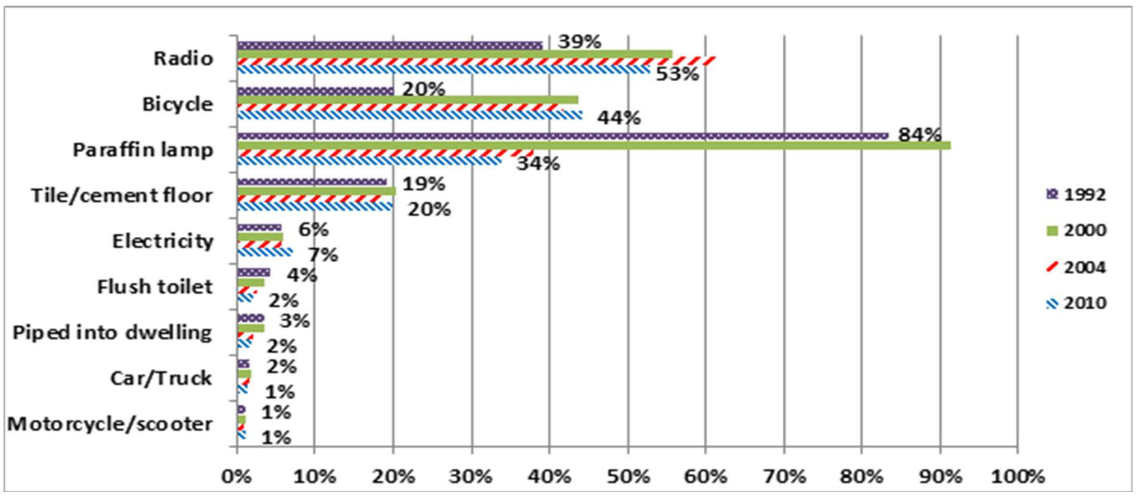
Description	Urban	Rural	Northern	Central	Southern	Total
Average household size	4.64	4.63	5.01	4.75	4.41	4.63
Private assets						
Radio	73.40%	51.00%	55.20%	52.50%	55.30%	54.30%
Bicycle	32.50%	42.90%	34.90%	42.00%	43.20%	41.40%
Paraffin lamp	61.70%	52.10%	65.60%	52.40%	50.20%	53.60%
Electricity	29.50%	2.30%	6.70%	5.00%	7.30%	6.40%
Car/truck	6.60%	0.70%	1.50%	1.60%	1.60%	1.60%
Motorcycle/scooter	2.10%	0.90%	0.90%	1.00%	1.10%	1.10%
Toilet facility						
Own flush toilet	14.00%	0.70%	2.40%	2.50%	2.90%	2.70%
Shared flush toilet	0.30%	0.10%	0.30%	0.00%	0.10%	0.10%
Traditional pit toilet	54.20%	44.30%	44.70%	45.20%	46.50%	45.70%
Ventilated improved pit latrine	7.00%	2.80%	2.60%	3.60%	3.60%	3.40%
No facility/bush	3.40%	17.40%	14.90%	17.20%	14.10%	15.30%
Other	21.20%	34.80%	35.20%	31.50%	32.90%	32.80%
Floor material						
Mud/earth	37.50%	85.40%	77.10%	79.80%	77.50%	78.20%
Cement	8.10%	0.70%	3.10%	1.50%	1.50%	1.80%
Bricks	0.10%	0.00%	0.20%	0.00%	0.00%	0.10%
Wood	0.30%	0.00%	0.00%	0.00%	0.10%	0.10%
Tiles	52.70%	11.60%	17.70%	14.80%	19.70%	17.70%
Other	1.20%	2.40%	1.90%	3.80%	1.10%	2.20%
Water source						
Piped into residence	13.40%	0.60%	2.70%	1.90%	2.90%	2.50%
Public tap	26.30%	2.50%	10.10%	4.50%	5.70%	6.00%
Piped into yard/plot	37.40%	9.90%	15.70%	8.70%	17.10%	14.00%
Well in residence	7.50%	32.60%	25.20%	29.80%	29.50%	28.90%
Public well	1.60%	3.80%	5.00%	4.00%	2.50%	3.40%
Protected well/borehole	2.60%	7.90%	4.50%	9.90%	6.10%	7.10%
Spring	1.90%	3.70%	2.70%	5.00%	2.60%	3.40%
River/stream	4.90%	18.60%	15.00%	19.70%	14.80%	16.50%
Pond/lake	3.00%	10.90%	7.00%	9.30%	10.90%	9.70%
Dam	0.20%	0.10%	0.30%	0.10%	0.10%	0.10%
Rainwater	0.10%	1.20%	1.10%	1.00%	1.10%	1.10%
Other	1.20%	8.20%	10.70%	6.20%	6.70%	7.20%

Source: Own computation from MDHS 1992, 2000, 2004 and 2010

We also show in Figure 2.8 access to assets by type and survey period. This analysis explains the sources of the movements in the asset index over time. The movements themselves could be due to so many factors including technological changes and changing tastes and preferences. Furthermore, showing access to assets is important because adjusting the asset index shifts the distribution to the right and this may be a limitation for poverty comparisons over time because the mean has changed. We use this as robustness checks for our pro-poor growth analysis results.

Between 1992 and 2010, ownership of radio and bicycle has expanded by 14 and 24 percentage points, respectively. On the other hand, ownership of paraffin lamp has dropped by 50 percentage points from 84% in 1992 to 34% in 2010. Access to tile/cement floor and electricity has marginally increased by around 1 percentage point over the period from 1992 and 2010. Ownership of car/truck and motorcycle/scooter has stagnated at 2% and 1%, respectively. Ownership of flush toilet and access to piped water into the dwelling has dropped by 2 percentage points.

Figure 2.8: Access to assets by type and survey period



Source: Own computation from MDHS 1992, 2000, 2004 and 2010; data labels for 1992 and 2010

2.21 Pro-poor growth in child nutritional status

Our analysis is based on HAZ owing to the fact that it is long term as already discussed. The HAZ scores are calculated using the new WHO (2006) child growth standards and, therefore, directly comparable over time. Table 2.18 provides summary statistics on HAZ in terms of the mean and standard deviation.

Table 2.18: Descriptive statistics for HAZ

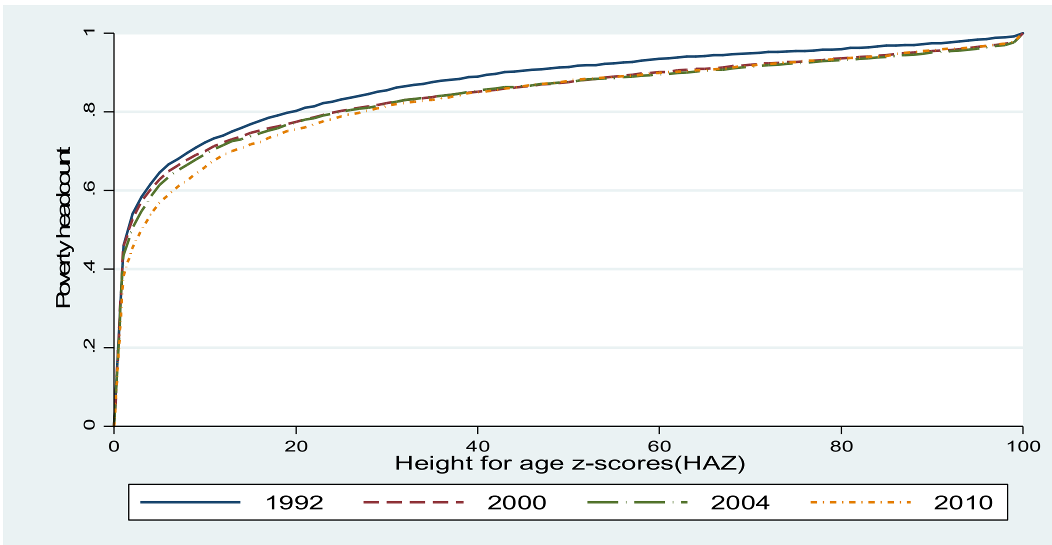
Survey year	Observations	Average	Standard deviation
1992	3,288	-2.01	1.57
2000	9,396	-1.93	1.79
2004	8,309	-1.94	1.80
2010	4,653	-1.76	1.66
Total	25,646	-1.91	1.75

Source: Own computation from MDHS 1992, 2000, 2004 and 2010

Just as with the asset index, there has been an improvement in the average HAZ scores from -2.01 in 1992 to -1.76 in 2010. The deviation in the HAZ scores has slightly increased.

The poverty incidence curves given in Figure 2.9 indicate that incidence of poverty headcount is unambiguously highest in 1992. For the rest of the years, the curves are compact except for specific lower ranges of the z-scores.

Figure 2.9: Poverty incidence curves for HAZ by DHS survey year



Source: Own computation from MDHS 1992, 2000, 2004 and 2010

Results for the calculated differences in the FGT headcount index over time and for successive survey periods are given in Table 2.19. As before, the poverty line is set at 2.3rd percentile which corresponds to the international cut-off of -2SD proposed in the WHO (2006) standards below which children are considered malnourished. The null hypothesis is that there exists no statistical difference in the HAZ scores between any two chosen periods against an alternative that there exist differences.

The results show that there has been an improvement in the nutritional status of children aged below 5 years in Malawi during the period between 1992 and 2010; the incidence of stunting levels amongst under-five aged children in Malawi has decreased from about 54% in 1992 and to about 46% in 2010.We also find significant improvements in the levels of child nutritional status for the successive survey periods apart from the 1992 and 2000 pair where the reduction in the incidence of malnutrition is not statistically different from zero.

Using the asset index as a measure of welfare, results indicated that poverty increased between 1992 and 2000 but not statistically significant. Therefore, it is concluded that the results are similar and that both asset poverty and child nutritional status remained the same over the period.

Table 2.19: Differences in the FGT poverty headcount index for HAZ

Description	Estimate	Std. Err.	T	P>|t|	[95% Conf.	interval]	Pov. line
1992	0.539	0.011	49.174	0.000	0.517	0.561	2.300
2000	0.519	0.006	82.173	0.000	0.506	0.531	2.300
Difference	-0.020	0.013	-1.601	0.111	-0.045	0.005	---
2000	0.519	0.006	82.173	0.000	0.506	0.531	2.300
2004	0.496	0.007	70.046	0.000	0.482	0.510	2.300
Difference	-0.023	0.009	-2.430	0.016	-0.042	-0.004	---
2004	0.496	0.007	70.046	0.000	0.482	0.510	2.300
2010	0.457	0.000	.	.	0.457	0.457	2.300
Difference	-0.039	0.007	-5.484	0.000	-0.053	-0.025	---
1992	0.539	0.011	49.174	0.000	0.517	0.561	2.300
2010	0.457	0.000	.	.	0.457	0.457	2.300
Difference	-0.082	0.011	-7.491	0.000	-0.104	-0.061	---

Source: Own computation from MDHS 1992, 2000, 2004 and 2010

We also find evidence of both absolute and relative pro-poor growth in child nutritional status for 2000-2004, 2004-2010 and 1992-2010 as shown in Table 2.20. Not only did absolute poverty decline but the poor also benefited more. For 1992 and 2000, the Ravallion and Chen (2003) index gives a negative estimate of 0.203 thereby making us unable to conclude absolute pro-poorness despite the fact the Kakwani and Pernia (2000) index and PEGR index indicate otherwise. We are also unable to conclude relative pro-poorness using all the three relative measures, namely the Ravallion and Chen (2003) index – g, PEGR index- g and the Kakwani and Pernia (2000)'s index.

Table 2.20: Indices of pro-poorness for HAZ

Pro-poor indices	1992-2000	2000-2004	2004-2010	1992-2010
Growth rate(g)	0.223	0.028	0.044	0.313
Ravallion and Chen (2003)	-0.203	0.360	0.946	1.042
Kakwani and Pernia (2000)	0.624	8.547	7.702	2.965
PEGR index	0.139	0.241	7.702	0.928
Ravallion and Chen (2003) – g	-0.426	0.332	0.902	0.729
PEGR - g	-0.084	0.213	0.295	0.615

Source: Own computation from MDHS 1992, 2000, 2004 and 2010

2.22 Conclusions

In this study, we have measured poverty and inequality in Malawi both spatially and temporally. Over the period between 1992 and 2010, we find evidence of pro-poor growth in both absolute and relative terms. We have also identified factors associated with asset poverty in Malawi. Results from the study indicate that while household size is positively correlated with asset accumulation, age dependency ratio has a negative association. There is no statistically significance difference in asset ownership between male and female headed households. Households with older household heads do better in terms of asset ownership than those headed by younger people. Shock to sickness in the household has a negative association with asset ownership. Education attainment by the members of the household has been found to have a positive association with the level of assets in the household. Rural areas have higher levels of asset poverty compared to urban areas. Significant regional differences between the northern and central regions have also been found in terms of asset poverty with the latter lagging behind.

We have also identified factors that are associated with the nutritional status of under-five children in Malawi. Using three anthropometric measures of nutrition, namely HAZ for stunting, WAZ for underweight and WHZ for wasting, it is shown that the incidence of stunting is the highest of the nutritional problems amongst under 5 children in Malawi followed by underweight. On the one hand, there no large differences between regions and areas in terms of child nutritional status. On the other hand, when assets are used, the welfare gap between regions and areas is bigger.

2.23 Policy discussion

Non-monetary pro-poor growth analysis is important for policy because it allows us to identify the progress made in poverty reduction over time as well as analyse the redistribution effects that arise with growth in living standards. Specifically, as shown in the paper, the poor people in Malawi have benefited the most from the changes in assets and child-nutritional status that have occurred over time. This means that policy makers should not only focus on income growth when analysing progress towards goals set in the national development agenda such as those in the Malawi Growth and Development Strategy (MGDS) and the Sustainable Development Goals (SDGs).

Subgroup poverty comparisons are important for targeting development policies towards specific groups that have been identified as the poorest. For example, our study has shown that poverty is higher in rural areas than urban areas. As a response to these findings, deliberate policy measures might be put in place to specifically target the identified poorest groups of Malawi. Similarly, policies can be developed to address the problem of inequality as informed by the results.

A number of policy conclusions can be also be drawn from the regression analysis. Firstly, the negative association between asset ownership and age dependency ratio suggests that increasing the income generating opportunities for the economically active population could help in reducing asset poverty in Malawi. Younger household members could, for example, be equipped with the necessary skills as they wait to enter the working age group. Secondly, the significance of education attainment and the fact that the sizes of coefficients increase with the level of education imply that post-primary education should be emphasised in Malawi. Currently, the focus of education policy is on basic education (completion of primary education). Thirdly, since shocks to illnesses are negatively associated with asset ownership, it may be important to consider looking into policies that improve access to health and assist households to withstand shocks thereby preventing them from being pushed into poverty traps.

The multivariate analysis of child malnutrition reveals that characteristics such as age, sex and twin status matter in the outcomes of nutritional status at the child level. Child birth order, mother education and economic status have a positive association with child nutritional status. These findings have policy implications too. Firstly, based on the literature linking recovery from stunting to cognitive outcomes, the study suggests that timing of nutritional inputs is critical in the cognitive development of children. Secondly, since male children tend to have weaker nutritional status compared to their female counterparts, nutritional feeding programmes might consider addressing this gender imbalance. Thirdly, the importance of female education for the nutritional status of children suggests that training women in child nutrition may require emphasis. Fourthly, related to the second implication, the significance of higher levels of education suggests that post-primary education should be emphasised in general. Currently, the focus of education policy is to achieve basic or primary education. Finally, since higher levels of economic status are associated with better nutritional status, policy should continue to be geared towards improving households' living conditions.

Chapter 3

Externalities and returns to education in Malawi: Panel data evidence

3.1 Introduction

Malawi remains a poor country despite registering gains in poverty reduction over the past two decades. Estimates based on Malawi's first, second and third integrated surveys (IHS1, IHS2 and IHS3) show that the incidence of poverty based on household percapita consumption has marginally fallen from 52.4% in 2005 to 50.7% in 2011. In 1998, the poverty rate was 65.3%. Rural areas, which make up about 85% of the population, have significantly higher poverty rates than urban areas, although the gap between the two is closing over time. The Gini coefficient shows that inequality increased from 0.401 in 1998 to 0.390 in 2005 to 0.452 in 2011 (National Statistical Office, 2012).

On the one hand, the labour participation rate for Malawi, defined as the share of the population aged 15 and above working or seeking employment, is very high and stands at 88%. On the other hand, education levels are low. About 74% of the population aged 15 years and above do not have any qualification at all and 21% reported never having attended education. Therefore, literacy remains a challenge in Malawi given the high rate of labour force participation. As at 2011, the literacy rate (defined as the ability to read and write with understanding in any language) amongst people aged 15 years and above stood at 65%, which is an insignificant improvement from 64% reported in 2005 (National Statistical Office, 2012).

In the Malawi Growth and Development Strategy (MGDS), education is identified as one of the themes necessary for growth and socio-economic development. Malawi's formal education system consists of primary, secondary and tertiary or post-secondary education. The country's education policy has been largely focussed towards increasing access to primary education. In recognition of this, primary school education was universally made free in 1994 for all government schools. According to National Statistical Office (2012), as a response to these policy changes, the net primary enrolment rate has increased to 86% in 2011 from 80% in 2004 while the primary dropout rate has dropped from 5% to 1% over the same period. In addition to universal free primary education, tertiary education has been subsidised in order to make it more affordable, although the country is yet to realise the benefits in terms of expansion in tertiary education.

The Malawian labour market can be categorised into the formal and informal sectors. As is the case in many other developing countries, the formal sector in Malawi only consists of a small percentage of the labour force. Consequently, most people are involved in either self-employment activities or paid employment in the informal sector (Chirwa & Matita, 2009). The informal sector is, therefore, important to Malawi's economy; and it accounts for 78% of total employment (National Statistical Office, 2014).

Education has been identified as a tool for poverty reduction in Malawi (National Statistical Office, 2012). Through education attainment, the poor are said to be empowered and equipped for better opportunities in national development.

It is against this background that the study seeks to look at the role of education in poverty reduction in Malawi. The link between poverty and education is identified through the labour market. A number of studies that link education to poverty reduction have been conducted. Da Maia (2012) looks at the link between education and poverty reduction in Mozambique. The study estimates the probability of an individual getting employment in any of the given sectors conditioned on education and models the relationship between education and earnings. Several researchers have looked at the role played by education in earnings (e.g., Mincer, 1974; Psacharopoulos, 1994 & 2002; and in Malawi, Chirwa & Matita, 2009; Chirwa & Zgovu, 2002). Clearly, the relationship between education and earnings is not only of great interest to many scholars but also important for a policy perspective.

However, previous studies on Malawian labour markets (see Chirwa & Matita, 2009; Matita & Chirwa, 2009; Chirwa & Zgovu, 2002) have explored the link between education and earnings using cross-sectional data sets due to lack of panel data. This study expands on the available literature by taking advantage of the newly released panel data set. Use of panel data has many advantages and this is well acknowledged in the literature. Firstly, we can control for unobservable individual heterogeneity. As shown in the literature, failure to control for individual specific effects leads to bias in results. Secondly, panel data contains rich information about cross-sectional variations and dynamic behaviour of the subjects of interest. Thirdly, with panel data, we are able to identify time effects which cannot be estimated with cross-sectional data. Baltagi (2013) provides a detailed discussion on the advantages and limitations of panel data.

In addition to using panel data, the study distinguishes between the formal and informal sectors of the economy to see if the role played by education in earnings differs by sector. This decomposition is important considering the large size of the informal sector. Furthermore, the study corrects for sample selection bias which arises as a result of selection into economic sectors as well as working with only selected individuals who end up entering employment instead of the full working age-population. With regard to distinguishing between sectors, we improve on previous studies (e.g. Chirwa & Matita, 2009) by conducting coefficient comparison tests between groups by gender and sector. The test is routinely implementable in Stata 13.1 via the *suest* command (see Appendix A1). The need for conducting coefficient comparison and balancing tests has been motivated in Pischke and Schwandt (2014).

The study also explores whether earnings from non-farm household enterprises are affected by the educational attainment of other members of the household. Basu and Foster (1998) argue that the presence of a literate person in a given household may generate positive intra-household education externalities. Unlike the isolated illiterate persons, illiterate individuals living in a household where at least one member is literate enjoy some of the literate person's capabilities. Therefore, education is said to have some positive spill-over effects in that the literate members of the household may help in the running of enterprises in the same manner agricultural extension workers or medical personnel help in their communities. We explore whether there are positive education externalities in household enterprises by the use of maximum number of years of schooling in a household. This specification is also motivated by Jolliffe (2002) who finds that in household level income functions such as non-farm enterprises, the maximum or average level of education in the household is a better explanatory variable of household income. Furthermore, it is argued that in most developing countries income is largely earned at the household rather than the level of the individual, hence our focus on household enterprises. Moreover, according to data from IHS3, the enterprise sector is large in Malawi and affects a significant percentage of households in Malawi. Specifically, about 20% of households are involved in non-farm employment activities in Malawi (National Statistical Office, 2012).

The chapter addresses three main specific issues (i) to investigate the factors influencing labour force participation and employment likelihood in Malawi; (ii) to estimate returns and externalities to education in Malawi; and (iii) to identify and deal with inconsistencies in the IHPS data. This includes conducting comparisons of earnings and consumption data in addition to robust treatment of outliers.

3.2 Methodology

3.2.1 Theoretical framework

There are two main competing groups of theories for explaining labour market outcomes. First is the traditional neoclassical model of labour supply which argues for a simple competitive labour market whereby workers are paid their marginal product. In this competitive model of the labour market, individuals' wages are largely determined by their productivity (Kerr & Teal, 2015). According to this model, wage differences across sectors or occupations are competitively eliminated by workers moving away from underpaid sectors to highly paid sectors. One of the early works explaining occupational choice is by Roy (1951). Using a two-sector model, Roy (1951) argued that individuals choose sectors in which they have a comparative advantage such that wage differences between sectors are simply a reflection of the differences in unobserved ability between employees in the two sectors. However, investments in education and training were not modelled in Roy's model.

The simplest form of the competitive model does not distinguish between degrees of job informality or differences in individual investments that would enhance productivity. However, by allowing for investment in human capital, individuals are assumed to make investments in education and other productivity enhancing abilities in such a way as to maximise their expected lifetime utility. Despite contributions by other researchers, much of the human capital theory is based on the work of Becker (1962) and Mincer (1974) after which the human capital theory became the standard for empirical work in labour economics (Card, 1999). In this strand of literature, earnings are assumed to be dependent on education or training and experience in the labour market

The alternative is the segmented labour market hypothesis which observes labour market outcomes and wage differentials that are inconsistent with the competitive model. Specifically, the latter argues that people's choices of where to work and the wage received are determined by institutional factors such as government legislation (e.g. minimum wage legislation) and other institutional factors (see (Leontaridi, 1998; Dunlop, 1957; Kerr, 1954). It is, for example, argued that labour market institutional constraints prevent the bargaining down of wages in high wage sectors by workers in low-wage sectors. Furthermore, it was observed that workers with similar jobs were paid different wages as a result of such institutional constraints. A number of explanations have been given for the existence of institutional constraints and wage setting mechanisms between sectors. One of these explanations in the developing country context is the Harris-Todaro model of migration (Harris & Todaro, 1970). According to this model, while the urban sector is insulated from the forces of supply and demand as a result of minimum wage legislation, the rural sector was assumed competitive. The existence of a minimum wage in the urban sector meant that rural workers or the unemployed were unable to bid the urban wage down to an equilibrating level. Under these conditions, this ensured that workers who find employment in the urban sector would therefore be paid more than if they were employed in the rural agricultural sector.

Despite these criticisms, the human capital theory remains the leading theoretical model for explaining labour market outcomes and is used in this study. In addition to these two competing theoretical frameworks, recently, there have also been efforts towards improved understanding and definitions of informal markets as well as the linkages between the informal and formal economy. This has given rise to different theories or explanations as to what constitutes or gives rise to informal labour markets (Chen, 2012). This debate is relevant in the Malawi context because of the huge size of the informal employment, which makes up of 78% of total employment, as earlier stated.

The recognition of the existence of an informal market in the analysis of labour markets in developing countries largely came as a result of the 1972 International Labour Organization (ILO) study in Kenya (Singer & Jolly, 1972) where it was found that the traditional and informal sectors of the economy included profitable and efficient enterprises as well as marginal activities. In addition to this, work by British anthropologist Hart (1973) found that there was no wage employment among unskilled migrants who had migrated from Northern Ghana into the capital city. Instead, these were involved in low-income activities in the informal sector not related to the formal sector. Fields (1975) made another contribution to the empirical and theoretical research on the informal sector by extending the Harris-Todaro framework to include an urban informal sector. In this framework, three sectors are recognised, namely the urban formal sector, the urban informal sector and the traditional informal sector. The urban informal sector was made up of casual employment and was viewed as a temporary sector in which people freely enter as they look for higher paid jobs in the formal sector. However, the presence of union activity and minimum wage legislation prevented people from entering the formal sector and therefore kept workers in the informal sector (Kerr & Teal, 2005).

The definition of the informal sector has changed over the years. Previously, following the ILO Kenya report, the characteristics of informal activities included ease of entry, family ownership of enterprises and unregulated and competitive markets (Singer & Jolly, 1972). ILO now uses the expanded statistical criteria for distinguishing between formal and informal employment which was endorsed by the 2000 International Labour Conference and the 2003 International Conference on Labour Statisticians (Hussmanns, 2004). The formal sector primarily includes salaried employment in the private and government sectors as well as non-governmental organisations (NGOs). In this sector, the relationship between the employer and employees is governed by formal labour laws including employee benefits and income taxation. In contrast, the informal sector has two components, namely self-employment activities and informal wage employment.

Self-employment mainly consists of employers in informal enterprises, own account workers in informal enterprises, contributing family workers (in informal and formal enterprises) and members of informal producers' cooperatives. Informal self-employment includes enterprises that are not registered under any national legislative authority and not engaged in agricultural activities.

Informal wage employment includes employees of informal enterprises, casual or day labourers, temporary or part-time workers, paid domestic workers, contract workers, unregistered or undeclared workers and industrial outworkers (also called homeworkers). Of great importance to the Malawian economy is casual employment (locally known as ganyu[14]). Despite being largely seasonal, it is very important in both urban and rural areas.

[14] Ganyu is the dominant form of employment in the informal sector.

3.2.2 Estimating returns to education

Mincerian earning functions are the standard approach for estimating returns to education in labour markets. The methodology started with the work of Mincer (1974) and has been widely used in the literature (Psacharopoulos, 1994). It is based on the human capital theory which argues that investment in education improves workers' skills resulting in high productivity and, therefore, higher earnings (Mincer, 1974). The basic model is summarised below:-

$$\ln(Y_{it}) = \beta_1 + \beta_2 S_{it} + \beta_3 E_{it} + \beta_4 E_{it}^2 + \varepsilon_{it} \qquad (3.1)$$

where, for any individual (i), at time (t), Y is the earnings of individual, S is the number of years of schooling, E is the experience, E^2 is experience squared and ε is the error term. The coefficient β_2 is interpreted as the private rate of return to education (RORE) and $(\beta_2 * 100)$ gives the percentage return to one additional year of schooling. The above specification can be extended to include other control variables such as gender, location, etc.

As is the practice in the literature, we improve on the classic model presented in equation (3.1) in two main ways. Firstly, we account for the fact that returns to education may be heterogeneous rather than homogeneous and secondly, we address the problem of selection bias which arises as a result of using non-random data for analysis. Due to its importance, we will deal with sample selection in the section that follows.

As stated, the basic model disregards the differences in the level of educational achieved by looking at a single overall education level or years of schooling. This approach is called the one factor or homogeneous model since it assumes that there are no differential trends in the returns to education for different levels of education. There is little statistical evidence and causal empiricism for the homogenous model. The alternative approach, called the multiple factor approach or heterogeneous model, looks at the different levels of education as having separate effects on earnings. This involves replacing S with an educational dummy variable to represent the different educational categories. Seven educational qualifications are captured in the data, namely "None", "Primary School Leaving Certificate of Education (PSLCE)", "Junior Certificate of Education (JCE)", "Malawi School Certificate of Education (MSCE)", "Non-University Diploma", "University Diploma" and "Post-Graduate degree". In some cases of our analysis, we combine the last three categories into tertiary education[15].

[15] In Malawi, PSLCE, JCE and MSCE are equivalent to 8, 10 and 12 years of schooling. For tertiary education, diplomas are usually completed within two or three years while undergraduate degrees take four years except for

Our data set does not have information on the number of years of experience. Following the standard practice used in the literature (e.g., Matita & Chirwa, 2009; Kahyarara & Teal, 2008; Appleton, Bigsten, & Manda, 1999), we estimate the number of potential years of experience as age less years of schooling less preschool age. This assumes that once people complete their education, they immediately enter the labour market. Those without education are assumed to enter the labour market at the lowest labour market entry age of 15 years.

3.2.3 Sample selection

Sample selection[16] bias arises from a number of sources. Firstly, it could arise as a result of self-selection into different employment categories or sectors. Secondly, it could be due to non-random attrition in panel surveys when subjects drop out of the sample for some reason. However, in our data, attrition is not a major problem, as we will discuss in Section 3.3 on data. Consequently, we do not deal with sample selection caused by attrition. Thirdly, it may be as a result of working with a truncated sample. This raises the problem of sample selection because ideally we are interested in studying all individuals in the working age population (15-64) but end up with only those that actually entered employment. Fourth and finally, since earnings are only observed for individuals who are working at the time of the survey, our analysis is made on the sample selected on this basis.

Failure to correct for sample selection would bias our results. The Heckman (1979) two-step procedure, following Wooldridge (2002), has been used to correct for selection bias. In the first step, the probability of an individual selecting into an economic sector or attriting through a probit model is estimated as:

$$c_{it} = \beta x_{it} + v_{it} \tag{3.2}$$

where: i represents an individual, c denotes the binary response variable equal to either 1 or 0 and x is a set of regressors such as education level, age, sex, marital status, etc., including dummies for location. The error term is given by v.

From equation (3.2), we obtain the inverse of the Mills ratio and use it as an explanatory variable in the estimation of either the wage equation (3.1) or the likelihood of employment conditional on an

---

engineering, law and medicine which are completed in five years. A post-graduate degree usually takes two years to complete.

[16] Some may refer to this as self-selection. Whether we define it as sample selection or self- selection is trivial in our study as long as we achieve our goal of addressing the non-random nature of the sample when estimating our employment and earnings functions.

individual's participation in the labour market (see Section 3.6.2). If the coefficient of the inverse Mills ratio is statistically significant, then we are justified in correcting for selection bias. We then proceed to correct for and report heteroskedasticity consistent standard errors in the first stage.

3.2.4 Modelling unobserved heterogeneity

Let us consider the following model when time $T = 1,2$:

$$y_{it} = \beta x_{it} + u_{it} \tag{3.3}$$

Where y_{it} is the log of earnings for individual i at time t. Suppose that the error term is made up of two components as follows:

$$u_{it} = n_i + v_{it} \tag{3.4}$$

where n_i is time invariant and correlated with x_{it}, and v_{it} is time varying and uncorrelated with x_{it}. If the exogeneity assumption is violated, i.e. when $E(x_{it} n_i) \neq 0$, the OLS estimator will be biased in cross-section. On the other hand, panel estimators can be used to control for unobserved individual time-invariant heterogeneity and this allows us to obtain unbiased estimates of β. Even when the unobserved correlated effect is not time invariant, using panel data techniques reduces the magnitude of the bias.

The most commonly estimated models with panel data are the fixed effects and random effects models and several considerations will affect the choice between the two. The first consideration is the nature of omitted variables. If we think that there are no omitted variables from the model or that the omitted variables are uncorrelated with the explanatory variables in the model, then a random effects model is the best (Williams, 2015). A random effects model under these assumptions will produce unbiased estimates of the coefficients, use all the data available, and yield the smallest standard errors. However, it is more likely that omitted variables will produce at least some bias in the estimates. If there are omitted variables and these variables are correlated with the variables in the model, then a fixed effects model provides a means for controlling for omitted variable bias. In a fixed-effects model, subjects serve as their own controls. For this to work, the omitted variables must have time-invariant values with time-invariant effects. For example, gender does not change over time and its effect on the outcome in wave 1 is the same as the effect of gender in wave 2.

Secondly, researchers consider the amount of variability within subjects: If subjects do not change much, or not at all, across time, then fixed effects models may not work very well or even at all. There is need to have within-subject variability in the variables to justify the use of fixed effects. When there is little variability within subjects, the standard errors from a fixed effects model may be too large to tolerate. Conversely, random effects models will often have smaller standard errors (Clark & Linzer, 2015).

Third are the effects are we interested in studying. In fixed effects models, we are not interested in estimating the effects of variables that do not change or change very little over time. Rather, we control for them or "partial them out". On the other hand, with random effects models, we are able to estimate the effects of time-invariant variables such as gender, although the method is no longer controlling for omitted variables (Williams, 2015).

Given the above considerations, we choose to use the random effects model for three main reasons. First, education, whose effect we are interested in measuring, is generally a slow changing variable especially for a three year period over which we have data. Secondly, given that our panel is short (only two periods) there is not much within-subject variability in most of our variables. Third, a random effects model allows us to estimate the effects of time-invariant variables such as gender which is an important aspect in Malawi. With fixed effects, this is not possible since the variable gets dropped off after demeaning.

Therefore, using a random effects model naturally comes at a cost and the trade-off is that their coefficients are more likely to be biased than the fixed effects estimates. Nevertheless, according to Wooldridge (2002), panel data techniques reduce the magnitude of bias compared to OLS.

3.3 Description of the data

This section provides a brief description of the Malawi Integrated Household Panel Survey (IHPS) data, a two-wave panel conducted in 2010 and 2013. The survey was implemented by the National Statistical Office (NSO) of Malawi. The 2010 wave was part of the third nationally representative Integrated Household Survey (IHS3)[17] conducted between March 2010 and March 2011 during which 3247 households were selected as a panel subcomponent to be resurveyed in 2013. The second wave, carried out between April and December 2013, saw the panel sample increase to 4,000 households because split-off members who formed new households were also included.

[17] IHS3 consists of four questionnaire types, namely the household, agricultural, fisheries and community questionnaires. In the household questionnaire, individuals are asked if they were involved in agricultural and fishing activities over the past 12 months. Those who answered "Yes" are then further administered the agriculture and fisheries questionnaire where more details on crop, livestock and fishing practices are collected. Our analysis is based on the household questionnaire only where the scope of our study lies.

The household formed the primary unit of analysis in the IHPS surveys. An attempt was, therefore, made to track all baseline households as well as members that moved away from the baseline dwellings between 2010 and 2013. Servants and guests at the time of the IHS3 were excluded and only individuals who were expected to be at least 12 years of age and known to be residents in mainland Malawi[18] were tracked. Of the 3,247 households initially chosen in 2010, some could not be located while others split into new households. Our analysis is based on the 3,104 households that are available in both waves. The rate of attrition at the household level was only 3.78 %.

The 2010 baseline had 15,597 individuals of all ages, of which 14,232 are available in both waves, representing an overall attrition rate of 7.42 % at the individual level. Given these low rates of attrition, which also seem random, we pursue this issue no further because we believe the representativeness of the sample has not been affected[19]. For purposes of this study, our sample is restricted to economically active available in both waves. All calculations include survey weights.

There are two sources of earnings as captured in the household questionnaire, namely wage employment and self-employment activities. Earnings from self-employment activities are provided as profit from non-farm enterprises over a period of 30 days while earnings from wage employment are given with an indication of the period over which earnings are earned, i.e. day, week, two weeks or month. Ganyu wages are given as daily earnings with an indication of the number of days worked in a week. All earnings are converted into real monthly figures in 2013 constant prices.

3.3.1 Work and non-work activities of the employed

The surveys collected information on both work and non-work activities of individuals over the seven days prior to the administration of the questionnaire. Non-work activities or domestic tasks are given in (a) and (b) while the rest constitute work activities.

a. How many hours did you spend yesterday collecting water?
b. How many hours did you spend yesterday collecting firewood (or other fuel materials)?
c. How many hours in the last seven days did you spend on household agricultural activities (including livestock and fishing-related activities) whether for sale or for household food?
d. How many hours in the last seven days did you run or do any kind of non-agricultural or non-fishing household business, big or small, for yourself?

[18] Excluding Likoma district, which is an Island on Lake Malawi.

[19] In the literature, the most common ways of addressing attrition are Inverse Probability Weighting (IPW) and Heckman selection correction (e.g., Wooldridge, 2002).

e. How many hours in the last seven days did you help in any of the household's non-agricultural or non-fishing household businesses, if any?
f. How many hours in the last seven days did you engage in casual, part-time or ganyu labour?
g. How many hours in the last seven days did you do any work for a wage, salary, commission, or any payment in kind, excluding ganyu?
h. How many hours in the last seven days did you engage in an unpaid apprenticeship?

All individuals and households involved in household agricultural activities identified from (c) are administered the household questionnaire where more information related to agriculture is collected. Questions (d) and (e) are further explored in the enterprise module where information on enterprise owners, customers and profits is collected. Questions (f) and (g) are further explored in the module for time use and labour where information on wages and occupations is collected.

3.3.2 Describing employment structure and hours worked

There are about nine mutually exclusive employment types that can be identified in the data set as shown in Table 3.1. Public works programme (PWP) and church/religious organisations can be jointly referred to as non-governmental organisations (NGO), considering the nature of their activities.

Table 3.1: Employment and occupation structures

	2010			2013		
Occupations	Frequency	Percent	Cumulative	Frequency	Percent	Cumulative
Private Company	270	8.9%	8.9%	303	8.0%	8.0%
Private Individual	261	8.6%	17.4%	225	6.0%	14.0%
Government	170	5.6%	23.0%	173	4.6%	18.6%
Parastatal	14	0.5%	23.5%	15	0.4%	19.0%
PWP	12	0.4%	23.9%	16	0.4%	19.4%
NGO	45	1.5%	25.3%	63	1.7%	21.1%
Other	6	0.2%	25.5%	18	0.5%	21.6%
Casual/Ganyu	1,620	53.2%	78.7%	2,056	54.5%	76.0%
Enterprises	650	21.3%	100.0%	904	24.0%	100.0%
Total	3,048	100.0%		3,773	100.0%	

Source: Own computation from IPHS data

On the one hand, the structure of employment observed in the table reflects a major similarity that Malawi shares with other less developed African countries, i.e. that casual employment and self-employment activities (household enterprises) make up the largest share of total employment both in and outside of agriculture. This observation holds in both 2010 and 2013. Moreover, attrition levels are very low such that the structure observed in the panel is largely a reflection of the national figures. On the other hand, the observed structure is a contrast to what is seen in countries like South Africa where there is a very small informal sector (e.g., Kingdon & Knight, 2004; Kerr & Teal, 2015).

The dominance of the informal sector is reflected in the number of observations and accompanying percentages in each of the employment types. Regular formal employment (private and government) have registered slight declines and possibly these partly account for the growth in the informal employment numbers. Total informal employment has increased as a result of growth in both casual employment and self-employment activities. The IHPS report states that the percentage of households in Malawi that operate non-agricultural enterprises has grown by 9 percentage points from 21% in 2010 to 30% in 2013.

The survey collected information on the number of hours worked[20] during the 7 days prior to the survey. We provide a breakdown of the average number of hours by occupation in Table 3.2.

The table reveals that, on average, individuals in the panel worked for about 25 hours in a week in 2013. This is an insignificant increase from the 24 hours registered in 2010. The maximum working hours regulated by the government is 48 per week in all jobs. Most of the occupations registered increases in the number of hours worked. PWP and ganyu employment registered significant declines. The former is a safety net scheme targeting the poor who mainly utilise it when there is a need. The latter employees (ganyu) usually work for food and basic needs and their labour supply is largely dependent on meeting daily needs. Once their daily needs are met, there is little motivation to work longer. Overall, the private sector (companies and individuals) worked the longest number of hours in both 2010 and 2013. The government and parastatal worked for a similar number of hours.

Table 3.2 also shows the average years of schooling by occupation. On average, number of years of schooling is about 7.5 and 7.4 years for 2010 and 2013, respectively. This is just below the official 8 years of primary education. What we find here is reflective of what the focus of government education policy has been in Malawi over the past two decades- the provision of free primary education. Government employees have the highest number of years of schooling across the two years while ganyu and public programme workers have the least. The levels of education partly explain the large earning differentials observed earlier, consistent with the human capital theory which states that education is positively correlated with earnings.

[20] Includes agricultural work but excludes unpaid domestic activities: water and firewood collection.

Table 3.2: Changes in the average weekly hours worked and years of education by year

Occupations	2010 Average weekly hours worked	2010 Years of education	2013 Average weekly hours worked	2013 Years of education
Private Company	40.20	8.80	44.10	9.00
	(25.50)	(4.50)	(23.60)	(04.50)
Private Individual	32.80	7.50	41.30	7.70
	(24.60)	(3.10)	(28.0)	(3.30)
Government	32.90	13.40	34.10	14.50
	(23.10)	(5.10)	(23.20)	(5.60)
Parastatal	36.30	11.20	35.20	10.60
	(21.0)	(4.80)	(29.30)	(5.20)
Public Works Programme	26.20	7.70	13.60	5.80
	(27.70)	(3.50)	(15.10)	(3.30)
Church/Religious Organisation	30.90	10.50	40.30	10.80
	(23.80)	(5.30)	(27.0)	(5.50)
Other	47.30	10.90	30.00	9.40
	(24.40)	(3.60)	(17.80)	(4.20)
Casual/Ganyu	16.40	6.50	15.10	6.40
	(17.60)	(2.60)	(18.10)	(2.70)
Non-agricultural enterprises	29.61	7.43	33.28	7.42
	(21.40)	(3.40)	(22.10)	(3.40)
Total	24.10	7.50	24.80	7.40
	(22.10)	(3.70)	(23.50)	(3.80)

Source: Own computations from IHPS data; standard deviations in parenthesis.

3.3.3 Treatment of outliers, missing data and zero earnings

Missing earnings, for example, because individuals refused to answer or the respondent did not know, are not imputed. Negative and zero earnings were dropped since their natural logs are undefined[21]. As is the practice in the literature, we only look at economically active individuals aged between 15 and 64 years. We also restrict our sample to observations with non-missing values in any of the variables reported in our tables. The resulting samples are 5,377 individuals and 1,702 households with positive real earnings. Next, we explain how outliers have been treated, using wage earners as an example.

Observations that are substantially different from the rest can make a difference to the regression results obtained. It is, therefore, important to not only investigate these unusual observations but also find ways of dealing with them. Wittenberg (2014) and Burger and Yu (2007) provide a good discussion on dealing with outliers. We discuss four approaches in the ensuing paragraphs.

[21] Researchers sometimes consider an alternative of adding a very small number to earnings before taking the log, which would allow considering zero incomes.

The first approach is to take out millionaires which are clear outliers. In our data set, there were 24 millionaires with average monthly earnings of MWK2,269 702 compared to MWK 38,471 for the rest of the 5353 individuals. However, the choice of millionaires is arbitrary and has a potential of removing genuine earners, especially considering that the minimum education level of these millionaires is MSCE- above the typical education level of those in the labour force.

The second approach is to remove outliers by identifying observations with extreme regression residuals. In linear regression, an outlier is an observation with a large residual. This is achieved by estimating a simple Mincerian type wage regression of the log of real monthly wages on education, age, age squared, gender and occupation. After running this regression, studentised or standardised residuals were created. In this approach, studentised residuals with absolute values greater than five are flagged as extreme and corresponding observations dropped. Using this approach, only five observations were flagged as extreme. We reduced the cut-off to 4 resulting in 14 individuals being flagged as having earnings that were too high or low for their characteristics.

The third approach is robust regression. When data is contaminated with outliers, using studentised residuals has been found to be insufficient in identifying the 'bad' observations. Robust regression is an alternative to least squares regression when this is the case, i.e. when data is contaminated with outliers or influential observations (Wittenberg, 2014). This approach is easily handled in Stata 13.1 and observations are given weights depending on whether they are outliers or not. Outliers are assigned zero weights and consequently identified as not belonging in the regression. In total, robust regression identified 35 observations as being extreme and this included all the outliers also identified through the studentised residuals.

The final approach is to remove observations in the 100th percentile. We generated a new variable containing percentiles of real monthly earnings. This was used to identify and then drop 56 observations in the 100th percentile with average real monthly earnings of MWK1,394,771 compared to MWK355,924 in the 99th percentile. However, this results in a loss of 56 observations, which we deemed excessive. One can equally consider dropping observations in the bottom percentile but we are more concerned with outliers in the higher percentiles. The median of earnings is low (about MWK8,400) while the mean is high (about MWK26,253), suggesting the data distribution is positively skewed by the presence of large outliers.

Considering sample size issues, we used the second approach with results from the other three approaches presented as robustness checks (see Section 3.8.1.3)

.

3.3.4 Dealing with inconsistencies

We noted some inconsistencies in terms of the period over which salaries were paid. About 6% of the total sample in the formal sector reported a payment period different from that reported in 2010. Similarly, about 24% of those in ganyu employment reported a different number of days spent on ganyu per week in 2013 than they previously reported in 2010. One may reason that some of these changes may be genuine considering that people switch jobs and occupations. However, before applying the changes, we first checked for unusually large swings in wages and earnings. Moreover, the inconsistencies were not unique to a specific occupation but rather spread across all occupations-confirming our argument. These inconsistencies were dealt with by creating a period that is consistent in both waves[22].

In order to reduce noise introduced by individual heterogeneity and arrive at a fairly consistent data set, we only worked with a balanced panel and also dealt with outliers which were partly responsible for unrealistically large increases in earnings. Some of the outliers could have been as a result of the manner in which we generate our monthly earnings, i.e. (ganyu days in a week*daily wage*4). In some cases, we found some few ganyu millionaires. Our analysis showed that of the 56 total outliers, 7 were ganyu earners. We had to deal with the outliers because we believe that a ganyu earner who had a large windfall payment on a single day will have their earnings overstated after converting the daily earnings into monthly wages. Nevertheless, our ganyu earnings compare very well with data from FISP as explained in Section 3.7.3.

3.4 Labour force participation

The IHS questionnaire contains four questions which allow us to compute labour force participation and unemployment rates for Malawi. The questions are:

a. Did you do work for any number of hours in the past 7 days?
b. Even though you did not do any activities in the last seven days, do you have a job or business or any economic activity to return to?
c. If you were offered a job, would you be willing to accept the job?
d. In the past four weeks, have you taken any action to look for any kind of work or start any kind of business / income generating activity?

The labour force is the sum of the employed and unemployed. Two standard definitions of unemployment exist in the literature, namely broad and narrow definitions. Broad unemployment is where a person is without work and available for work during the reference period. In this case, it

[22] See Appendix A2 for description of the Stata code.

includes responses in (a), (b) and (c). Narrow unemployment is where a person is without work and available for work during the reference period and is seeking work. This would be captured in (a), (b), (c) and (d). The broad definition counts all the jobless people who want jobs even though they did not search for one. The labour force participation comprises those employed and seeking work.

3.4.1 Size of labour force and labour force participation rates

Table 3.3 shows labour force participation for the balanced panel according to some selected characteristics. One key finding is that, regardless of the measure used, the labour force participation for the balanced panel has increased by almost 10 percentage points over the three-year period. For the broad measure, the labour force participation rate has increased from 81% in 2010 to 91% in 2013. The results indicate very high labour force participation rates - a key characteristic of low-income agricultural economies[23]. We also note some gender gaps in favour of males in terms of both labour force participation and employment rates. The gap has, however, marginally declined over three years.

Based on human capital theory, education does not only improve an individual's chances of getting employment but also positively impacts on earnings. Based on the table results, this theory seems to hold amongst those with some level of education, i.e. if we exclude those without formal education. The analysis shows that the labour force participation rate increases (due to higher employment rates) with the level of education and is highest amongst those with university education, reaching around 90% in both 2010 and 2013. It is, however, worth noting that those without formal education have equally high levels of labour force participation - registering rates of as high as 81% in 2010 and 91% in 2013 when measured in broad terms. Nonetheless, most of these are probably those who immediately enter the informal sector of the labour market because they do not have formal education.

With respect to age-groups, we note that compared to any other age groups, the (15-19)[24] age group has the lowest participation rate, at 58% in 2010 and about 77% in 2013. Consequently, as we will see, this age-group has the largest rates of unemployment although the situation is improving. The rise in the labour force participation rate between 2010 and 2013 is partly due to the fact that we are working with a balanced panel such that no new young entrants are captured in the second wave, i.e. wave 2 would only cover the 18 and 19-year olds.

[23] In Malawi as well as the rest of Sub-Saharan Africa, the traditionally high rates of labour force participation can be linked to the existence of very large agricultural sectors.

[24] Based on the IHS3 figures, about 57% of the Malawian population is aged between 0 and 19 years. In our data set, almost 22% is in the 15-19 age-group (National Statistical Office, 2012).

Table 3.3: Labour force participation rates according to characteristics

Description	Working age population				Broad labour force				Narrow labour force			
	2010	Share	2013	Share	2010	Share	2013	Share	2010	Share	2013	Share
Malawi	7,086	100%	8,106	100%	5,737	100%	7,336	100%	5,033	100%	6,578	100%
Sex												
Male	3,331	47%	3,869	48%	2,791	49%	3,526	48%	2,537	50%	3,241	49%
Female	3,755	53%	4,237	52%	2,946	51%	3,810	52%	2,496	50%	3,337	51%
Education												
None	5,275	74%	5,841	72%	4,278	75%	5,326	73%	3,790	75%	4,787	73%
PSLCE	754	11%	958	12%	589	10%	832	11%	498	10%	744	11%
JCE	576	8%	719	9%	449	8%	636	9%	376	7%	560	9%
MSCE	386	5%	463	6%	335	6%	430	6%	288	6%	385	6%
Tertiary	96	1%	125	2%	86	2%	112	2%	81	2%	102	2%
Age groups												
15-19	1,487	21%	1,796	22%	869	15%	1,387	19%	691	14%	1,195	18%
20-24	1,209	17%	1,311	16%	988	17%	1,179	16%	829	16%	1,000	15%
25-29	1,102	16%	1,203	15%	954	17%	1,142	16%	833	17%	1,026	16%
30-34	868	12%	1,023	13%	774	13%	986	13%	681	14%	894	14%
35-39	688	10%	762	9%	625	11%	734	10%	578	11%	685	10%
40-44	460	6%	604	7%	415	7%	583	8%	374	7%	541	8%
45-49	466	7%	442	5%	427	7%	420	6%	402	8%	399	6%
50-54	311	4%	407	5%	272	5%	391	5%	259	5%	362	6%
55-59	255	4%	288	4%	220	4%	271	4%	203	4%	250	4%
60-64	239	3%	269	3%	194	3%	242	3%	182	4%	224	3%
Regions												
Northern	924	13%	1,055	13%	740	13%	964	13%	665	13%	861	13%
Central	3,054	43%	3,541	44%	2,467	43%	3,174	43%	2,212	44%	2,875	44%
Southern	3,109	44%	3,510	43%	2,530	44%	3,198	44%	2,156	43%	2,842	43%

Source: Own calculations from IHPS data

3.4.2 Changes in the labour force according to background characteristics

In the next few paragraphs, we examine the patterns in the working age-population and labour force over the three years under study for individuals in the panel. The results are presented in Table 3.4 and show that there has been an increase of 14% in the number of people making up the working age population between 2010 and 2013 in the balanced sample. This increase is higher than the average GDP growth rate of around 3% experienced during the same period, raising fears about the economy's capacity to absorb the increasing labour supply. The labour force has grown both in broad and narrow terms by about 28% and 31%, respectively. These observed increases in the labour force could either be genuine or simply as a result of better capturing of the labour force in the administration of the questionnaire.

With respect to gender, there was a larger increase in the working age population for males compared to females. However, this pattern is reversed when we look at the labour force where females now have a higher growth rate than males for both broad and narrow definitions.

Amongst the education categories, the largest growth rate for the working age population was largest amongst those with tertiary education although the growth is from a very small number. However, the growth in the labour force has been the largest amongst those with PSCLE and JCE. This observation consistently holds for both the broad and narrow definitions of the labour force.

Table 3.4: Changes in the labour force according to characteristics

Description	Working age population			Broad labour force			Narrow labour force		
	2010	2013	% change	2010	2013	% change	2010	2013	% change
Malawi	7,086	8,106	14%	5,737	7,336	28%	5,033	6,578	31%
Sex									
Male	3,331	3,869	16%	2,791	3,526	26%	2,537	3,241	28%
Female	3,755	4,237	13%	2,946	3,810	29%	2,496	3,337	34%
Education									
None	5,275	5,841	11%	4,278	5,326	25%	3,790	4,787	26%
PSLCE	754	958	27%	589	832	41%	498	744	49%
JCE	576	719	25%	449	636	42%	376	560	49%
MSCE	386	463	20%	335	430	28%	288	385	34%
Tertiary	96	125	30%	86	112	30%	81	102	26%
Age groups									
15-19	1,487	1,796	21%	869	1,387	60%	691	1,195	73%
20-24	1,209	1,311	8%	988	1,179	19%	829	1,000	21%
25-29	1,102	1,203	9%	954	1,142	20%	833	1,026	23%
30-34	868	1,023	18%	774	986	27%	681	894	31%
35-39	688	762	11%	625	734	18%	578	685	19%
40-44	460	604	31%	415	583	41%	374	541	44%
45-49	466	442	-5%	427	420	-1%	402	399	-1%
50-54	311	407	31%	272	391	44%	259	362	40%
55-59	255	288	13%	220	271	23%	203	250	23%
60-64	239	269	12%	194	242	25%	182	224	23%
Regions									
Northern	924	1,055	14%	740	964	30%	665	861	29%
Central	3,054	3,541	16%	2,467	3,174	29%	2,212	2,875	30%
Southern	3,109	3,510	13%	2,530	3,198	26%	2,156	2,842	32%

Source: Own computations from IHPS data.

There have been positive growth rates across all age-groups except the (45-49) group which has registered a decline of about 5% in the working age population and a corresponding drop of 1% for both the broad and narrow definitions of the labour force.

All the three regions have registered increases in the working- age population numbers and this is also reflected in the broad and narrow labour force numbers. Using the broad definition, the Northern region has registered the largest growth rate in proportion to the rest. On the other hand, when the narrow definition is used, the largest increase is observed in the Southern region.

3.4.3 Shares in the labour force

According to results presented in Table 3.5, females make up the majority of the labour force regardless of whether the broad or narrow definition is used. The proportion of females in the labour force ranges between 50% and 53% and this reflects well with the population shares observed in the 2008 census data. However, proportionately fewer females enter the labour market compared to males. We will see this clearly when we discuss unemployment rates in Section 3.5.

Table 3.5: Shares of working-age population and labour force according to characteristics

Description	Working age population		Broad labour force participation				Narrow labour force participation			
			2010		2013		2010		2013	
	2010	2013	Number	Rate	Number	Rate	Number	Rate	Number	Rate
Malawi	7,086	8,106	5,737	81%	7,336	91%	5,033	71%	6,578	81%
Sex										
Male	3,331	3,869	2,791	84%	3,526	91%	2,537	76%	3,241	84%
Female	3,755	4,237	2,946	78%	3,810	90%	2,496	66%	3,337	79%
Education										
None	5,275	5,841	4,278	81%	5,326	91%	3,790	72%	4,787	82%
PSLCE	754	958	589	78%	832	87%	498	66%	744	78%
JCE	576	719	449	78%	636	88%	376	65%	560	78%
MSCE	386	463	335	87%	430	93%	288	75%	385	83%
Tertiary	96	125	86	90%	112	90%	81	84%	102	82%
Age groups										
15-19	1,487	1,796	869	58%	1,387	77%	691	46%	1,195	67%
20-24	1,209	1,311	988	82%	1,179	90%	829	69%	1,000	76%
25-29	1,102	1,203	954	87%	1,142	95%	833	76%	1,026	85%
30-34	868	1,023	774	89%	986	96%	681	78%	894	87%
35-39	688	762	625	91%	734	96%	578	84%	685	90%
40-44	460	604	415	90%	583	97%	374	81%	541	90%
45-49	466	442	427	92%	420	95%	402	86%	399	90%
50-54	311	407	272	88%	391	96%	259	84%	362	89%
55-59	255	288	220	86%	271	94%	203	80%	250	87%
60-64	239	269	194	81%	242	90%	182	76%	224	84%
Regions										
Northern	924	1,055	740	80%	964	91%	665	72%	861	82%
Central	3,054	3,541	2,467	81%	3,174	90%	2,212	72%	2,875	81%
Southern	3,109	3,510	2,530	81%	3,198	91%	2,156	69%	2,842	81%

Source: Own computations from IHPS data.

As far as education is concerned, we make two important observations. First, is that the labour force in Malawi is highly uneducated with percentages as high as between 72% and 75%. This is despite two decades of free primary education in Malawi. The second observation is that despite remaining highly uneducated, the labour force has become slightly more educated on average. As the table shows, the shares of those without formal education (None) have dropped by two percentage points from 74% to 72%. A similar percentage point decrease has been registered in the broad and narrow labour force figures. We also note that the shares of individuals with some education have increased over the three-year survey period. However, much of the increase in the levels of education is accounted for by PSLCE and JCE. MSCE and Tertiary have remained constant at 8% and 2%, respectively.

The table also shows that the majority of the work force in Malawi is made up of the youth and this is reflective of Malawi's population pyramid. Using the International Labour Organisation (ILO) definition, the youth (15-24) make up 38% of the total working-age-population. This proportion has been stable for both 2010 and 2013. On the other hand, using the Southern African Development Community (SADC) definition, the youth (15-34) consist of 66% in both 2010 and 2013. Regardless of the definition used, the youth make up the majority of the labour force.

We observe consistent shares in the working age population among regions for both the broad and narrow labour force definitions. The Northern region has the lowest shares across the survey periods partly because the population is also low there.

3.4.4 Multivariate analysis of labour force participation

We analyse the likelihood of labour force participation by the use of a probit regression. The results are presented in Table 3.6. Both models generate correctly predicted probabilities of at least 78% and this is quite high and good. Our dependent variable is a binary variable equal to '1' if an individual is in the labour force and '0', if not. For both the broad and narrow definitions, the explanatory variables are: sex of individuals (reference group: male), five year age categories (reference group: (15-19), education levels (reference group: none), marital status (reference group: married or non-formal union), age dependency ratio[25], region (reference group: Northern), year (reference group: 2010).

Under both the broad and narrow definitions, females are less likely to enter the labour market compared to males. Compared to the reference category, older individuals are more likely to participate in the labour market. We also note that labour force participation is generally concave with respect to age - first increasing before reaching a turning point.

After controlling for other factors and regardless of whether we use the broad or narrow definition, individuals with primary and junior secondary education are significantly less likely to participate in the labour market than the reference group who has no education. This is perhaps because they choose to continue with their education unlike those without education who might as well join the market straight away. In the narrow definition, we pick up an interesting and expected result where those with tertiary

[25] Age dependency ratio is given as the ratio of dependents (people younger than 15 or older than 64) to the working-age population (those aged 15-64).

education are more likely to enter the labour market compared to those without education, but this is not the case for the broad definition. As the lower labour force participation rate amongst those with some education could be the result of longer time spent studying, it is instructive to also see whether those without education also have a higher rate of participation if the sample is limited to those above 25 years.

Table 3.6: Probit regressions on labour force participation

Description	Broad		Narrow	
Female	-0.215***	(0.029)	-0.365***	(0.025)
Age categories				
20-24	0.496***	(0.040)	0.317***	(0.037)
25-29	0.743***	(0.050)	0.571***	(0.043)
30-34	0.833***	(0.058)	0.686***	(0.049)
35-39	0.852***	(0.067)	0.806***	(0.056)
40-44	0.793***	(0.075)	0.719***	(0.062)
45-49	0.870***	(0.080)	0.906***	(0.069)
50-54	0.782***	(0.086)	0.786***	(0.072)
55-59	0.609***	(0.091)	0.692***	(0.080)
60-64	0.459***	(0.084)	0.598***	(0.077)
Education level				
Primary	-0.076*	(0.042)	-0.099***	(0.037)
Junior secondary	-0.130***	(0.044)	-0.198***	(0.039)
Senior secondary	0.058	(0.057)	-0.043	(0.046)
Tertiary	0.078	(0.098)	0.209**	(0.085)
Not married	-0.521***	(0.033)	-0.398***	(0.030)
Age dependency ratio	0.051**	(0.021)	0.099***	(0.019)
Region				
Central	-0.110***	(0.037)	-0.026	(0.033)
Southern	0.005	(0.037)	-0.053*	(0.032)
Year=2013	0.532***	(0.027)	0.377***	(0.022)
Constant	0.760***	(0.052)	0.457***	(0.047)
Percent correctly predicted probability	86.13%		78.08%	
R-squared	0.162		0.123	
Observations	15,192		15,192	

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

Individuals who are unmarried (separated, divorced, widowed or never married) are less likely to enter the labour force compared to those that are married or living with a partner. Those who come from households with more children and elderly (high age dependency ratio) are more likely to participate in the labour market. In the broad and narrow definitions, individuals residing in the Central region and Southern region have a significantly lower likelihood of labour force participation compared to those in the Northern region. The conditional likelihood of labour force participation is greater in 2013, either indicating the presence of more job opportunities in 2013 than in 2010 or simply the fact that there are now more people looking for jobs.

3.5 Unemployment

From Table 3.4, we noted that there was a growth in the labour force regardless of whether the broad or narrow definition is used. However, the numbers of unemployed and unemployment rates shown in Table 3.7 have been relatively stable over the three years. Broad unemployment rates have remained constant at 13% while narrow unemployment has only marginally increased from 2% in 2010 to 3% in 2013. This means that much of the labour force growth was absorbed, probably on account of existing vacancies or new jobs created.

Table 3.7: Unemployment shares and rates by year

Description	Broadly unemployed						Narrowly unemployed					
	2010			2013			2010			2013		
	Number	Share	Rate	Number	Share	Rate	Number	Share	Rate	Number	Share	Rate
Malawi	743	100%	13%	938	100%	13%	118	100%	2%	222	100%	3%
Sex												
Male	273	37%	10%	371	40%	11%	59	50%	2%	104	47%	3%
Female	470	63%	16%	566	60%	15%	59	50%	2%	118	53%	4%
Education												
None	500	67%	12%	666	71%	13%	73	62%	2%	160	72%	3%
PSLCE	95	13%	16%	104	11%	12%	12	10%	2%	18	8%	2%
JCE	79	11%	18%	93	10%	15%	11	10%	3%	20	9%	4%
MSCE	60	8%	18%	59	6%	14%	17	14%	6%	18	8%	5%
Tertiary	8	1%	10%	15	2%	14%	4	4%	5%	6	3%	6%
Age groups												
15-19	175	24%	20%	208	22%	15%	8	7%	1%	26	12%	2%
20-24	183	25%	18%	218	23%	19%	36	31%	4%	43	20%	4%
25-29	144	19%	15%	154	16%	13%	35	29%	4%	43	20%	4%
30-34	97	13%	13%	124	13%	13%	15	13%	2%	38	17%	4%
35-39	47	6%	8%	61	7%	8%	9	8%	2%	17	8%	3%
40-44	37	5%	9%	52	6%	9%	3	3%	1%	15	7%	3%
45-49	23	3%	5%	30	3%	7%	5	5%	1%	11	5%	3%
50-54	10	1%	4%	43	5%	11%	2	1%	1%	18	8%	5%
55-59	17	2%	8%	24	3%	9%	3	2%	1%	5	2%	2%
60-64	10	1%	5%	23	2%	9%	2	1%	1%	7	3%	3%
Regions												
Northern	72	10%	10%	136	15%	14%	8	7%	1%	39	17%	4%
Central	268	36%	11%	355	38%	11%	47	40%	2%	70	32%	2%
Southern	403	54%	16%	446	48%	14%	63	53%	3%	113	51%	4%

Source: Own computations from IHPS data.

As also earlier discussed, there is an evident gender gap in favour of males as reflected in higher rates of unemployment for females. The gap is partly due to the fact that there are more females in the population but could also be as a result of gender discrimination. Nevertheless, the gap seems to be closing as shown by the decline in the share of females in broad unemployment from 63% in 2010 to 60% in 2013.

With respect to education, those without education make up the largest share of unemployment numbers and the proportions decline as the level of education increases, consistent with the human capital theory. However, there seems to be no clear pattern between unemployment rates and levels of education since unemployment is prevalent at all levels of education. While the unemployment rates are lower for those without education (because they can enter the 'labour' market earlier), they increase for individuals with PSLCE, JCE and MSCE, probably because these people are still in school. The level and rates of unemployment are the lowest amongst those with tertiary education.

The younger age groups not only make up the largest shares of unemployment figures but also experience the highest rates of unemployment, ranging between 13% and 20% when broadly measured and between 1% and 4% using narrow measurement.

3.6 Employment trends and characteristics

This section looks at employment trends between 2010 and 2013 as well as the characteristics of those in employment in Malawi. We also examine the work activities of the employed.

3.6.1 Employment shares and growth rates

Table 3.8 shows that the shares of employment have fairly remained stable by gender although females dominate in 2013 on account of growth. We observe a drop in the share of those without education from 76% in 2010 to a still high 73% in 2013. The youth, regardless of how we define them, dominate the shares of employment numbers. We note, however, that in 2013, the (15-19) age group dominates the share of employment at 18%, overtaking the (20-24) and (25-29) age groups which previously dominated in 2010.

On average, the number employed has grown by 28% and 29% for broad and narrow measures, respectively. A breakdown by gender shows that there was a larger increase for females than males regardless of the measure used. The largest growth in the number employed occurred amongst those with either PSLCE (49%) or JCE (48%) as their highest level of education. All age-groups, except those in the (45-49) group, have registered growth between 2010 and 2013. The largest increase (at least 70%) has been registered in the (15-19) age group. These are likely to be individuals who drop out of the education system in search of employment.

Table 3.8: Broad and narrow employment shares and growth by year

Description	Broadly employed					Narrowly employed				
	2010		2013		Growth	2010		2013		Growth
	Number	Share	Number	Share	(2010-2013)	Number	Share	Number	Share	(2010-2013)
Malawi	4,994	100%	6,398	100%	28%	4,915	100%	6,356	100%	29%
Sex										
Male	2,518	50%	3,155	49%	25%	2,478	50%	3,137	49%	27%
Female	2,476	50%	3,243	51%	31%	2,437	50%	3,219	51%	32%
Education										
None	3,777	76%	4,660	73%	23%	3,717	76%	4,627	73%	24%
PSLCE	494	10%	728	11%	47%	486	10%	725	11%	49%
JCE	370	7%	543	8%	47%	364	7%	540	8%	48%
MSCE	275	6%	371	6%	35%	271	6%	368	6%	36%
Tertiary	78	2%	97	2%	24%	77	2%	96	2%	25%
Age groups										
15-19	694	14%	1,179	18%	70%	683	14%	1,170	18%	71%
20-24	805	16%	961	15%	19%	793	16%	957	15%	21%
25-29	811	16%	989	15%	22%	798	16%	983	15%	23%
30-34	677	14%	862	13%	27%	666	14%	856	13%	29%
35-39	578	12%	673	11%	16%	569	12%	668	11%	17%
40-44	377	8%	531	8%	41%	371	8%	526	8%	42%
45-49	403	8%	390	6%	-3%	397	8%	389	6%	-2%
50-54	262	5%	348	5%	33%	258	5%	345	5%	34%
55-59	203	4%	247	4%	22%	200	4%	245	4%	23%
60-64	184	4%	219	3%	19%	181	4%	218	3%	20%
Regions										
Northern	668	13%	828	13%	24%	657	13%	823	13%	25%
Central	2,200	44%	2,819	44%	28%	2,165	44%	2,805	44%	30%
Southern	2,127	43%	2,751	43%	29%	2,093	43%	2,728	43%	30%

Source: Own computations from IHPS data

3.6.2 Multivariate analysis of employment likelihood

Not all people who are in the labour force end up being employed. This raises the problem of sample selection, which we address using the Heckman (1979) two-step procedure as discussed earlier in Section 3.2.3. In the first step, we estimate the probability of an individual either participating in the labour force or not. From the first step in Section 3.4.4, we obtained the inverse of the Mills ratio and used it as an additional explanatory variable in the second step, the probit on the likelihood of employment. Our results are presented in Table 3.9.

Table 3.9: Two-step Heckman probit results on employment likelihood

Description	Broad		Narrow	
Female	-0.220***	(0.029)	-0.076*	(0.045)
Age categories				
20-24	-0.171**	(0.067)	-0.070	(0.058)
25-29	-0.038	(0.083)	-0.019	(0.086)
30-34	0.073	(0.089)	0.040	(0.099)
35-39	0.204**	(0.092)	0.118	(0.109)
40-44	0.144	(0.092)	0.088	(0.104)
45-49	0.243**	(0.10)	0.140	(0.118)
50-54	0.156	(0.098)	0.071	(0.112)
55-59	0.204**	(0.098)	0.108	(0.112)
60-64	0.200**	(0.091)	0.075	(0.106)
Education level				
PSLCE	-0.051	(0.037)	-0.006	(0.039)
JCE	-0.132***	(0.040)	-0.035	(0.047)
MSCE	-0.164***	(0.044)	-0.085*	(0.045)
Tertiary	-0.018	(0.078)	-0.080	(0.078)
Not married	-0.118**	(0.048)	-0.108**	(0.051)
Region				
Central	0.063*	(0.033)	0.022	(0.032)
Southern	-0.072**	(0.031)	-0.034	(0.032)
Year=2013	0.002	(0.046)	0.026	(0.047)
Inverse mills ratio	-1.710***	(0.235)	-1.654***	(0.244)
Constant	1.284***	(0.113)	1.428***	(0.141)
Percent correctly predicted probability	75.63%		75.65%	
R-squared	0.106		0.106	
Observations	15,192		15,192	

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

First, we note that the coefficient of the inverse Mills ratio is statistically significant, indicating the presence of sample selection. It is, therefore, important to correct for selection bias. We then proceed to correct for selection bias and report heteroskedasticity consistent standard errors in the first stage. Under the broad definition, the results show also that the following groups are less likely to enter employment: females, those aged (20-24), those with JCE and MSCE, individuals who are not married, and people living in the Southern region, while those aged 35 and above as well as those from the Central region are more likely to be employed. Fewer variables are significant in the narrow definition, where we find that the following groups are significantly less likely to be employed: females, individuals with MSCE and those that are not married.

3.7 Earnings and changes in employment status

Table 3.10 shows that overall, the total average real monthly wages have increased by 37% from MWK 20,186 in 2010 to MWK27,704 in 2013. In the ensuing paragraphs, we attempt to breakdown and explain the sources of the increase. The first step is to examine how the employment status of individuals has changed between the two waves. Conditional on missing earnings, we came up with four employment statuses, unemployed in both waves (not shown in table), employed in either 2010 or 2013 only, and employed in both periods.

Table 3.10: Mean monthly total wages by employment status and survey period

	2010				2013			
	Mean	SD	N	Percent	Mean	SD	N	Percent
Employment status								
Wave 1 only	13,123	(56,494)	914	38%	.	.	.	.
Wave 2 only	.	.	.	.	18,552	(49,050)	1,411	43%
Both waves	24,696	(75,160)	1,520	62%	36,119	(96,957)	1,532	57%
Total	20,186	(68,713)	2,434	100%	27,704	(78,263)	2,943	100%

Source: Own computation from IHPS data, earnings expressed in constant 2013 prices.

3.7.1 Employed in either wave

The table shows that part of the increase in earnings is explained by the new entrants into the labour market (n=1,411) with average real monthly earnings of MWK18,552 compared to the MWK13,123 in 2010 of those who have now exited (n=914) the labour market. Between these two groups, average earnings have increased by 41%. Consequently, those that have exited the market have been replaced by higher earning individuals.

3.7.2 Employed in both waves

We observe that earnings are higher amongst individuals employed in both years compared to those only employed in either period. The gap in earnings between these two groups is stable, i.e. MWK13,123 versus MWK24,696 (1.88 times higher) in 2010 and MWK18,552 versus MWK36,119 (1.95 times more) in 2013. Moreover, those employed in both periods also experienced an increase in earnings; their earnings increased from MWK24,696 in 2010 to MWK36,119 in 2013, representing an increase of 46% over three years.

Considering the importance of ganyu employment in Malawi, the analysis in Table 3.10 is repeated for ganyu earners. The results are given in Table 3.11 where a similar pattern is observed. First, the largest increase in earnings is observed for individuals employed in both waves. Second, those employed in wave 2 only (new entrants) earn more compared to individuals in wave 1 only.

Table 3.11: Mean monthly ganyu wages by employment status and survey period

	2010				2013			
	Mean	SD	N	Percent	Mean	SD	N	Percent
Employment status								
Wave 1 only	9,234	(32,415)	711	47%	.	.	.	.
Wave 2 only	.	.	.	.	15,733	(42,866)	1,136	58%
Both waves	10,013	(16,498)	792	53%	20,231	(54,093)	835	42%
Total	9,649	(25,222)	1,503	100%	17,718	(48,183)	1,971	100%

Source: Own computation from IHPS data, earnings expressed in constant 2013 prices.

We further examine if there is any relationship between employment status and education attainment. Table 3.12 shows that of those without education qualification ("None"), 47% were unemployed in both waves and this forms the majority. On the other hand, 61% and 60% of those with university diploma and post-graduate degree, respectively, were employed in both periods. New entrants into the labour market (wave 2 only) have more education compared to those that have dropped out of the labour market (wave 1 only).

Table 3.12: Employment status and education attainment of individuals

Description	None	PSCLE	JCE	MSCE	Non-University Diploma	University Diploma	Post-graduate degree	Total
Unemployed in both waves	5,253	960	756	350	26	11	3	7,358
	47%	*56%*	*58%*	*41%*	*20%*	*16%*	*13%*	*48%*
Employed in wave 1 only	1,616	173	122	74	18	6	4	2,013
	15%	*10%*	*9%*	*9%*	*14%*	*10%*	*19%*	*13%*
Employed in wave 2 only	2,106	268	160	126	17	9	2	2,687
	19%	*16%*	*12%*	*15%*	*13%*	*13%*	*8%*	*18%*
Employed in both waves	2,143	311	256	300	70	41	14	3,135
	19%	*18%*	*20%*	*35%*	*54%*	*61%*	*60%*	*21%*
Total	11,118	1,712	1,295	849	130	67	23	15,194
	100%	*100%*	*100%*	*100%*	*100%*	*100%*	*100%*	*100%*

Source: Own computation from IHPS data, percentages in italics.

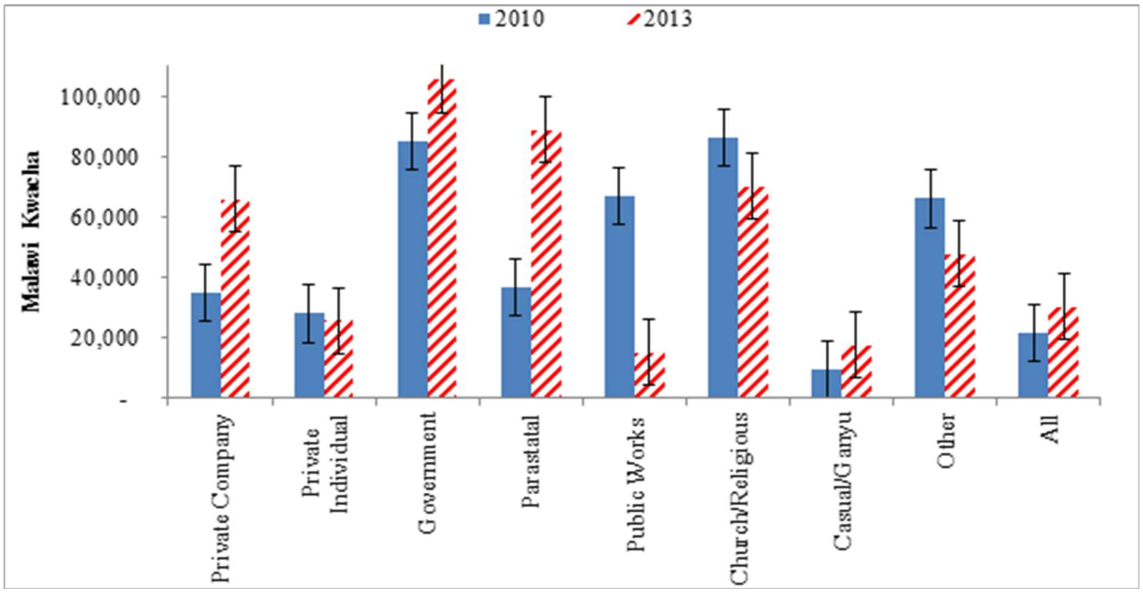
3.7.3 Identifying sources of increases in earnings

Figure 3.1 shows real monthly wages by occupation and survey year. Although we do not fully know the reasons behind the changes, it is important to point out two observations. Firstly, as we will note in the figure, the average earnings in the NGO sector (PWP and churches) have dropped between 2010 and 2013. These might have been negatively affected by donor aid withdrawals that Malawi faced recently.[26]

[26] The "Other" category is made up of 7 and 19 individuals in 2010 and 2013, respectively. Perhaps we need to add them up to the dominant category or a category with similar characteristics.

Secondly, our results compare well with Farm Input Subsidy Programme (FISP) evaluation panel data currently being analysed by other researchers in Malawi where large increases have been observed in real ganyu wages (nominal ganyu daily wage rates divided by maize prices) for some districts between 2012 and 2015 (see Figure A1 in the appendix).

Figure 3.1: Histogram of real monthly earnings by survey year and occupation



Source: Own computation from IHPS data, bars represent standard errors.

Given this comparison, one can argue that the increases in the earnings may be genuine although it may be difficult to isolate the main factors from the many drivers behind this rise. Gross Domestic Product (GDP) growth may be a contributor but the economy only grew at an average of about 3% per year. As for ganyu wages, it is possible that with the availability of food in many parts of the country, the ganyu workers tended to have higher bargaining power and asked for more wages.

It is also worth noting that in real terms, minimum wages were adjusted upwards twice between 2010 and 2013; first by 50% effective 1st January 2011 and second by 34% effective 1st July 2013. These could explain the increases in earnings in the private and government sectors, although this largely depends on effective implementation and monitoring and would only affect low-income workers. Despite working the same number of hours, those in government received significantly more per month than their counterparts in parastatal companies. Nevertheless, parastatal organisations more than doubled their earnings between 2010 and 2013. There is some catch up between the two groups.

3.8 Econometric analysis of returns to education

As explained in Section 3.3, the IHS questionnaire collects information on wage employment and household non-farm enterprises. The former is analysed at the individual and the latter at the household level. It is, however, possible to link household to individual data by enterprise owner.

Earnings from household non-farm enterprises are analysed at a household rather than the individual level for three reasons. Firstly, we reason that wages and enterprise earnings (profit) may not necessarily be explained by the same factors. This is confirmed because the R-squares for the wage regressions improve after splitting the analysis (for wages and household earnings separately). Secondly, the IHS data have a separate module for household enterprises with background characteristics such as industry, customers and age of enterprises which allows us to achieve this purpose. Thirdly, looking at enterprises separately enables us to explore the existence of education externalities within the household. We analyse externalities by using maximum education in the household. Average household education is used for robustness checks and also works well. We begin our analysis by looking at wage employment before moving on to enterprise earnings in Section 3.8.2.

3.8.1 Wage employment

Recall that according to the human capital theory, investment in education improves workers' skills resulting in high productivity and, therefore, higher earnings (Mincer, 1974). We begin our analysis with ordinary least squares estimation (OLS) for 2010, 2013 and the pooled sample. The results are given in Table 3.13 where the dependent variable is the log of real monthly earnings[27]. The explanatory variables are years of schooling, years of experience, square of experience, sex (base category: male), region (base: Northern) and occupation (base: private individual).

Across the three models, the results show that there are large returns to education of between 6.8% and 7.7%. The strongly positive returns to education are consistent with other findings in Malawi (Chirwa & Matita 2009; Chirwa & Zgovu 2002). Similar results have been found in other African countries (e.g., in Cameroon by Ewoudou & Vencatachellum, 2006; in Rwanda by Lassibille & Tan, 2005); Bennell, 1996, for Sub-Saharan Africa). The negative and significant gender dummy is consistent with the general finding that females earn less than their male counterparts (Chirwa & Matita, 2009). We also find that average earnings for private individuals are significantly lower than those in private companies. As expected, ganyu earnings are consistently significantly lower in all the models. It is only in the 2013

[27] The interpretation of the coefficients is the percentage change in the monthly earnings given a unit change in the explanatory variable. For dummy variables the percentage effect of a change from the base category.

model that workers in public works programme earn significantly less than those in private companies, albeit only at the 10% level of significance.

Table 3.13: OLS results for log of real monthly wages

Description	OLS		
	2010	2013	Pooled
Years of schooling	0.077***	0.068***	0.072***
	(0.008)	(0.009)	(0.008)
Experience	0.037***	0.055***	0.047***
	(0.006)	(0.006)	(0.005)
Experience squared	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
Female	-0.260***	-0.276***	-0.267***
	(0.045)	(0.042)	(0.032)
Central	-0.101	-0.019	-0.054
	(0.081)	(0.098)	(0.075)
Southern	-0.257***	-0.144	-0.202***
	(0.074)	(0.096)	(0.066)
Private Individual	-0.210*	-0.555***	-0.392***
	(0.108)	(0.126)	(0.094)
Government	0.206	0.232	0.206*
	(0.130)	(0.162)	(0.122)
Parastatal	0.179	0.386*	0.305**
	(0.196)	(0.197)	(0.145)
Public Works Program	0.555	-0.262*	0.094
	(0.376)	(0.157)	(0.193)
Church/Religious Organisation	-0.024	0.171	0.107
	(0.335)	(0.234)	(0.231)
Other	0.135	0.036	0.128
	(0.490)	(0.283)	(0.284)
Casual/Ganyu	-0.803***	-0.805***	-0.795***
	(0.094)	(0.105)	(0.083)
Constant	8.962***	9.080***	9.015***
	(0.153)	(0.168)	(0.137)
R-squared	0.340	0.282	0.295
Observations	2,434	2,943	5,377

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

3.8.1.1 Homogeneous returns to education

Next, we discuss findings from three models, namely OLS, fixed effects and random effects beginning with estimation results presented in Table 3.14. In each of these models, we assume that returns to education are homogeneous and also ignore selectivity bias[28].

Table 3.14: OLS, Fixed effects and random effects results for log of real monthly wages

Description	Pooled OLS			Fixed effects			Random effects		
	1	2	3	4	5	6	7	8	9
Years of schooling	0.109***	0.072***	0.072***	0.009	0.055***	0.015	0.115***	0.086***	0.083***
	(0.007)	(0.008)	(0.007)	(0.008)	(0.010)	(0.015)	(0.004)	(0.005)	(0.005)
Experience		0.047***	0.047***		0.068***	0.018		0.043***	0.041***
		(0.005)	(0.005)		(0.011)	(0.017)		(0.004)	(0.004)
Experience squared		-0.001***	-0.001***		-0.000*	0.000		-0.001***	-0.001***
		(0.000)	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)
Female		-0.267***	-0.268***					-0.277***	-0.280***
		(0.032)	(0.032)					(0.030)	(0.030)
Central		-0.054	-0.05		0.567***	0.418**		-0.030	-0.036
		(0.075)	(0.074)		(0.208)	(0.203)		(0.037)	(0.037)
Southern		-0.202***	-0.194***		0.365	0.281		-0.179***	-0.183***
		(0.066)	(0.066)		(0.265)	(0.256)		(0.037)	(0.037)
Private Individual		-0.392***	-0.369***		-0.113	-0.111		-0.302***	-0.292***
		(0.094)	(0.096)		(0.081)	(0.079)		(0.056)	(0.056)
Government		0.206*	0.216*		0.085	0.070		0.246***	0.253***
		(0.122)	(0.122)		(0.145)	(0.151)		(0.070)	(0.070)
Parastatal		0.305**	0.298**		0.290**	0.275**		0.355***	0.371***
		(0.145)	(0.139)		(0.126)	(0.127)		(0.111)	(0.109)
Public Works Program		0.094	0.061		-0.247	-0.305		0.177	0.143
		(0.193)	(0.210)		(0.251)	(0.241)		(0.180)	(0.186)
Church/Religious Organisation		0.107	0.091		0.172	0.114		0.155	0.144
		(0.231)	(0.225)		(0.169)	(0.166)		(0.124)	(0.124)
Other		0.128	0.064		0.352**	0.308**		0.522***	0.443**
		(0.284)	(0.280)		(0.164)	(0.151)		(0.178)	(0.175)
Casual/Ganyu		-0.795***	-0.805***		-0.487***	-0.495***		-0.731***	-0.754***
		(0.083)	(0.085)		(0.089)	(0.088)		(0.049)	(0.048)
Year=2013			0.297***			0.263***			0.240***
			(0.039)			(0.053)			(0.021)
Constant	8.418***	9.015***	8.868***	9.306***	7.816***	8.918***	8.382***	8.846***	8.791***
	(0.057)	(0.137)	(0.133)	(0.066)	(0.253)	(0.374)	(0.034)	(0.085)	(0.084)
R-squared	0.137	0.295	0.313	0.638	0.672	0.685	0.233	0.374	0.383
Observations	5,377	5,377	5,377	5,377	5,377	5,377	5,377	5,377	5,377

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

The fixed effects model does not work well, indicating very low within variation[29] in our variables especially after the inclusion of the time dummy. The time dummy further reduces the variation in the data and this somehow causes the results to change dramatically[30]. As the results show, the coefficient for education in the OLS regression is almost 5 times as bigger as that of fixed effects- a small coefficient

[28] Selection bias is explained in Section 3.2.3.

[29] Low within variation refers to a situation where independent variables change very gradually over time.

[30] We tested as to whether we need to be using a time dummy and the results support its inclusion.

for the fixed effect compared to OLS is a sign that there is little within variation. Moreover, in the fixed effects model, the coefficient for education is insignificant. We, therefore, concentrate on OLS and random effects which yield consistent results except for differences in the magnitudes in selected cases. The Breusch Pagan LM test[31] yields significant results, indicating that the random effects model is more appropriate compared to OLS. The time dummy shows that average monthly earnings are higher in 2013 compared to 2010.

3.8.1.2 Heterogeneous returns to education

The results from the basic models discussed above disregard the differences in the level of educational attainment by looking at a single overall education level- years of schooling. This homogeneous model assumes that there are no differential trends in the returns to education for different levels of education. As earlier discussed, there is little statistical evidence and causal empiricism for the homogenous model. The heterogeneous model provides the alternative and looks at the different levels of education as having separate effects on earnings. Using this model specification, we replace years of schooling (S) with an educational dummy variable to represent the different educational categories discussed in Section 3.2.2.

The results are presented in Table 3.15 and we provide a graphical illustration in Figure 3.2 but based on the random effects only. Regardless of gender, the returns to education increase with the level of education supporting a convex relationship between education and earnings. Our results confirm the finding that the returns increase with the level of schooling in Sub-Saharan Africa but are also contrary to the literature supporting concave rates of returns such as Psacharopoulos (1994).

Also consistent with the international literature is the finding that female workers tend to have much higher returns on education than male workers, particularly at higher levels of education. Similar findings have been established in Malawi by Chirwa and Matita (2009). Although more men enter the labour market, they tend to have lower returns than their female counterparts with similar education levels suggesting that female education is more effective in generating returns in Malawi. The high rates of return to higher education for females and tertiary education may mean that gender discrimination favouring men is reduced at higher levels of education. It is worth noting that the education system in Malawi has greatly emphasised on primary education especially with the introduction of free primary education in 1994. There has been a lot of expansion in primary education attainment and this possibly explains the low returns.

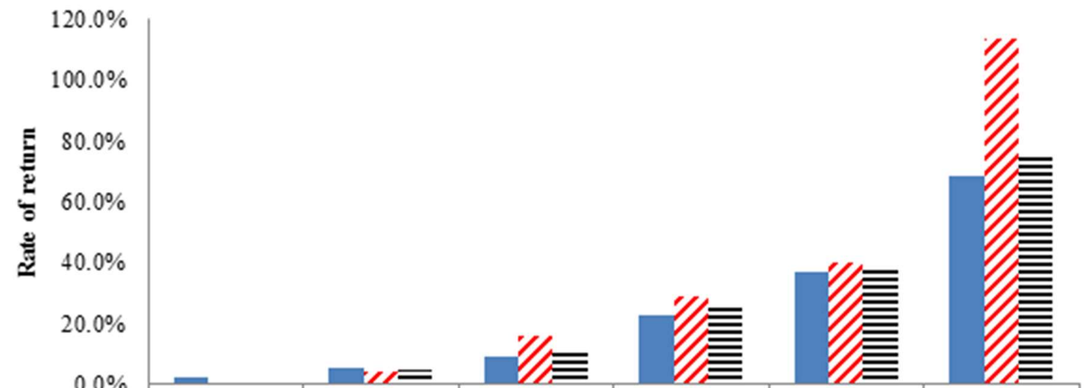
[31] The test helps a researcher decide between a random effects regression and a simple OLS regression (e.g., Baltagi, B.H. & Li, Q., 1990). The null hypothesis is that variances across individuals are zero. Put differently, that there is no significant difference across units (i.e. no panel effect).

Table 3.15: OLS and random effects results for log of monthly wages using education categories

Description	Pooled OLS			Random effects		
	Male	Female	All	Male	Female	All
PSLCE	0.168**	-0.022	0.120*	0.162***	0.052	0.141***
	(0.076)	(0.117)	(0.067)	(0.051)	(0.082)	(0.043)
JCE	0.341***	0.428**	0.356***	0.411***	0.351***	0.401***
	(0.088)	(0.208)	(0.087)	(0.057)	(0.103)	(0.050)
MSCE	0.670***	1.184***	0.753***	0.726***	1.071***	0.800***
	(0.106)	(0.166)	(0.094)	(0.063)	(0.132)	(0.057)
Non-University Diploma	1.430***	1.617***	1.498***	1.435***	1.612***	1.503***
	(0.149)	(0.206)	(0.135)	(0.093)	(0.162)	(0.082)
University Diploma	1.891***	2.085***	1.995***	1.932***	1.999***	1.968***
	(0.173)	(0.330)	(0.160)	(0.132)	(0.20)	(0.109)
Post-Graduate Diploma	2.773***	3.220***	2.819***	2.590***	3.069***	2.673***
	(0.299)	(0.295)	(0.314)	(0.206)	(0.290)	(0.186)
Experience	0.056***	0.034***	0.045***	0.046***	0.027***	0.037***
	(0.005)	(0.006)	(0.004)	(0.005)	(0.005)	(0.004)
Experience squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.000***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female			-0.273***			-0.282***
			(0.032)			(0.029)
Central	-0.101	-0.07	-0.09	-0.088*	-0.085	-0.087**
	(0.081)	(0.087)	(0.074)	(0.046)	(0.057)	(0.036)
Southern	-0.214***	-0.238***	-0.227***	-0.194***	-0.234***	-0.214***
	(0.068)	(0.083)	(0.063)	(0.045)	(0.058)	(0.036)
Private Individual	-0.268***	-0.339*	-0.292***	-0.198***	-0.254*	-0.208***
	(0.096)	(0.193)	(0.090)	(0.060)	(0.134)	(0.055)
Government	0.119	-0.004	0.113	0.154**	0.061	0.149**
	(0.129)	(0.179)	(0.110)	(0.075)	(0.135)	(0.066)
Parastatal	0.379***	-0.188	0.247**	0.279***	0.307	0.311***
	(0.139)	(0.186)	(0.122)	(0.102)	(0.252)	(0.102)
Public Works Program	0.052	0.162	0.139	0.136	0.211	0.204
	(0.275)	(0.229)	(0.204)	(0.266)	(0.236)	(0.170)
Church/Religious Organisation	-0.233	0.53	0.009	0.015	0.263	0.1
	(0.165)	(0.379)	(0.210)	(0.143)	(0.221)	(0.119)
Other	0.066	0.01	0.044	0.349*	0.281	0.324**
	(0.293)	(0.316)	(0.238)	(0.203)	(0.218)	(0.145)
Casual/Ganyu	-0.637***	-0.784***	-0.698***	-0.576***	-0.744***	-0.632***
	(0.094)	(0.160)	(0.086)	(0.054)	(0.122)	(0.049)
Year=2013	0.282***	0.298***	0.289***	0.237***	0.220***	0.234***
	(0.038)	(0.054)	(0.038)	(0.027)	(0.035)	(0.021)
Constant	9.100***	9.124***	9.234***	9.103***	9.155***	9.232***
	(0.119)	(0.183)	(0.111)	(0.083)	(0.145)	(0.069)
R-squared	0.308	0.316	0.337	0.385	0.424	0.423
Observations	3,244	2,133	5,377	3,244	2,133	5,377

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

Figure 3.2: Rates of return on education by gender



	PSLCE	JCE	MSCE	Non-Uni. Diploma	Uni. Diploma	P.G. Diploma
Male	2.2%	5.1%	8.9%	22.9%	36.9%	68.5%
Female	0.7%	4.2%	16.0%	28.7%	39.9%	114.0%
All	1.9%	4.9%	10.2%	25.0%	38.5%	74.9%

Source: Own computation from IHPS data.

3.8.1.3 Robustness checks

We conduct five main robustness checks to see if our main results are preserved under different conditions and assumptions. These include decomposition of results by gender, analysis of returns to education only based on individuals employed in both waves, alternative treatment of outliers, sample selection and distinguishing the results by sector. We discuss these in the paragraphs that follow.

Firstly, we decompose our results by gender. We report OLS and random effects regression results in Table 3.16 having split the sample by the gender of earners. A version of this segregation is already discussed in Section 3.8.1.2 but assuming heterogeneous returns to education. The table shows that the key results after dividing the sample into male and female sub-samples are preserved. The difference for the coefficient comparison test (male minus female) reveals that the coefficients for females are smaller compared to men in most cases as shown by positive and significant differences for years of schooling, experience, government and ganyu. The difference is the largest for ganyu employment. The only case where females have a larger coefficient is within church/religious organisations where the magnitude is very surprisingly large.

Secondly, we consider individuals employed in both waves only. As we earlier noted in Section 3.7, these individuals experienced the largest jump in earnings over three years and also have the highest levels of education. The results from Table 3.16 (last two columns) do not differ much compared to those based on the full sample (which includes those only employed in either waves 1 or 2).

Table 3.16: Regressions of log monthly wages by gender and employees in both waves

Description	Pooled OLS			Random effects		Employed in both waves	
	Male	Female	Difference	Male	Female	OLS	Random
Years of schooling	0.078***	0.056***	0.022**	0.089***	0.068***	0.076***	0.083***
	(0.008)	(0.011)		(0.006)	(0.008)	(0.008)	(0.006)
Experience	0.056***	0.035***	0.021**	0.049***	0.030***	0.039***	0.031***
	(0.006)	(0.007)		(0.005)	(0.006)	(0.006)	(0.006)
Experience squared	-0.001***	-0.001***	0.000**	-0.001***	-0.000***	-0.001***	-0.000***
	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
Female						-0.174***	-0.194***
						(0.048)	(0.043)
Central	-0.059	-0.034	-0.025	-0.036	-0.034	-0.059	-0.023
	(0.080)	(0.090)		(0.047)	(0.059)	(0.084)	(0.053)
Southern	-0.173**	-0.227***	0.054	-0.158***	-0.216***	-0.166**	-0.139***
	(0.070)	(0.085)		(0.046)	(0.060)	(0.076)	(0.051)
Private Individual	-0.303***	-0.618***	0.315	-0.237***	-0.508***	-0.374***	-0.263***
	(0.097)	(0.20)		(0.060)	(0.139)	(0.105)	(0.065)
Government	0.209	0.163	0.046**	0.246***	0.196	0.207	0.254***
	(0.138)	(0.201)		(0.078)	(0.152)	(0.137)	(0.080)
Parastatal	0.430***	-0.182	0.612	0.324***	0.414*	0.376**	0.412***
	(0.155)	(0.253)		(0.117)	(0.234)	(0.153)	(0.117)
Public Works Program	0.024	-0.117	0.141	0.12	-0.056	-0.204	-0.245*
	(0.280)	(0.237)		(0.294)	(0.243)	(0.186)	(0.147)
Church/Religious Organisation	-0.174	0.591	-0.765**	0.026	0.372	0.122	0.208
	(0.179)	(0.392)		(0.144)	(0.245)	(0.248)	(0.139)
Other	0.029	0.036	-0.007	0.421*	0.43	-0.179	0.264
	(0.337)	(0.431)		(0.242)	(0.287)	(0.270)	(0.180)
Casual/Ganyu	-0.692***	-1.108***	0.416**	-0.646***	-1.057***	-0.742***	-0.678***
	(0.088)	(0.167)		(0.053)	(0.115)	(0.086)	(0.057)
Year=2013	0.286***	0.297***	-0.011	0.240***	0.221***	0.328***	0.256***
	(0.039)	(0.056)		(0.027)	(0.035)	(0.042)	(0.026)
Constant	8.636***	9.101***	-0.465**	8.566***	9.026***	8.923***	8.922***
	(0.135)	(0.227)		(0.103)	(0.161)	(0.145)	(0.111)
R-squared	0.287	0.278		0.348	0.376	0.348	0.411
Observations	3,244	2,133		3,244	2,133	3,052	3,052

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

Thirdly, we compare results based on alternative treatment of outliers. Table 3.17 gives results based on four alternative ways of dealing with outliers discussed in Section 3.3.3. Overall, the results do not change much considering both magnitude and direction, implying that the method of outlier treatment does not really matter. However, failure to account for outliers yields larger returns to education and this would be more evident if we decompose by groups such as sector.

Table 3.17: Random effect results based on alternative treatment of outliers

Description	Random effects				
	Outliers included	Millionaires	100th percentile	Robust regression	Extreme residuals
Years of schooling	0.087***	0.081***	0.077***	0.085***	0.083***
	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
Experience	0.042***	0.041***	0.040***	0.040***	0.041***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Experience squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.275***	-0.279***	-0.274***	-0.273***	-0.280***
	(0.030)	(0.030)	(0.029)	(0.029)	(0.030)
Central	-0.04	-0.024	-0.043	-0.042	-0.036
	(0.038)	(0.036)	(0.036)	(0.036)	(0.037)
Southern	-0.181***	-0.168***	-0.189***	-0.189***	-0.183***
	(0.038)	(0.036)	(0.035)	(0.036)	(0.037)
Private Individual	-0.300***	-0.285***	-0.278***	-0.296***	-0.292***
	(0.057)	(0.055)	(0.054)	(0.055)	(0.056)
Government	0.228***	0.180***	0.162***	0.235***	0.253***
	(0.072)	(0.064)	(0.061)	(0.069)	(0.070)
Parastatal	0.298**	0.362***	0.344***	0.362***	0.371***
	(0.125)	(0.107)	(0.103)	(0.108)	(0.109)
Public Works Program	0.123	0.134	0.055	0.139	0.143
	(0.186)	(0.186)	(0.165)	(0.185)	(0.186)
Church/Religious Organisation	0.114	0.095	0.045	0.066	0.144
	(0.125)	(0.115)	(0.110)	(0.113)	(0.124)
Other	0.410**	0.435**	0.417**	0.431**	0.443**
	(0.174)	(0.174)	(0.176)	(0.173)	(0.175)
Casual/Ganyu	-0.760***	-0.748***	-0.746***	-0.762***	-0.754***
	(0.049)	(0.048)	(0.047)	(0.048)	(0.048)
Year=2013	0.246***	0.242***	0.243***	0.233***	0.240***
	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)
Constant	8.759***	8.786***	8.826***	8.786***	8.791***
	(0.085)	(0.083)	(0.081)	(0.083)	(0.084)
R-squared	0.380	0.370	0.370	0.397	0.383
Observations	5,391	5,367	5,335	5,356	5,377

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

Fourthly, we address sample selection which as indicated can bias results if not addressed in the earnings regression. We, therefore, correct for sample selection using the Heckman selection model as earlier explained. In the first step, we model the probability of labour force participation. In the second step, we model earnings as explained in equation 2 where the inverse mills ratio is now added as an additional explanatory variable. The sample size is reduced by two observations due to the omission of missing values in the process of generating the mills ratio from the first stage.

The sample selection corrected results presented in Table 3.18 and do not differ much from the uncorrected findings in Table 3.13 and Table 3.14. A simple interpretation of this is that sample selection is not a very serious issue in our data.

Table 3.18: Wage functions corrected for sample selection

Description	OLS (2010)	OLS (2013)	Pooled OLS	Random
Years of schooling	0.067***	0.061***	0.065***	0.077***
	(0.009)	(0.008)	(0.008)	(0.005)
Experience	0.014*	0.030***	0.027***	0.024***
	(0.008)	(0.008)	(0.006)	(0.005)
Experience squared	-0.000*	-0.001***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Female	-0.183***	-0.257***	-0.224***	-0.242***
	(0.048)	(0.042)	(0.031)	(0.030)
Central	-0.091	0.018	-0.029	-0.018
	(0.081)	(0.099)	(0.075)	(0.037)
Southern	-0.266***	-0.139	-0.195***	-0.186***
	(0.075)	(0.096)	(0.066)	(0.037)
Private Individual	-0.200*	-0.563***	-0.365***	-0.287***
	(0.106)	(0.125)	(0.094)	(0.056)
Government	0.218*	0.231	0.220*	0.256***
	(0.131)	(0.162)	(0.122)	(0.070)
Parastatal	0.184	0.377*	0.295**	0.377***
	(0.187)	(0.201)	(0.138)	(0.110)
Public Works Programme	0.528	-0.251	0.054	0.129
	(0.371)	(0.157)	(0.205)	(0.181)
Church/Religious Organisation	-0.033	0.167	0.089	0.137
	(0.337)	(0.229)	(0.225)	(0.124)
Other	0.156	0.016	0.059	0.443**
	(0.479)	(0.277)	(0.275)	(0.175)
Casual/Ganyu	-0.788***	-0.804***	-0.796***	-0.743***
	(0.095)	(0.105)	(0.084)	(0.048)
Inverse mills ratio_2010	-0.713***			
	(0.195)			
Inverse mills ratio_2013		-1.031***		
		(0.330)		
Inverse mills ratio			-0.722***	-0.675***
			(0.182)	(0.132)
Year=2013			0.208***	0.162***
			(0.044)	(0.025)
Constant	9.401***	9.503***	9.274***	9.155***
	(0.187)	(0.204)	(0.163)	(0.107)
R-squared	0.346	0.287	0.317	0.387
Observations	2,434	2,941	5,375	5,375

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

However, it is important to correct for sample selection as we have done given that the inverse mills ratio is significant in each of the four models. Without sample correction, the returns to education tend to be larger by between 0.7 and 0.8 percentage points than with sample selection. Our estimates are, therefore, upwardly biased when sample selection is not addressed. Self-selection between sectors is addressed in the section that follows.

The final robustness check is to distinguish between formal and informal sectors. Just like in many other developing countries, the formal sector in Malawi only absorbs a small percentage of the labour force (Chirwa & Matita, 2009). In our data set, about 77.54% of individuals aged between 15 and 64 years with positive earnings are employed in the informal sector. Due to the large size of the informal sector, studying earning differentials between economic sectors becomes important for policy.

The process of accounting for self-selection is the same as before; we first run a probit on the choice of employment sector (see Table 3.19) from which we obtain the inverse mills ratio used in the second stage. The first stage results show that the following are less likely to enter the formal sector: females, those not married and individuals from households with high age dependency ratio. The rest of the explanatory variables are associated with a higher probability of entering the formal sector.

Table 3.19: Probit on choice of employment sector

Explanatory variables	Coefficient	Standard error
Female	-0.439***	(0.045)
Age groups		
20-24	0.278***	(0.098)
25-29	0.607***	(0.096)
30-34	0.911***	(0.099)
35-39	0.913***	(0.103)
40-44	1.079***	(0.110)
45-49	1.040***	(0.113)
50-54	1.093***	(0.124)
55-59	0.906***	(0.140)
60-64	0.737***	(0.141)
PSLCE	0.512***	(0.062)
JCE	0.990***	(0.067)
MSCE	1.867***	(0.077)
Non-University Diploma	2.343***	(0.166)
University Diploma	2.691***	(0.267)
Post Graduate degree	2.420***	(0.412)
Not married	-0.096*	(0.051)
Age dependency ratio	-0.212***	(0.033)
Central	0.160***	(0.056)
Southern	0.296***	(0.055)
Year=2013	-0.169***	(0.041)
Constant	-1.276***	(0.104)
Percent correctly predicted	80.10%	
Pseudo R-squared	0.315	
Observations	5,416	

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

For each of the two sectors, we present two sets of results, namely without and with self-selection correction. To achieve this, we distinguish between the formal and informal sectors. Sector decomposition analysis of earnings sheds some light on the importance of distinguishing between the different types of employment sector in estimating returns to education in developing countries. Studies that fail to take this into account tend to overstate the returns to education by assuming that returns are the same in both the formal and informal sectors. We with uncorrected results (without sample selection) in Table 3.20 where we also conduct the coefficient comparison test.

Table 3.20: Regression results on log of monthly wages by sector without sample selection

Description	Pooled OLS			Random effects	
	Formal	Informal	Difference	Formal	Informal
Years of schooling	0.099***	0.037***	0.062***	0.086***	0.047***
	(0.011)	(0.009)		(0.006)	(0.007)
Experience	0.041***	0.048***	-0.007	0.027***	0.048***
	(0.008)	(0.006)		(0.006)	(0.004)
Experience squared	-0.001***	-0.001***	0.000	-0.000**	-0.001***
	(0.000)	(0.000)		(0.000)	(0.000)
Female	-0.079	-0.345***	0.266***	-0.117*	-0.368***
	(0.076)	(0.033)		(0.066)	(0.032)
Central	0.030	-0.113	0.143	0.085	-0.122***
	(0.136)	(0.079)		(0.073)	(0.040)
Southern	-0.072	-0.264***	0.192	-0.052	-0.263***
	(0.132)	(0.070)		(0.071)	(0.040)
Private Individual	-0.328***			-0.243***	
	(0.092)			(0.056)	
Government	0.082			0.171**	
	(0.128)			(0.068)	
Parastatal	0.262**			0.347***	
	(0.130)			(0.098)	
Public Works Programme	0.054			0.113	
	(0.216)			(0.182)	
Church/Religious Organisation	0.031			0.158	
	(0.215)			(0.126)	
Other	0.025			0.353**	
	(0.276)			(0.139)	
Year=2013	0.204***	0.322***	-0.118	0.164***	0.280***
	(0.060)	(0.047)		(0.030)	(0.028)
Constant	8.576***	8.348***	0.228	8.767***	8.308***
	(0.221)	(0.115)		(0.135)	(0.078)
R-squared	0.238	0.120		0.297	0.132
Observations	1,909	3,468		1,909	3,468

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

The returns to education are positive in both sectors but with larger magnitudes in the formal sector for both OLS and random effects. It is not surprising for the informal sector to have positive returns to education. Previous research in Malawi by Chirwa and Zgovu (2002) has also shown that casual employment (informal sector wage employment) on peak labour tasks may be well paid, above the daily minimum wage, but short-lived.

The convex relationship between earnings and experience is also maintained in both sectors except that the coefficient is bigger in the informal sector than the formal sector. This implies that experience matters more in the informal sector. The time dummy coefficient is also positive in both models but larger in the informal sector, an indication that earnings grew much stronger within the informal sector. Next, we look at results corrected for sample selection in Table 3.21.

Table 3.21: Regression results on log of monthly wages by sector with sample selection

Description	Pooled OLS		Random effects	
	Formal	Informal	Formal	Informal
Years of schooling	0.058***	0.012	0.044***	0.015*
	(0.014)	(0.011)	(0.007)	(0.008)
Experience	0.029***	0.030***	0.012*	0.027***
	(0.010)	(0.008)	(0.006)	(0.005)
Experience squared	-0.001***	-0.001***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.070	-0.207***	0.041	-0.210***
	(0.087)	(0.048)	(0.065)	(0.041)
Central	-0.005	-0.154*	0.023	-0.169***
	(0.137)	(0.082)	(0.073)	(0.040)
Southern	-0.153	-0.329***	-0.145**	-0.337***
	(0.130)	(0.071)	(0.072)	(0.041)
Private Individual	-0.275***		-0.178***	
	(0.092)		(0.057)	
Government	0.060		0.111*	
	(0.126)		(0.066)	
Parastatal	0.266*		0.324***	
	(0.137)		(0.099)	
Public Works Programme	0.116		0.209	
	(0.218)		(0.187)	
Church/Religious Organisation	0.015		0.148	
	(0.215)		(0.126)	
Other	0.062		0.366***	
	(0.269)		(0.131)	
Inverse mills ratio	-0.459***	-0.323***	-0.570***	-0.363***
	(0.146)	(0.082)	(0.071)	(0.059)
Year=2013	0.249***	0.371***	0.215***	0.331***
	(0.064)	(0.045)	(0.029)	(0.029)
Constant	9.472***	9.144***	9.799***	9.224***
	(0.330)	(0.256)	(0.174)	(0.163)
R-squared	0.251	0.127	0.311	0.142
Observations	1,909	3,466	1,909	3,466

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

We note a number of things that we did not observe without sample selection. Although the key results are preserved in terms of direction and significance after correcting for self-selection, the magnitudes of the coefficients significantly become smaller. This suggests that failure to account for sample selection upwardly biases the results reported in studies. For example, after controlling for sample selection, we find a smaller statistically significant coefficient for gender in the formal sector compared to the formal sector. In addition, while the coefficients for regions are insignificant in the formal sector, the differences are not only significant but also larger in the informal sector.

3.8.2 Household non-farm enterprise earnings

In this section, we explore two main ideas. Firstly, we would like to establish if there are positive 'externalities' in education in the running of household enterprises. Secondly, we establish if the observed externalities have a turning point beyond which they either increase or decrease.

Table 3.22 presents the results where we establish that strong education externalities exist within the households in non-farm enterprises. After controlling for other factors, we find returns to education of up to 12.8% and 15.8% for OLS and random effects, respectively. The control factors are experience, the gender of the owner, location of the enterprise, enterprise industry, the age of the enterprise, type of enterprise customers, enterprise registration status with an association and time dummy.

We add a square term of maximum education to establish if education externalities have a turning point beyond which they either increase or decrease. Some studies have established that education externalities have been found to be monotonic. For example, Mussa (2014) uses quartiles of average household education and finds that the marginal effects do not switch signs with respect to efficiency and production uncertainty of maize in Malawi. This finding is confirmed in this study; the coefficient of the square of maximum education is not significant, implying that education externalities are monotonic, i.e. they do not increase with levels of education.

Our study finds contrary evidence to that by Matita and Chirwa (2009), who find that enterprises owned by females tend to be less profitable by about 73% compared to those owned by males. We find no statistically significant earning differences between males and females. Enterprises in rural areas earn almost 60% lower than their counterparts operating in urban areas. This is probably due to the fact that the demand for goods and services is higher in urban areas than in rural areas. Matita and Chirwa (2009) find similar results.

With respect to industry, the following enterprises are more profitable when compared to agricultural enterprises: mining and quarrying (OLS only), construction and transport, storage and communication sector. Older enterprises tend to generate more profit probably because they are more likely to have a loyal customer base accumulated over time and have accumulated the necessary experience. As expected, enterprises in the informal sector are less profitable compared to those in the formal sector. Informal sector enterprises are typically micro and small enterprises with limited access to capital and entrepreneurial skills.

Table 3.22: Regressions for monthly household non-farm enterprise earnings

Explanatory variables	OLS				Random effects			
	1	2	3	4	5	6	7	8
Maximum years of schooling	0.144***	0.189***	0.128**	0.128**	0.143***	0.229***	0.163***	0.158***
	(0.018)	(0.059)	(0.053)	(0.053)	(0.012)	(0.046)	(0.045)	(0.045)
Maximum years of schooling squared		-0.002	-0.002	-0.002		-0.003*	-0.003*	-0.003
		(0.002)	(0.002)	(0.002)		(0.002)	(0.002)	(0.002)
Owner experience			0.008**	0.008**			0.008***	0.008**
			(0.003)	(0.003)			(0.003)	(0.003)
Owner experience squared			-0.000***	-0.000***			-0.000***	-0.000***
			(0.000)	(0.000)			(0.000)	(0.000)
Enterprise owner is female			0.055	0.053			-0.008	-0.016
			(0.069)	(0.068)			(0.064)	(0.063)
Rural			-0.567***	-0.567***			-0.526***	-0.531***
			(0.117)	(0.118)			(0.078)	(0.078)
Enterprise industry								
Mining and Quarrying			0.806*	0.804*			0.896	0.908
			(0.454)	(0.463)			(0.565)	(0.565)
Manufacturing			-0.299*	-0.279			0.052	0.078
			(0.171)	(0.179)			(0.284)	(0.285)
Construction			1.316***	1.338***			1.675***	1.699***
			(0.402)	(0.40)			(0.476)	(0.474)
Wholesale, Retail and Trade			-0.078	-0.061			0.269	0.291
			(0.184)	(0.193)			(0.284)	(0.286)
Transport, Storage and Communication			0.334	0.346			0.808**	0.821**
			(0.274)	(0.279)			(0.347)	(0.347)
Financing, Insurance and Business			0.32	0.374			0.753	0.793
			(0.917)	(0.912)			(0.825)	(0.817)
Community, Social and Personnel			0.029	0.049			0.173	0.201
			(0.236)	(0.242)			(0.316)	(0.317)
Informal sector			-1.182***	-1.183***			-1.097***	-1.096***
			(0.123)	(0.124)			(0.113)	(0.113)
Age of household enterprise			0.018***	0.017***			0.020***	0.019***
			(0.005)	(0.005)			(0.004)	(0.004)
Main enterprise customers								
Traders			0.228	0.230			0.301**	0.304**
			(0.172)	(0.173)			(0.152)	(0.153)
Other small businesses			0.645***	0.661***			0.840***	0.855***
			(0.216)	(0.218)			(0.237)	(0.238)
Large established businesses			1.486***	1.502***			0.905*	0.939*
			(0.355)	(0.354)			(0.547)	(0.539)
Marketing board (ADMARC)			4.467***	4.550***			4.031***	4.120***
			(0.240)	(0.263)			(0.152)	(0.160)
Other			0.102	0.138			0.197	0.231
			(0.177)	(0.177)			(0.210)	(0.206)
Not registered with enterprise association			-0.376*	-0.377*			-0.23	-0.234
			(0.202)	(0.203)			(0.215)	(0.215)
Year=2013				0.108				0.119**
				(0.079)				(0.060)
Constant	7.312***	7.059***	9.455***	9.382***	7.446***	6.943***	8.709***	8.664***
	(0.189)	(0.336)	(0.470)	(0.476)	(0.120)	(0.276)	(0.465)	(0.467)
R-squared	0.101	0.101	0.256	0.257	0.130	0.131	0.288	0.290
Observations	1,702	1,702	1,702	1,702	1,702	1,702	1,702	1,702

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; robust standard errors in parenthesis

The data also allows us to control for the type of enterprise customers. We use final consumers as the base category since we expect enterprises selling to these customers to be less profitable compared to institutional customers such as other small businesses, large established businesses and the national marketing board (ADMARC). This is confirmed in our study for both OLS and random effects models. Enterprises that are not registered with any enterprise association make less profit compared to those that belong to an association according to OLS but in the random effects model, the result is not significant.

3.9 Measurement error using panel data

In panel data, measurement error, just like non-random attrition bias, is of concern and an attempt is usually made in the literature to arrive at results that are robust to these concerns (Deaton, 1997). We discuss two types of measurement error as discussed in Wooldridge (2002), namely measurement in the dependent variable and measurement in the independent variable.

The assumptions we make about the measurement error are important. First is the usual assumption that the measurement error has zero mean. However, if this is not the case, then the estimation of the intercept is affected. The second assumption relates to the relationship between the measurement error and the explanatory variables included in the model. If the measurement error in the earnings is statistically independent or uncorrelated with each explanatory variable, then the OLS estimators from equations are consistent (and possibly unbiased as well). Consequently, measurement error does not bias the coefficients but only leads to larger standard errors than when the dependent variable is not measured with error, i.e. it leads to loss of efficiency.

In our model, one can reasonably argue that measurement error is correlated with education since people with more education tend to report their earnings more accurately. However, in the absence of additional information, it is difficult to establish if measurement error in earnings is related to any of the explanatory variables. One solution to measurement error in earnings is to collect more data because more observations imply a better estimator of variance and consequently reduces the errors in inferences. This solution is beyond the researcher's control considering that the data is secondary data[32].

Measurement error in the independent variables is considered a more serious problem than measurement error in the dependent variable. In panel data, the most common method of dealing with measurement error is first differencing (short and longer differencing). However, our data comes from a panel of two

[32] Alternatively, we can use the Malawi Labour Force (2013) cross-sectional data section for comparison purposes. Nevertheless, there is no guarantee that this is 'better' quality data.

periods such that longer differencing is not possible. Moreover, first differencing when T=2 yields the same results as fixed effects presented in Table 3.14. We are thus unable to correct for the possibility of such measurement error.

3.10 Comparing income and consumption

There is a good discussion in the literature as to which is a better measure between income and consumption in the money-metric approach to the measurement of living standards. In this section, we want to find out if our consumption and income data tell a consistent story as pointed out in Section 3.1.

In general, consumption is considered a better measure of long-term living standards than income because it is less volatile on an annual basis for most households. Unlike income, consumption is said to fluctuate much less dramatically. This argument particularly holds for low-income countries (including Malawi) where there is a large informal sector. Generally, it is difficult and less straightforward to estimate income from self-employment activities and casual employment in the informal sector (Haughton & Khandker, 2009). Furthermore, Malawi is largely agricultural based and incomes are largely vulnerable to seasonal and weather patterns. On the other hand, income is relatively easy to measure in many high-income countries where salaries and wages form the largest sources of income. Another advantage of using consumption expenditure over income is that while income usually fluctuates over an individual's lifetime, consumption remains relatively stable as a result of smoothing by saving and borrowing (Blundell & Preston, 1998); McKay, 2000; Duclos & Araar, 2006; Haughton & Khandker, 2009) as explained in the permanent income hypothesis.

In the IHPS survey, consumption seems to have been collected in a greater detail than income; only two modules of the household questionnaire are dedicated to collecting information on earnings (Modules E and N) compared to five modules aimed at collecting information about consumption (Modules G through K). In order to reduce recall associated with consumption and aspects of agricultural activities, it was planned that households be visited twice in 2013[33]. The visits were only approximately three months apart which we believe does not adversely affect recall bias.

The reference period for consumption depends on the items (e.g., the reference period is one week for food consumption and either one week, one month, three months and 12 months for non-food expenditures). However, the reference period for earnings is over the past twelve months. According to Deaton (1997), large increases in prices (inflation) will tend to overstate consumption relative to income due to the differences in the periods of reporting as noted in the questionnaire. Following this line of

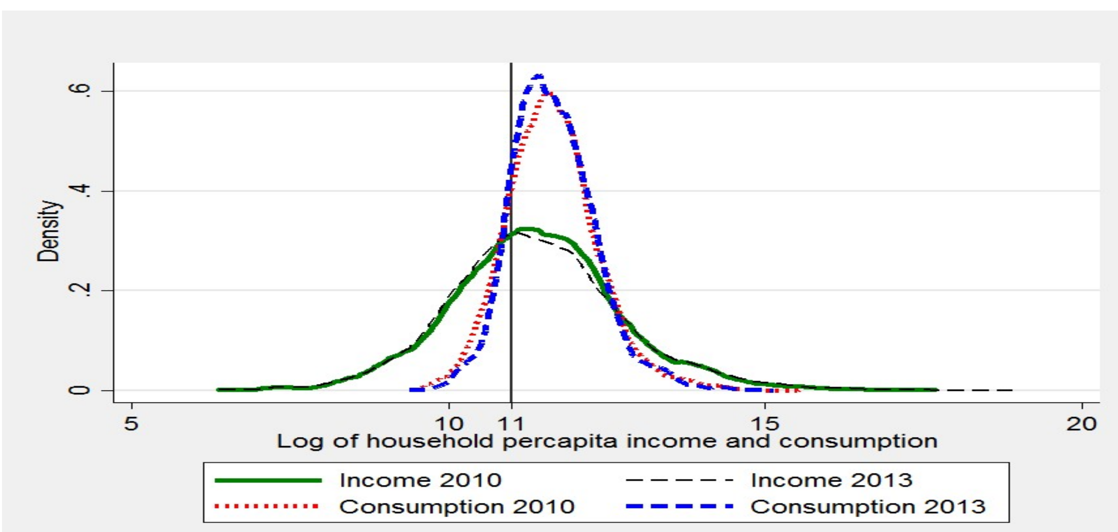
[33] About 92% of the households in the survey were visited twice in 2013.

argument, consumption is more likely to be reported in more recent and now higher prices than is income. However, inflation has not increased a lot over the period under study. Moreover, this inflation bias (if any) is partly reduced due to the fact that our consumption estimates are both spatially and temporally deflated.

We compare income and consumption using kernel densities as presented in Figure 3.3. Our analysis excludes households that split to form new households although the picture does not change even with the inclusion of the split-off households. Based on the figure, we can see that both consumption and earnings data seem to tell a consistent story. The data shows that, on average, both percapita income and percapita consumption have remained the same amongst households that were available in both waves. Nevertheless, we note that on the lower tail, consumption is generally larger than income meaning that some households to consume more than their income. This could be due to dissaving or receipts of grants and grants.

However, using household total income and consumption and, therefore, ignoring household size, the data shows that income and consumption have increased by 6.3% and 3.0%, respectively over three years. This shows that adjusting for household size is important especially when there are fewer earners in the household. Household percapita income (consumption) is calculated by diving total annual household income (consumption) by household size. We loosely define income as the sum of wages and earnings from self-employment activities. Household income is the sum of the incomes of all the earners in the household. The income was then converted into annual figures and adjusted into percapita figures by dividing by household size.

Figure 3.3: Kernel densities for annual household percapita consumption and income by year



Source: Own computation from IHPS data; the vertical line is the annual poverty line

3.11 Dynamics of household percapita consumption

Table 3.23 gives a breakdown of household percapita consumption by survey year. Despite consumption remaining unchanged, we note that some consumption components such as housing have declined by almost 4 percentage points. An analysis of housing expenditure by urban-rural areas shows that the largest decline came from urban areas (34%) compared to (12%) in rural areas. This disparity is perhaps due to the fact that most households in rural areas live in their own houses. Surprisingly, households spend equal proportions on alcohol and education.

Table 3.23: Average household percapita consumption by components and year

	2010		2013	
Description	Amount	Percent of total	Amount	Percent of total
Food	84,962	54%	88,939	56%
Alcohol	2,949	2%	4,508	3%
Clothing	4,340	3%	4,978	3%
Housing	34,001	22%	27,853	18%
Furnishings	5,493	3%	5,400	3%
Health	2,420	2%	1,958	1%
Transport	7,730	5%	9,865	6%
Communication	5,890	4%	3,972	3%
Recreation	1,484	1%	1,070	1%
Education	2,459	2%	2,707	2%
Hotels and restaurants	1,706	1%	2,478	2%
Miscellaneous	3,855	2%	4,180	3%
Total	157,287	100%	157,907	100%

Source: Own computation from IHPS data

In order to analyse the household consumption poverty dynamics, we categorise households into three groups with reference to the poverty line[34]. The first group identifies households below the food poverty line to capture the extent of extreme poverty (0- K53,262). The second group gives households with percapita consumption between food and absolute poverty line (K53,262-K85,852). The second group captures movements with poverty. Finally, the third group consists of non-poor households with consumption above absolute poverty (>K85,852).

[34] Households whose total percapita household consumption is below K53,262 (food poverty line)are considered ultra-poor while those with consumption below K85,852 are poor. The absolute poverty line is the sum of the food and non-food poverty lines.

The resulting consumption transition matrix presented in Table 3.24 shows that in 2010, the proportion below the food poverty line was about 12%. This proportion has declined to about 9% in 2013. Conversely, the proportion above the food poverty line has increased from about 88% in 2010 to about 91% in 2013. This change is reflected in two parts. First, there is a slight increase in the proportion above the absolute poverty from 66% in 2010 to about 67% in 2013. Second, there has also been movement with poverty as reflected in the increase in the proportion between the food and absolute poverty line from 22% in 2010 to 24% in 2013.

Despite the improvement in household welfare between 2010 and 2013, there are some considerable transitions, into, within and out of poverty. Out of the 12% that was below the food poverty line, about 4.7% moved upwards but still within absolute poverty while 4.4% moved above the absolute poverty line. About 2.7% households have been stuck in extreme poverty over the three years.

Table 3.24: Consumption transition matrix

		2013			
		Below food poverty line (K53,262)	Between food and absolute poverty line	Above absolute poverty line (K85,852)	Total
2010	Below food poverty line (K53,262)	2.7% (0.004)	4.7% (0.006)	4.4% (0.005)	11.8% (0.011)
	Between food and absolute poverty line	3.6% (0.004)	8.0% (0.008)	10.6% (0.007)	22.2% (0.011)
	Above absolute poverty line (K85,852)	3.0% (0.004)	11.3% (0.008)	51.8% (0.017)	66.0% (0.016)
	Total	9.3% (0.008)	24.0% (0.013)	66.8% (0.016)	100.0%

Source: Own computations from IHPS data; figures in parenthesis are standard errors

Amongst those within poverty (22.2%), about 8% remained within poverty while 10.6% moved out of poverty and a further 3.6% slipped below the food poverty line into extreme poverty. Of those that were above absolute poverty line (66.0%) in 2010, 3% ended up in extreme poverty, 11.3% became absolutely poor while 51.8% of the households remained out of poverty.

In the IHS panel survey, respondents were asked about the shocks that affected their households and how they responded to the shocks. According to the results, the major shocks were increases in the prices of food and agricultural inputs. In order to smooth their consumption, households embarked on a number of mitigation measures against the shocks. In 2013, nearly more than one-third (about 35%) of households reported having used their own savings to cope. This is an increase compared to only 21% of the households in 2010. There has also been an increase in the number of households using other mitigation measures between 2010 and 2013. These measures include seeking help from relatives and friends, help from the government, changes in dietary patterns, selling of assets, use of credit and spiritual help.

3.12 Conclusion and policy implications

The study sought to examine the returns and externalities to education in Malawi using the IHS3 panel data set. The random effects results based on the standard Mincerian earnings functions show that the average rate of return to years of schooling in Malawi is around 8.3% for the full sample. Consistent with the existing literature, we obtain returns to education that increase with the level of education and are also higher for females than males. Decomposition of the sample by economic sector reveals that the rate of return on education is lower in the informal sector at 8.6% in the formal sector and 4.7% in the informal sector. The rates of return found in this study compare favourably with those observed in other studies in Malawi and other Sub-Saharan African countries including Ghana, Cameroon and Rwanda. Sectoral analysis of earnings gives us some more interesting results different than when the sample is not split. First, we observe that there is a smaller statistically significantly difference in earnings between males and females in the formal sector than in the informal sector. Second, we find that returns to education are positive in both sectors but lower in the informal sector. These two findings highlight the importance of distinguishing between the different types of employment sectors in estimating rates of return on education in developing countries. The assumption that returns are the same in both sectors of the economy is not realistic and suggests that studies that fail to take this into account tend to overstate the returns to education. The coefficient comparison tests help us arrive at a better way of comparing estimates across models. The results are robust to different model specifications and sample selection. After accounting for sample selection, the returns to education drop by between 0.7 and 0.8 percentage points. Our results show that education externalities play a significant role in non-farm enterprises. Education in Malawi has wider social benefits which should not be underestimated. We also conduct data consistency checks to ensure data quality. This is important if we are to construct stable and meaningful comparable data for further analysis. Specifically, the chapter highlights issues related to data quality and inconsistencies in measuring returns to education. Unless the sources of inconsistencies are explained, it is difficult to conduct further meaningful analysis with the data.

The results have a number of implications with respect to policy. First, the positive and large returns from schooling suggest that education is a good investment. This is particularly supported by the presence of education externalities. Access to education, therefore, is important for all. Second, since returns increase with the level of education, it may be important to invest more resources into higher level education while not neglecting primary education. The current policy makes primary education universally free. This is good given the positive social returns of education established in the operation of non-farm enterprises. We, however, note with concern that the government has recently reduced its subsidy contribution being offered for tertiary education yet the findings from this study show that returns to education are highest at the tertiary level. Perhaps, the government should simply redirect these resources to other areas within tertiary education rather than reduce the contribution. Third, education policy should encourage female education considering that females with similar skills to males tend to have higher returns. This may be both economically efficient and equitable although currently there are fewer females entering the labour market in Malawi. Therefore, the focus of policy should not only be primary education but also tertiary education where the returns are the highest.

Chapter 4

Patterns of migration and employment in Malawi: Spatial data analysis

4.1 Introduction

There is growing interest in spatial analysis in the literature and it is now widely recognised that the standard non-spatial econometric techniques produce biased results in the presence of spatial autocorrelation and heterogeneity. Therefore, spatial analysis is at the research frontier and has recently become more central in both theoretical and applied econometrics including the fields of labour and agricultural economics (Anselin, 2003). This growing interest has allowed the development of econometric techniques that can handle spatial data. There has also been a rapid spread of geographic information systems (GIS) and the associated availability of data sets containing the location of observations.

Taking advantage of these developments as well specific literature gaps in Malawi, this study pursues three main aims: (i) to understand the spatial and temporal patterns of employment and migration in Malawi through the application of spatial data econometric techniques; (ii) to analyse how long-term changes in age structure affect labour force participation; and; (iii) to explore the effects of land reform policy on migration and employment through the use of a difference-in-difference estimation strategy.

The study contributes to the literature in three main ways. First, we add a gender dimension to our analysis and this has implications for development policy in a poor country such as Malawi. In the literature, it is argued that economic growth and structural change help remove barriers faced by women in development and, therefore, improve their labour force participation. However, little attention has been given to the empirical analysis of such assertions especially in the context of developing countries (Gaddis & Klasen, 2014). The study contributes to this strand of literature. The second contribution of this study relates to the application of spatial panel data econometric methods to the study of employment and migration in Malawi. This is important because the economic phenomena we are studying are essentially spatial in nature in the sense that we expect their values in a given location to be determined by the values in neighbouring locations. To our knowledge, such a study has not yet been done in Malawi. With respect to this objective, we compare results obtained with the application of spatial econometric models with estimation results obtained from non-spatial panel data models. Finally, by matching census data GIS codes that are consistent over time at the level of small geographical places, it is now possible to integrate census data with other data for similar spatial analysis. This is relevant because all nationally representative data sets collected by Malawi's National Statistical Office (NSO) use censuses as the sampling frame. For example, one can now merge census data with Demographic Health Survey (DHS) data aggregated at the district or small geographical area. This kind of data integration could be an area for further study.

4.2 Data and methods

4.2.1 Data

Using data from the Integrated Public Use Microdata series (IPUMS)-International, Minnesota Population Centre, we matched district and small geographical area level GIS data codes to create boundaries that are consistent over time, namely for 1987, 1998 and 2008. This matching gives us a panel of districts and small geographical areas (also called traditional authorities) over which we conduct our spatial analysis while also exploring the advantages of working with panel data. Based on previous work, most of the indicators are heterogeneous at small geographical areas rather than regions where there seems to be uniformity.

At this stage, it is important to provide a brief background on Malawi's administrative units. The country has three regions, namely Northern, Central and Southern regions. The regions are divided into districts which are further subdivided into small geographical areas called traditional areas. Figure 4.1 shows the administrative boundaries for Malawi.

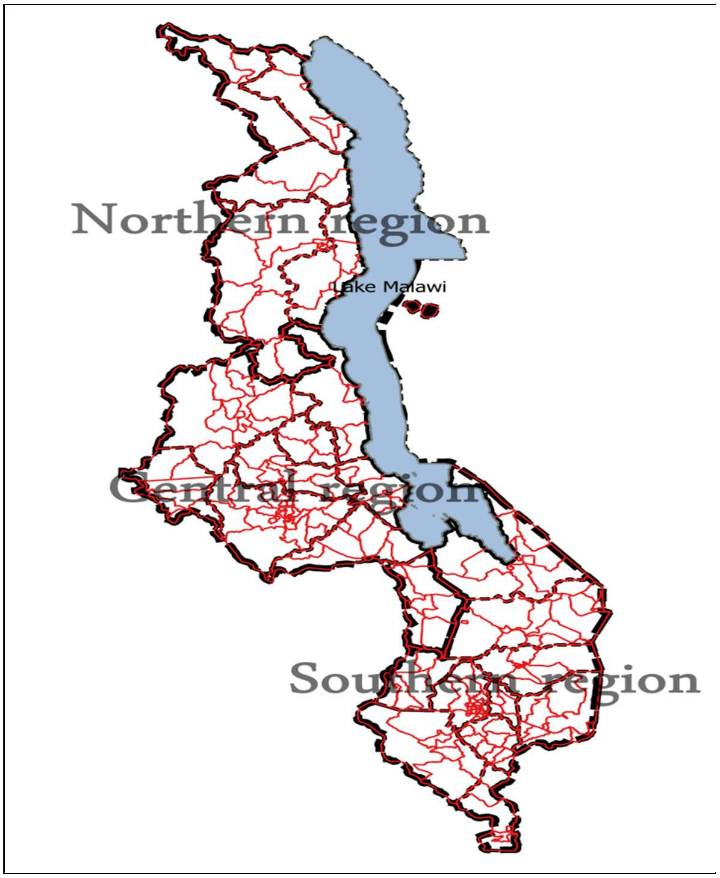
Malawi has four cities: Lilongwe, located in the Central region, Mzuzu in the Northern region, and Zomba and Blantyre in the Southern region. Since Lilongwe, Zomba and Blantyre are named after districts, the word "city" is added to the district name to signify the city. Consequently, "Lilongwe" refers to Lilongwe district while "Lilongwe City" to the city (National Statistical Office, 2008).

While the regions have remained stable over time, the number of districts and small geographical places in Malawi has been changing over time[35]. Through the changes, Machinga district was split into Machinga and Balaka, Mulanje was split into Mulanje and Phalombe, and Mwanza was split into Mwanza and Neno. Nkhata-Bay and Likoma are combined in the data (National Statistical Office, 2008). As expected, these changes create a problem of inconsistent comparability of the spatial units over time because we may not be working with the same observations as desired.

In order to allow for consistent comparison over time, we dissolved the new districts or traditional authorities and merged them into what they were as at 1987. The end result is data for a total of 24 districts and 178 traditional authorities instead of the 28 districts and 224 traditional authorities as at 2008. These form our units of analysis. Consequently, our variables of interest are collapsed by traditional authority as proportions by year and gender, depending on the level of analysis. A detailed list consisting of regions, districts and traditional areas is provided in Table A4 in the appendix.

[35] According to National Statistical Office (2008), there were 23 districts in 1966; 24 districts in 1977; 24 districts in 1987; 28 districts and 4 cities in 1998 and 27 districts and 4 cities in 2008.

Figure 4.1: Map of Malawi showing administrative boundaries



Source: Own computation from census data

4.2.2 Theoretical framework

Migration involves the movement of people from their usual places of residence to another area over a given period of time. Theories of migration try to explain why people migrate and where they come from and where they are going. All models of migration assume that migration is voluntary, although some forms of migrations are involuntary. The theoretical perspectives of migration from the field of economics and other disciplines can be broadly grouped into either of two disciplines, namely the disequilibrium and equilibrium perspectives (Greenwood, 2005). We now discuss these two alternative perspectives.

The disequilibrium framework has been the underlying model for understanding and the study of migration among economists dating as far as before the late 1970s. In this framework, migration was thought to be driven by differences in wages across regions or sectors. Consequently, people would migrate from one area to another in search of market opportunities in the form of wages and earnings. This framework has been criticised by the advocates of the equilibrium hypothesis who argue that the geographical differences in wages or incomes are compensating and do not reflect opportunities from which people can benefit. Instead, the equilibrium models argue that migration depends on imbalances in amenities. According to this theory, the patterns of migration are such that people move from amenity-poor areas into amenity-rich areas. However, as people migrate into amenity-rich areas, wages tend to decline and the prices of locally produced goods and services are driven up. Consequently, as wages and prices continue to diverge across regions until they compensate households for the differences in the amenity bundles supplied by the various regions. The common feature in both the disequilibrium and equilibrium models is that migration is motivated by regional imbalances, with the former dependent on wages while the latter on amenities. Furthermore, regardless of the model used, migration is essentially spatial (Cooke, 2013; Greenwood, 2005).

Within the equilibrium approach are the gravity and modified gravity models, which have foundations in the gravity law of spatial interaction. In the basic gravity model, migration is a function of the distance between any two places as well as the population sizes of the origin and destination. It has been argued that the basic gravity model does not fully capture all the factors behind migration. The modified gravity models, therefore, include several additional factors including income, unemployment rates, the degree of urbanisation, amenities and taxes, among other factors. Since the 1960s, the modified gravity models hold a prominent place in the literature and combine elements of both the disequilibrium and equilibrium hypotheses of migration (Greenwood, 2005).

Human capital theory is another theory used to explain migration. This model rests on the disequilibrium approach in which individuals are assumed to be economic agents seeking to maximise utility from leisure and income. Implicitly, this assumes that an individual's supply of labour is dependent on the wage rate. Individuals will, therefore, migrate only when the expected return exceeds the costs of migration among other costs. The human capital theory has been central to the understanding of migration and faced no competing theory for almost 20 years. Extensions of the human capital model are the ‘spatial job-search models’. Just like the human capital model, individuals migrate to other locations in search of opportunities but have reservation wages and would reject opportunities that give wages below the reservation wage. The reservation wage is itself dependent on several factors, including the rate of unemployment. Presumably, for unemployed people the reservation wage may fall over time (Willekens, 2008; Greenwood, 2005).

4.2.3 Spatial autocorrelation

The concept of spatial dependence[36] is based on the first law of geography by Waldo Tobler which states that “Everything is related to everything else, but near things are more related than distant things"

[36] We use spatial dependence and spatial autocorrelation interchangeably.

(Tobler, 1970, p. 236). Based on this law, two or more objects that are spatially close tend to be more similar to each other with respect to a given attribute Y than are spatially distant objects. In this case, we expect positive spatial autocorrelation among geographically close regions. Measures of spatial autocorrelation can be grouped into two broad categories, namely global and local indices.

Global indices express the overall level of similarity between spatially close locations in a given study area with respect to a variable Y. A global measure summarises the variable of interest in a single value and, therefore, only shows the average degree of the spatial distribution of the phenomena of interest. Local spatial autocorrelation measures overcome this limitation by detecting spatial clusters which provide greater detail than the global measures. Thus, local measures may help identify the locations that contribute most to the overall pattern of spatial clustering which some have referred to as hotspots (Sokal, Oden, and Thomson, 1998).

The most commonly used tests for spatial autocorrelation are the Moran's I and Geary's c statistics respectively named after their developers, Moran (1948) and Geary (1954). It is the practice in the literature to report both statistics as a means of robustness checks for each other in the detection of spatial dependence. A detailed discussion of these indices is provided in Pisati (2001) and Sokal et al. (1998). We begin our discussion with the global measures.

Moran's I is given by

$$\mathrm{I} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} Z_i Z_j}{S_0 m_2} \qquad (4.1)$$

Where: w_{ij} denotes the elements of the spatial weights matrix W corresponding to the location pair (i,j), $\mathrm{Z}_i = \mathrm{Y}_i - \bar{Y}$, Y_i represents the value taken by a variable of interest Y at location i; $\bar{Y}$ denotes the mean variable of Y; $S_0 = \sum_i \sum_j w_{ij}$ and $m_2 = \sum_i Z_i^2/N$.

Under the null hypothesis of no global spatial autocorrelation, the expected value of I is given by

$$\mathrm{E}(I) = -1/(N-1) \qquad (4.2)$$

On the one hand, if I is larger than its expected value, then the overall distribution of variable Y can be seen as characterised by positive spatial autocorrelation, meaning that the value taken on by Y at each location i tends to be similar to the values taken on by Y at spatially contiguous locations. On the other hand, if I is smaller than its expected value, then the overall distribution of variable Y can be seen as characterised by negative spatial autocorrelation, meaning that the value taken on by Y at each location i tends to be different from the values taken on by Y at spatially contiguous locations.

Inference is based on z-values, computed by subtracting $E(I)$ from I and dividing the result by the standard deviation of I as follows:

$$z_I = \frac{I - E(I)}{sd(I)} \tag{4.3}$$

Geary's c is calculated as

$$c = (N-1)\frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\left(Z_i - Z_j\right)^2}{2NS_0 m_2} \tag{4.4}$$

Under the null hypothesis of no global spatial autocorrelation, the expected value of c equals 1. If c is larger than 1, then the overall distribution of variable Y can be seen as characterized by negative spatial autocorrelation; on the other hand, if c is smaller than 1, then the overall distribution of variable Y can be seen as characterized by positive spatial autocorrelation. As in the case of Moran's I, inference is based on z-values, computed by subtracting 1 from c and dividing the result by the standard deviation of c:

$$z_c = \frac{c-1}{sd(c)} \tag{4.5}$$

Both z_I and z_c are assumed to follow a normal distribution (asymptotically) such that their significance can be evaluated through the use of a standard normal table (Pisati, 2001).

So far we have looked at global measures. We denote the local measures of autocorrelation by subscript i to denote location. Therefore, Moran's I_i and Geary's c_i detect positive and negative spatial autocorrelation around a given location. Positive I_i z-values and negative c_i z-values indicate clustering of similar values of Y around location i, that is, positive local spatial autocorrelation, while negative I_i z-values and positive c_i z-values indicate clustering of dissimilar values of Y around location i, that is negative local spatial autocorrelation.

4.2.4 Spatial panel data models

Spatial regression models estimate the relationship between a dependent variable y and one or more explanatory variables X while taking into account the spatial dependence among observations. In the literature, it is argued that when the observations are spatial units or locations, the standard regression models are usually misspecified because of the presence of spatial dependence (Pisati, 2001).

Several Stata tools exist for fitting spatial regression models and their application depends on whether one is working with cross-sectional[37] or panel data. In this study, we fit spatial panel data regressions through the use of xsmle by Belotti, Hughes, and Mortari (2013). This approach fits the models through the maximum likelihood (ML) estimation, which is the widely used method for spatial models. Anselin (2003) provides a good discussion on the different approaches for estimating spatial regression models including ML estimation, the method of moments estimators and spatial two-stage least squares.

Our discussion begins by considering the following general specification of spatial panel models from which specific types of spatial panel models are derived:

$$y_{it} = \tau y_{it-} \ + \rho W y_{it} + \beta X_{it} + \theta D Z_{it} + a_{i+} \gamma_{t+} \ v_{it} \ ; v_{it} = \lambda E v_{it} + \mu_{it} \qquad (4.6)$$

Where: y denotes an $N \ x \ 1$ vector of observations on the outcome variable; W denotes an $N \ x \ N$ row-standardised and inverse distance[38] spatial weights matrix; N denotes the number of locations; X denotes and $N \ x \ j$ matrix of observations on the explanatory variables; Wy denotes the spatially lagged dependent variable (DLAG) which assumes that the value taken by y in each geographical area is affected by the values taken by y in the neighbouring areas.

The term μ_{it} denotes an $N \ x \ 1$ vector of normally distributed homoscedastic errors uncorrelated to each other; D the spatial matrix for the spatially lagged independent variables; Z the spatially lagged regressors; and E the spatial matrix for the idiosyncratic error. The component a_i is the individual fixed or random effects and γ_t is the time effect. The terms ρ and λ denote the spatial autoregressive parameters. From the general model specification given in (4.6), we derive and discuss five spatial panel models which are normally considered in the literature.

First is the Spatial Auto-regressive model (SAR) with lagged dependent variable ($\theta = \lambda = 0$) and it is given by:

$$y_{it} = \tau y_{it-1} + \rho W \ \ _{it} + \beta X_{it} + a_{i+} \gamma_{t+} \ \mu_{it} \qquad (4.7)$$

where the standard SAR model is obtained by setting $\tau = 0$.

---

[37] For example, see Stata's *spatreg* by Pisati (2001) and *spmlreg* by Jenty (2010) for a discussion on spatial cross-sectional analysis.

[38] Row standardised means that all nonzero weights are rescaled so that they sum up to 1 within each row and inverse distance means that locations that are closer are given higher weights than those farther apart.

Second is the Spatial Durbin model (SDM) with time-lagged dependent variable ($\lambda = 0$) and it is specified as:

$$y_{it} = \tau y_{it-1} + \rho W y_{it} + \beta X_{it} + \theta D Z_{it} + a_{i+} \gamma_{t+} \mu_{it} \tag{4.8}$$

where the standard SDM model is obtained by setting $\tau = 0$. The package xsmle allows the use of a different weighting matrix for the spatially lagged dependent variable W and the spatially lagged regressors D together with a different set of explanatory and spatially lagged regressors Z_{it}. The default is to use $W = D$ and $X_{it} = Z_{it}$.

Third is the Spatial Autocorrelation Model (SAC) ($\theta = \tau = 0$) and it allows the use of a different weighting matrix for the spatially lagged dependent variable W and the error term E. It is given as:

$$y_{it} = \rho W y_{it} + \beta X_{it} + a_{i+} \gamma_{t+} v_{it} \text{ with } v_{it} = \lambda E v_{it} + \mu_{it} \tag{4.8}$$

The fourth model is the Spatial Error Model (SEM) ($\rho = \theta = \lambda = 0$) and it is given as:

$$y_{it} = \beta X_{it} + a_{i+} \gamma_{t+} v_{it} \text{ with } v_{it} = \lambda E v_{it} + \mu_{it} \tag{4.9}$$

Finally, we have the Generalised Spatial Random Effects (GSPRE) model ($\rho = \theta = \lambda = 0$) whose specification is as follows:

$$y_{it} = \beta X_{it} + a_{i+} \gamma_{t+} v_{it} \text{ with } a_i = \emptyset W a_i + \mu_i \text{ and } \mathrm{v}_{\mathrm{it}} = \lambda E v_{it} + \mu_{it} \tag{4.10}$$

Table 4.1 provides the summary of the spatial available fixed and random effects spatial models which are the most commonly estimators with panel data[39]. The SAR and SDM allow for a time-lagged dependent variable which may not be feasible for a panel with $T = 2$.

[39] Williams (2015), Clark and Linzer (2015); Wooldridge (2002) provide a detailed discussion on panel data methods and the choice between random and fixed effects.

Table 4.1: Summary of spatial panel model options

Description	Spatial Autoregressive Model (SAR)	Spatial Durbin Model (SDM)	Spatial Error Model (SEM)	Spatial Autocorrelation (SAC)	Generalised Spatial Panel Random Effects (GSPRE)
Random effects (RE)					
No time lagged dependent variable	Yes	Yes	Yes	N/A	Yes
With time lagged dependent variable	Yes, DLAG	Yes, DLAG	N/A	N/A	N/A
Fixed effects (FE)					
No time lagged dependent variable	Yes	Yes	Yes	Yes	N/A
With time lagged dependent variable	Yes, DLAG	Yes, DLAG	N/A	N/A	N/A

Source: Compiled from Belotti et al. (2013

We use fixed effects models because we are able to control for the effects of time-invariant variables with time-invariant effects. Our main results are based on SAR because it combines coefficients for both the weighting matrix and the matrix for the idiosyncratic error term, namely ρ and λ. SEM and SAC are used as robustness checks. Since SDM uses a time-lagged dependent variable, it is not possible to use for a panel with two periods of data. The non-spatial results are obtained with the help of the standard pooled ordinary least squares (OLS) and fixed effects estimation models.

4.3 Descriptive analysis

In this section, we show the spatial distribution of some of the variables of interest on the map of Malawi. This presents a first step in the spatial analysis of the data. Trends in fertility, population age structure, labour force participation, occupational mobility and migration are also discussed.

4.3.1 Spatial distribution of employment, education, assets and fertility

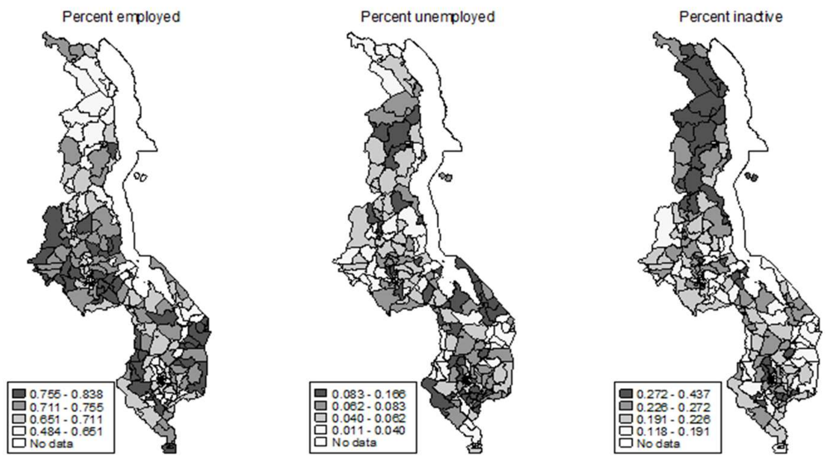
From Figure 4.2 to 4.5 we show choropleth maps for some of our indicators of interest drawn using two Stata tools, namely $shp2dta$ and $spmap$[40]. These maps reveal some spatial interaction and heterogeneity in Malawi in terms of proportions of employment status, type of employment, education, asset index, years of schooling and fertility rates. The long white strip on the map with no (or missing) data is Lake Malawi. The two dots near the Lake are Likoma and Chizumula Islands.

From the maps, we get an overall idea of the spatial distribution of the aggregated economic indicators under study. For example, most of the clustering in terms of the inactive population occurs in the

[40] $shp2dta$ converts shape boundary files to Stata datasets and $spmap$ is for visualisation of spatial data.

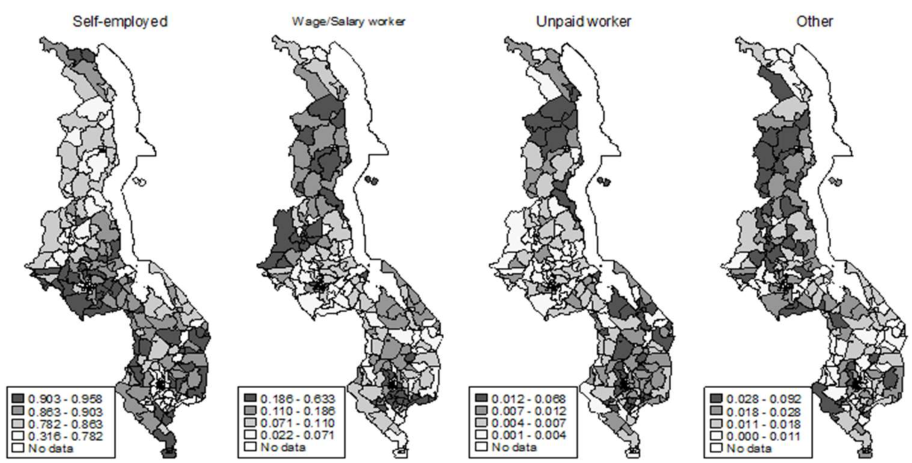
Northern Region. The two main reasons for being inactive cited in the data are schooling (52%) and involvement in housework (33%). We also observe some clusters of self-employment activity in some parts of the Central and Southern regions of the country. Education levels are lowest in the Southern region where there are also high levels of fertility and low asset ownership. The statistics generally agree with what is known about Malawi in terms of poverty from Chapter 2.

Figure 4.2: Spatial distribution of status in employment for 2008



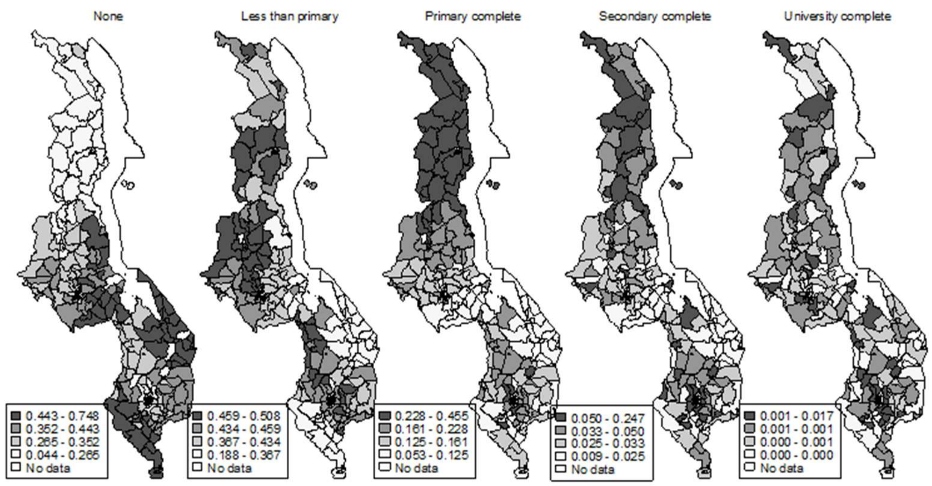
Source: Own computation from census data

Figure 4.3: Spatial distribution of employment type for 2008



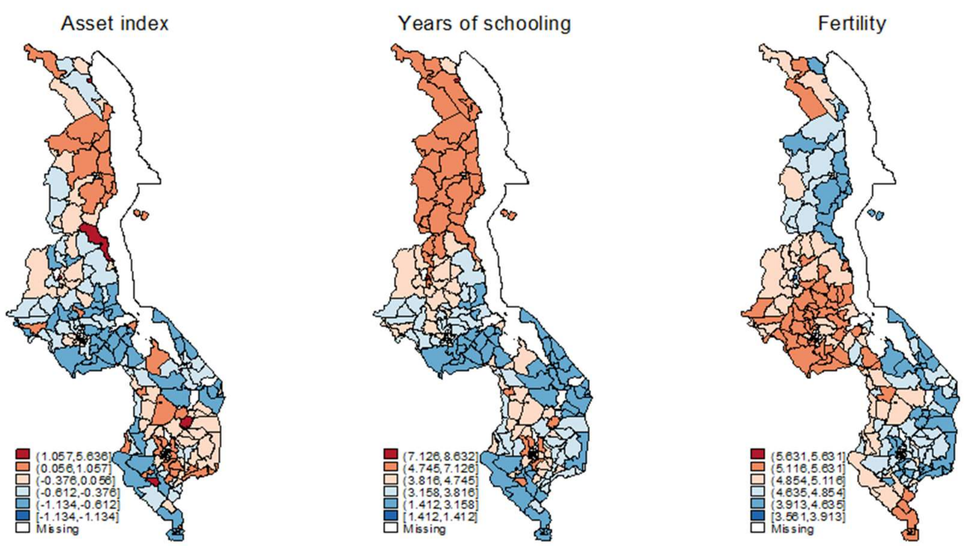
Source: Own computation from census data

Figure 4.4: Spatial distribution of educational attainment for 2008



Source: Own computation from census data

Figure 4.5: Spatial distribution of asset index, schooling and fertility for 2008



Source: Own computation from census data

4.3.2 Fertility trends

In Table 4.2, we present patterns of age-specific fertility rates[41] for three years of census data, namely 1987, 1998 and 2008. The unadjusted fertility first declined from 5.53 children per woman in 1987 to 4.75 in 1998 before slightly increasing to 5.05 in 2008. Overall, there has been a decline in total fertility rate over the two decades under study. With respect to the child-bearing ages, the patterns in the age-

[41] The age-specific fertility rate is calculated as the number of births in the last twelve months divided by the number of women for each age group. The result is then multiplied by 5 (the number in each of the age group).

specific fertility rates have not been uniform. They consistently remained high in the 20-24, 25-29 and 30-34 age groups but particularly increased for women in the 20-24 age group.

Table 4.2: Age-specific fertility rates (ASFR) and total fertility rates (TRF)

Year	Age group	No. of women	No. of births in last 12 months	Age-specific fertility rate
1987				
	15 – 19	37,732	4,477	0.59
	20 – 24	34,103	8,174	1.20
	25 – 29	28,211	6,497	1.15
	30 – 34	21,072	4,230	1.00
	35 – 39	20,147	3,314	0.82
	40 – 44	14,128	1,375	0.49
	45 – 49	12,538	677	0.27
	Total	167,931	28,744	5.53
1998				
	15 – 19	55,640	5,537	0.50
	20 – 24	54,149	12,029	1.11
	25 – 29	39,260	7,981	1.02
	30 – 34	29,465	5,049	0.86
	35 – 39	24,550	3,251	0.66
	40 – 44	17,782	1,415	0.40
	45 – 49	16,637	699	0.21
	Total	237,483	35,961	4.75
2008				
	15 – 19	68,134	7,558	0.55
	20 – 24	69,177	16,863	1.22
	25 – 29	57,291	12,797	1.12
	30 – 34	40,575	7,614	0.94
	35 – 39	29,428	4,252	0.72
	40 – 44	21,712	1,495	0.34
	45 – 49	17,057	513	0.15
	Total	303,374	51,092	5.05

Source: Own computation from census data

Studies have linked patterns of female employment to fertility rates. For example, the literature on feminisation U-hypothesis[42] suggests that the female labour force participation rate exhibits a U-shaped trend with respect to fertility rates as countries develop. At the early stages of economic development, when gross domestic product (GDP) is also still very low, most women tend to participate in the labour force even when fertility rates are still high. The women are said to be able to combine economic activity with child-bearing (Gaddis & Klasen, 2014). In the context of Malawi, much of the work is on household

[42] The feminisation hypothesis states that economic growth first lowers the participation of women in the labour market and increases it at later stages of economic development.

enterprises and small-scale household farms. However as the countries become richer, the shift in the structure of the economy towards manufacturing combined with the emergence of a formal sector results in lower levels of female labour force participation and employment while fertility is high.

Despite the improvement in economic activity and favourable structural change, the low levels of female education and the difficulty of combining wage employment with child-bearing contribute to the low participation of females in the emerging opportunities in industry and formal sector expansion. This may as well be influenced by social restrictions against females working in industries as well as allowing married women with children to get employment outside of the home.

The subsequent increase in female participation rate could be as a result of a combination of several factors. One possible explanation could be the expansion of education among females. However, it has generally been the case that labour force participation is also quite high amongst those without education. Therefore, improvement in education does not explain everything. Another explanation could be the emergence of attractive employment opportunities for women in the white-collar service industries which give women a chance to enter the labour force without many restrictions. In this case, a decline in fertility rates possibly frees up the time women spend on child-bearing which can be in the end reallocated to other work. Improvements and increased access to child-care facilities may also have allowed women to combine work outside the home with raising children.

4.3.3 Changes in population age structure and labour supply

Data for this section are drawn from the October 2011 International Labour Organisation (ILO) Economically Active Population, Estimates and Projections (EAPEP) (6th ed., October 2011). ILO's EAPEP database contains country estimates and projections of the total population, the activity rates and the economically active population by sex and age groups. In Table 4.3, we use the EAPEP data to show changes in population and labour force between 1987 and 2008.

Table 4.3: **Changes in economically active population and labour force by sex (1987-2008)**

	Population				Labour force			
Age group	1987	1998	2008	Change (1987-2008)	1987	1998	2008	Change (1987-2008)
Male								
15-19	407,691	558,786	782,638	92%	139,838	165,401	378,598	171%
20-24	343,770	493,273	629,291	83%	249,233	350,717	369,706	48%
25-29	284,125	372,529	533,372	88%	262,531	347,942	515,089	96%
30-34	223,757	298,164	449,713	101%	215,254	290,114	438,683	104%
35-39	192,948	246,465	314,596	63%	187,160	241,536	311,360	66%
40-44	157,269	195,192	236,352	50%	152,708	191,483	233,980	53%
45-49	120,458	167,955	192,212	60%	116,965	165,100	190,379	63%
50-54	97,986	147,204	155,313	59%	94,850	144,407	153,884	62%
55-59	77,201	110,054	137,155	78%	74,808	107,743	135,824	82%
60-64	57,122	87,802	119,605	109%	54,387	85,344	118,235	117%
65+	92,185	153,561	199,020	116%	81,330	144,040	193,049	137%
Total	2,054,512	2,830,985	3,749,267	82%	1,629,065	2,233,826	3,038,788	87%
Female								
15-19	420,190	557,397	775,215	84%	208,414	242,468	485,026	133%
20-24	363,609	495,638	626,106	72%	267,253	367,763	439,787	65%
25-29	306,338	386,990	522,124	70%	241,088	315,784	510,079	112%
30-34	247,318	318,279	426,374	72%	203,295	268,627	419,522	106%
35-39	219,557	267,212	299,434	36%	185,745	231,138	295,232	59%
40-44	178,186	214,836	241,890	36%	155,022	189,915	239,644	55%
45-49	135,505	181,539	216,010	59%	120,193	163,204	214,815	79%
50-54	110,125	155,631	185,938	69%	98,562	140,846	178,502	81%
55-59	88,609	113,442	162,105	83%	79,659	103,119	155,785	96%
60-64	68,982	90,758	137,274	99%	59,351	81,682	131,613	122%
65+	116,213	173,485	229,905	98%	87,863	145,727	202,316	130%
Total	2,254,632	2,955,207	3,822,375	70%	1,706,446	2,250,274	3,272,321	92%

Source: Own computation from EAPEP data

Firstly, we observe that the labour force growth is higher than that of the population. The magnitude is larger for females than for males. For males (females), the labour force has grown by 87% (92%) compared to 82% (70%) for the population. Secondly, the percentage of the youth has been on the increase. Using the International Labour Organisation (ILO) definition, the youth (15-24) made up 37.7% of the total working-age population in 2008 compared to 36.6% in 1987[43]. This explains some of the large increases that have taken place within the youth groups in terms of both the economically active population and labour force participation.

These observed changes in the age structure of the population have significant implications on the supply of labour and growth. As we see in Figure 4.6 the population pyramid for 2008 is different from the

[43]Estimates from the 2010 IHS3 put the figures at 38% and 66% for the ILO and SADC definitions, respectively.

other two in terms of labour force participation rates especially for the youth (ages 15-29) where the numbers have expanded, suggesting that a majority of the youths now participate in the labour market.

Figure 4.6 : Pyramids showing populations and labour force participation by sex (1987)

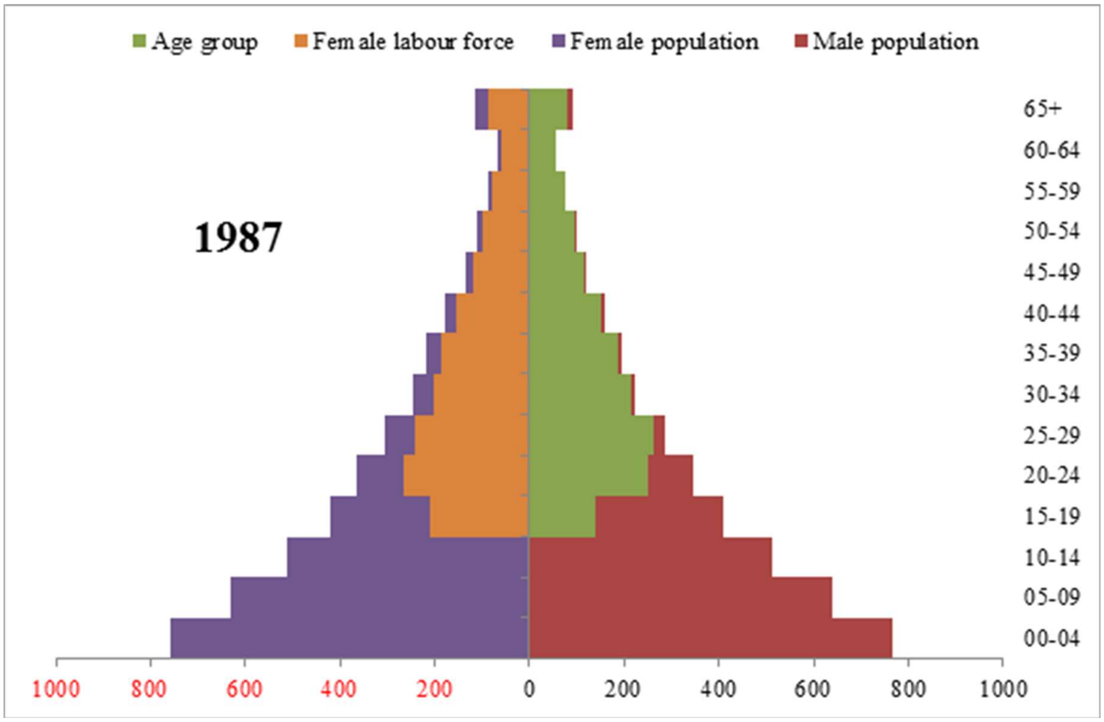


Figure 4.7: Pyramids showing populations and labour force participation by sex (1998)

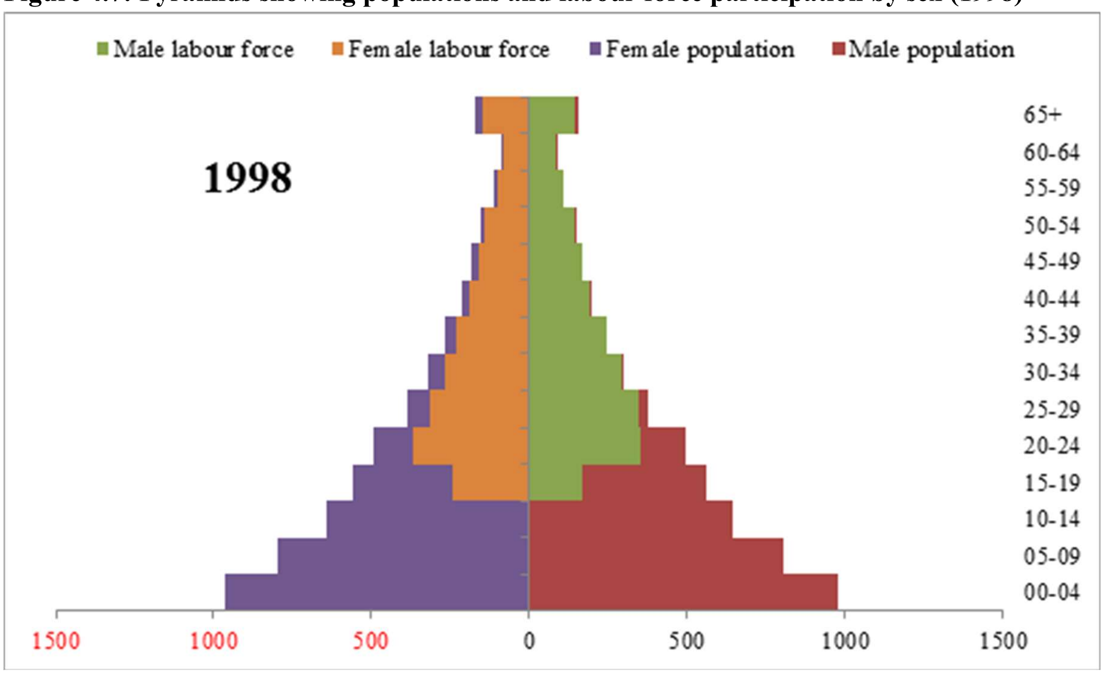
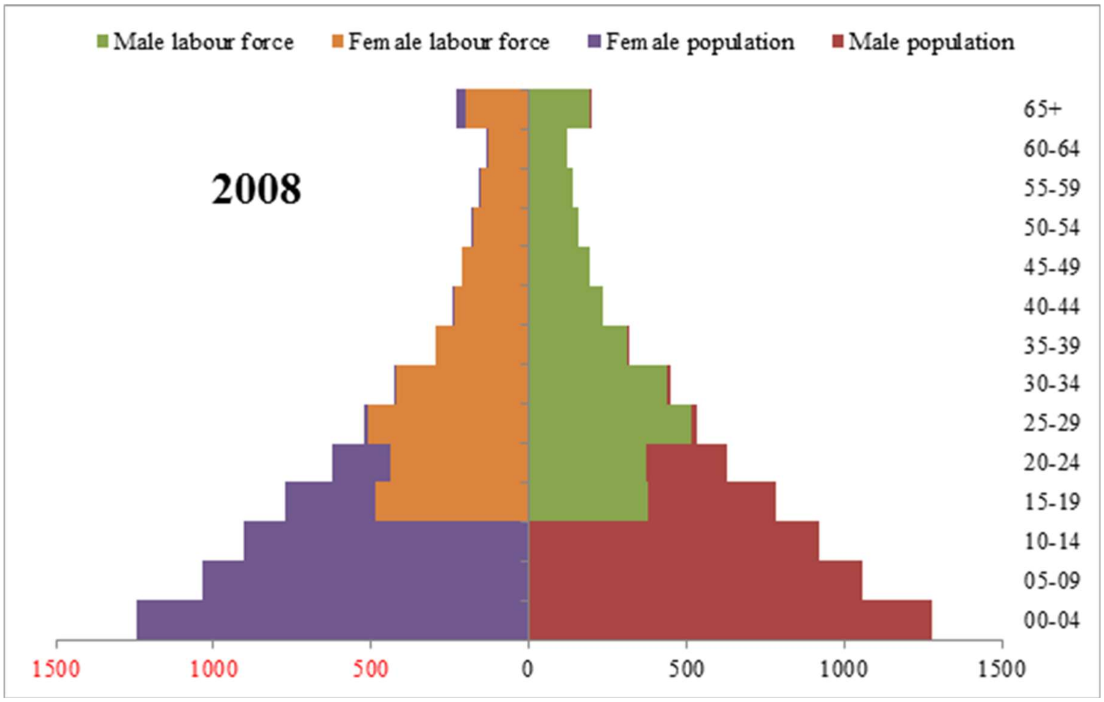


Figure 4.8: Pyramids showing populations and labour force participation by sex (2008)



Source: Own computation from census data

Changes in population depend on migration, fertility and mortality rates. The picture is, however, not complete because education levels are also increasing and this has implications on fertility and the time people spend in education (and therefore entry into the labour market). Fertility rates have not fallen significantly in Malawi between 1987 and 2008 as earlier shown. The mortality rate (measured by crude death rate) has dropped over time from 20.39 deaths per 1000 in 1987 to about 12.13 per 1000 in 2008.

It is not straightforward to know whether the combined changes of fertility, mortality and migration translate into significant differentials in the size, growth and structure of the labour force. In Table 4.4 we show the proportions of the population and labour force belonging to each of the age groups. With the exception of the (20-24) age group, the overall changes in the population proportions seem to mirror those in the proportions in the labour force.

Table 4.4: Proportions of population and labour force by age group

Age groups	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65+	Total
Population												
1987	19.2	16.4	13.7	10.9	9.6	7.8	5.9	4.8	3.8	2.9	4.8	100
1998	19.3	17.1	13.1	10.7	8.9	7.1	6.0	5.2	3.9	3.1	5.7	100
2008	20.6	16.6	13.9	11.6	8.1	6.3	5.4	4.5	4.0	3.4	5.7	100
Labour force												
1987	10.4	15.5	15.1	12.5	11.2	9.2	7.1	5.8	4.6	3.4	5.1	100
1998	9.1	16.0	14.8	12.5	10.5	8.5	7.3	6.4	4.7	3.7	6.5	100
2008	13.7	12.8	16.2	13.6	9.6	7.5	6.4	5.3	4.6	4.0	6.3	100

Source: Own computation from EAPEP data

We also examine trends in the age dependency ratio and the proportions of the total labour force by age groups. We define dependency ratio as the ratio of dependents (people younger than 15 or older than 64) to the working-age population (those aged 15-64). Based on this definition, the dependency ratio was 98% in 1987, declined to 94% in 1998 before increasing again to 96%.

4.3.4 Spatial and temporal patterns of migration

Migration data is only available for 1987 and 2008 and these can be consistently compared over time. The censuses collected migration data at the district level and geographical mobility was only defined where an individual's district of residence one year ago was different from the district of residence at the time of the census. Movements within a district or city are not considered as migration. However, for districts which share boundaries with cities, if a person moves from rural areas into a city (e.g. from Lilongwe rural to Lilongwe city) and vice versa is recorded as a movement.

Migration is thought to be responsible for changes in spatial distribution of populations and labour force in countries. For example, in Malawi, the urban population amongst individuals aged between 15 and 64 years has steadily increased from 11.7% in 1987 to 16.0% in 1998 to 16.6% in 2008. Although these are not very large changes, the trend is likely to increase and this may have implications for a wide range of social, economic and political dynamics in the country, including the general employment situation. The expansion of the labour force must be matched by the creation of new productive jobs into which migrants can be absorbed. While the Northern and Central regions have registered increases, the proportion of people living in the Southern region has dropped from 49.3% in 1987 to 44.5% in 2008.

Our analysis assumes, for the sake of simplicity, that Malawi is a closed system. We, therefore, only concentrate on internal migration. Moreover, the census data provide information only on immigrants into the country and not on Malawians now abroad but who lived in the country a year earlier. Two main patterns of migration can be identified as taking place in the country, namely regional and district level migration.

4.3.4.1 Regional movements

The first level of analysis involves movements across Malawi's three administrative regions. The transition frequencies are given in Table 4.5 and accompanying probabilities in Table 4.6.

As one might expect, the majority of the people remained in their home region with only a small percentage moving across. Approximately, 96.73% of the males and 97.71% of females resident in the Northern region in 1986 were still there in 1987. A slightly higher percentage than in the North remained in the Centre and South. The patterns largely remained the same in 2008 although slightly more people had migrated from their original region of residence. Much of the people are migrating towards the Centre and in the long-run, it might be expected to have the majority of the people. The reasons for migration are not captured in the census data. However, most migration seems to be undertaken as a way of improving incomes and economic status. Other reasons might include education, marriage and retirement.

Table 4.5: Inter-regional migration transition frequencies by gender

Year		Male				Female			
		North	Centre	South	Total	North	Centre	South	Total
1987	North	204,270	4,030	2,870	211,170	231,620	3,280	2,150	237,050
	Centre	4,990	720,780	12,790	738,560	3,470	778,960	9,390	791,820
	South	4,270	16,460	876,540	897,270	2,700	9,800	1,027,740	1,040,240
	Total	213,530	741,270	892,200	1,847,000	237,790	792,040	1,039,280	2,069,110
2008									
	North	365,230	22,780	15,510	403,520	404,750	22,810	14,470	442,030
	Centre	24,830	1,312,940	54,470	1,392,240	22,870	1,381,350	52,400	1,456,620
	South	17,580	93,200	1,389,170	1,499,950	13,130	70,840	1,537,130	1,621,100
	Total	407,640	1,428,920	1,459,150	3,295,710	440,750	1,475,000	1,604,000	3,519,750

Source: Own computation from census data

Table 4.6: Inter-regional migration transition probabilities by gender

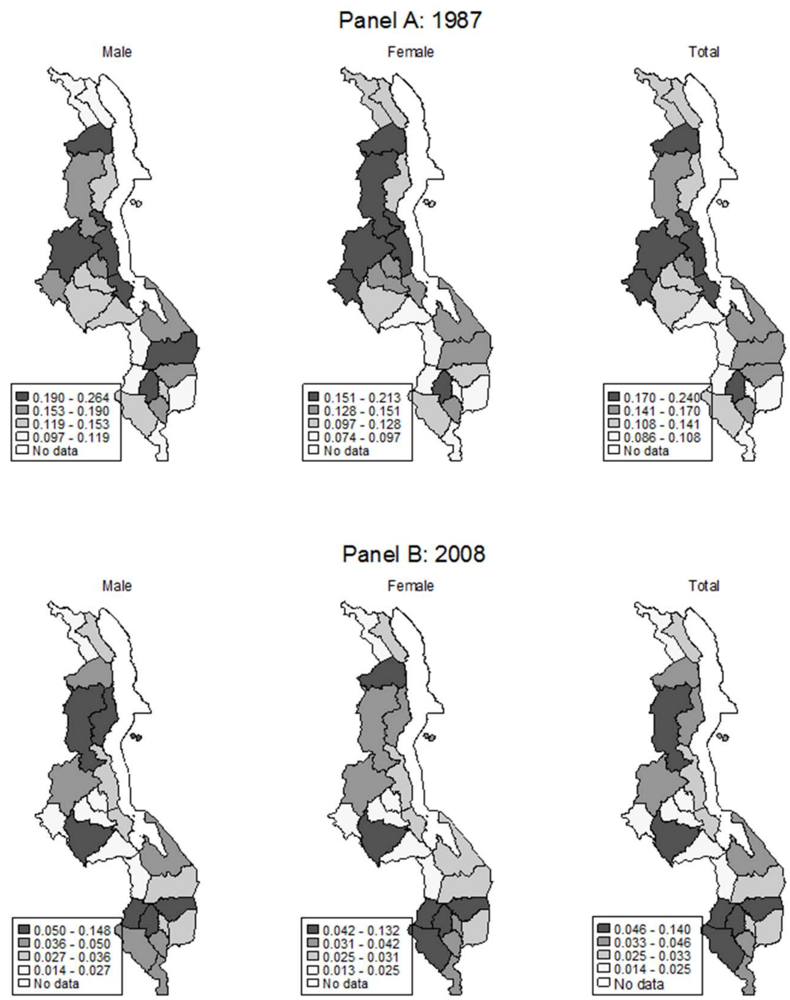
Year		Male				Female			
		North	Centre	South	Total	North	Centre	South	Total
1987	North	96.73	1.91	1.36	100	97.71	1.38	0.91	100
	Centre	0.68	97.59	1.73	100	0.44	98.38	1.19	100
	South	0.48	1.83	97.69	100	0.26	0.94	98.80	100
	Total	11.56	40.13	48.31	100	11.49	38.28	50.23	100
2008									
	North	90.51	5.65	3.84	100	91.57	5.16	3.27	100
	Centre	1.78	94.30	3.91	100	1.57	94.83	3.60	100
	South	1.17	6.21	92.61	100	0.81	4.37	94.82	100
	Total	12.37	43.36	44.27	100	12.52	41.91	45.57	100

Source: Own computation from census data

4.3.4.2 District level migration

As we will see, the district level migration patterns reveal another important aspect of migration in Malawi namely, rural-rural migration driven by people searching for agricultural land, considering the country's dominantly agricultural base and high fertility rates. Just like with regional movements, we capture district movements as a transitional probability matrix obtained by a cross tabulation of the current and previous district of residence. The population proportions that migrated in each district by gender and year are shown in Figure 4.9. We also provide a detailed analysis of population figures for each district by gender and year, the number of people that moved between districts and the accompanying proportions in the appendix in tables A6, A7 and A8.

Figure 4.9: Proportions that moved between districts



Source: Own computation from census data

Malawi has had policies that can potentially affect migration patterns such as those shown in the figure. We discuss two of these government policies. Firstly, the government has been supporting the establishment of rural growth centres which act as focal points for development within rural areas. This project falls under the Integrated Rural Development Programme outlined in the MGDS II. This might have the effect of keeping people in rural areas. Secondly, the Malawi Community Based Rural Land Development Project (CBRLDP), which was launched in 2004 and ran through 2009, enabled landless or land-poor households to voluntarily purchase plots from fallow estates and resettle there. The historical problem of land shortage is evident in Thyolo and Mulanje districts, which are traditionally home to tea estate farms since the early 1900s. The land reform policy was aimed at addressing emergent social conflicts related to unequal access to land. As a pilot programme, about 15,000 households from these rural districts were allowed to resettle in other rural districts of Machinga and Mangochi, but later it was expanded to Ntcheu and Balaka to ease pressure on land prices (World Bank, 2012). The policy has implications for migration patterns we observe post-2004 in the Southern region districts affected by CBRLDP and adjacent areas. The observed patterns of migration due to this policy are largely rural-rural migration since the programme targeted Malawi's rural poor. We do not have data on these specific 15,000 households but in the sections that follow we simply take advantage of the natural experiment by looking at conditions of the affected districts (as a whole) before and after the policy.

4.3.5 Spatial autocorrelation in variables

Table 4.7 extends the spatial comparisons in the maps to show the calculated Moran's I and Geary's c statistics for our variables of interest. For lack of space, we only report statistics based on the pooled sample, but we have also computed them on year by year basis where we find that they are generally stable over time as shown in the appendix in Tables A3, A4 and A5. The table shows that there is similarity between spatially close traditional authorities as shown by positive statistically significant Moran's I and Geary's c values. Technically, spatial dependencies justify spatial regression analysis.

Table 4.7: Results showing spatial dependencies in variables (pooled 1987, 1998 and 2008)

Variables	Moran's I	SD	p-value	Geary's c	SD	p-value
Migration	0.397	0.009	0.000***	0.612	0.011	0.000***
Employed	0.063	0.009	0.000***	0.930	0.013	0.000***
Years of schooling	0.167	0.009	0.000***	0.793	0.013	0.000***
Sex	0.153	0.009	0.000***	0.857	0.013	0.000***
Age	0.056	0.009	0.000***	0.954	0.015	0.002***
Married	0.127	0.009	0.000***	0.863	0.012	0.000***
Dependency ratio	0.075	0.009	0.000***	0.927	0.013	0.000***

Note: *, **, *** denote significance at 10%, 5% and 1% levels

We compute and plot Moran's *I* spatial correlograms for employment and migration based on cumulative distance bands, namely (0; 100,000 metres], (0; 200,000 metres],…, (0; 700,000 metres] presented in Figure 4.10 and Figure 4.11, respectively.

Figure 4.10: Spatial autocorrelation for employment

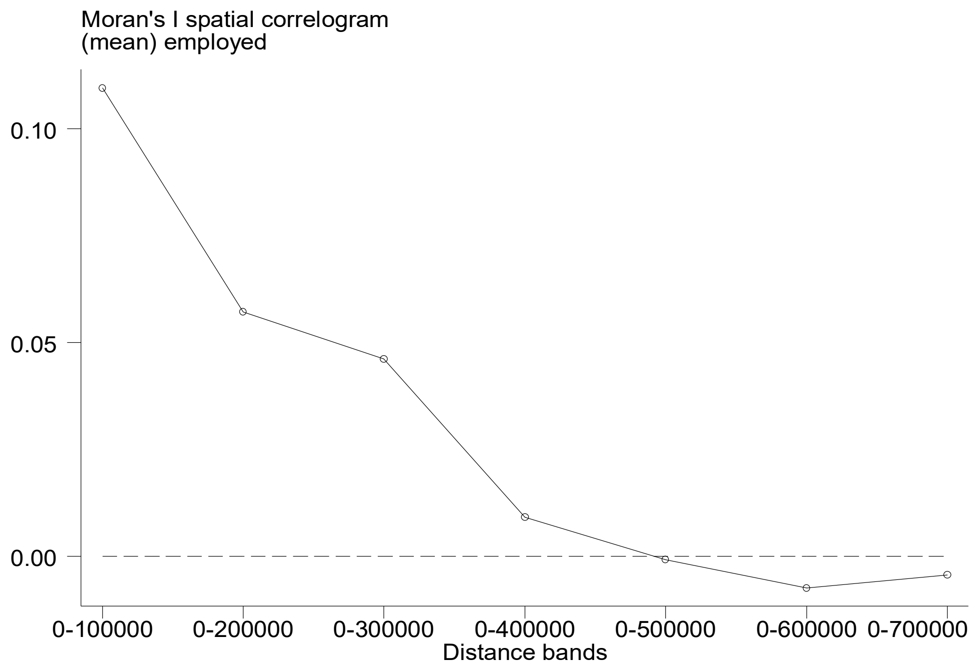
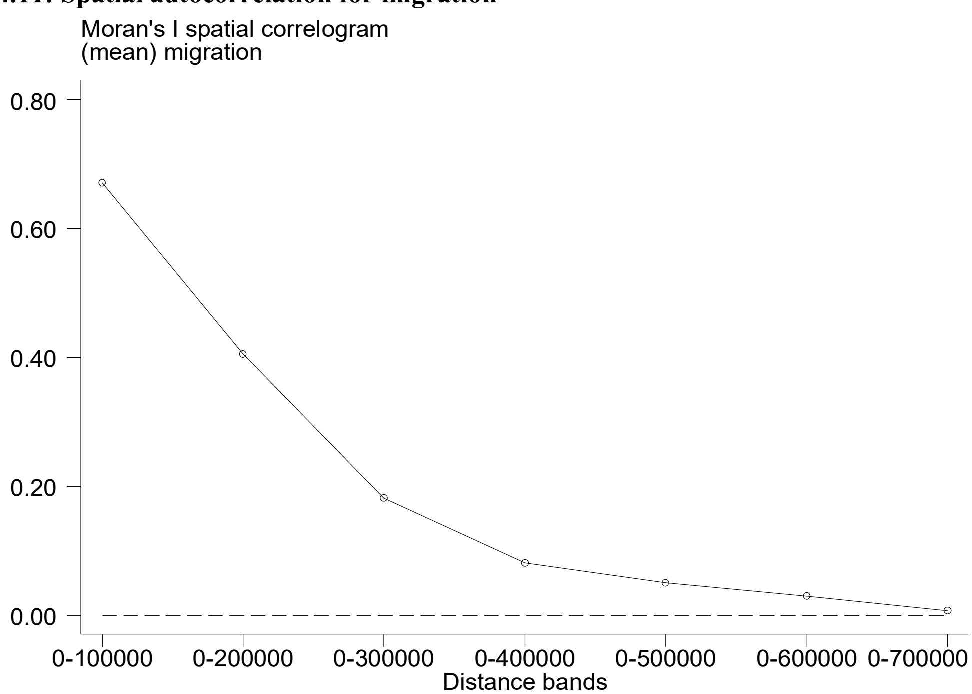


Figure 4.11: Spatial autocorrelation for migration

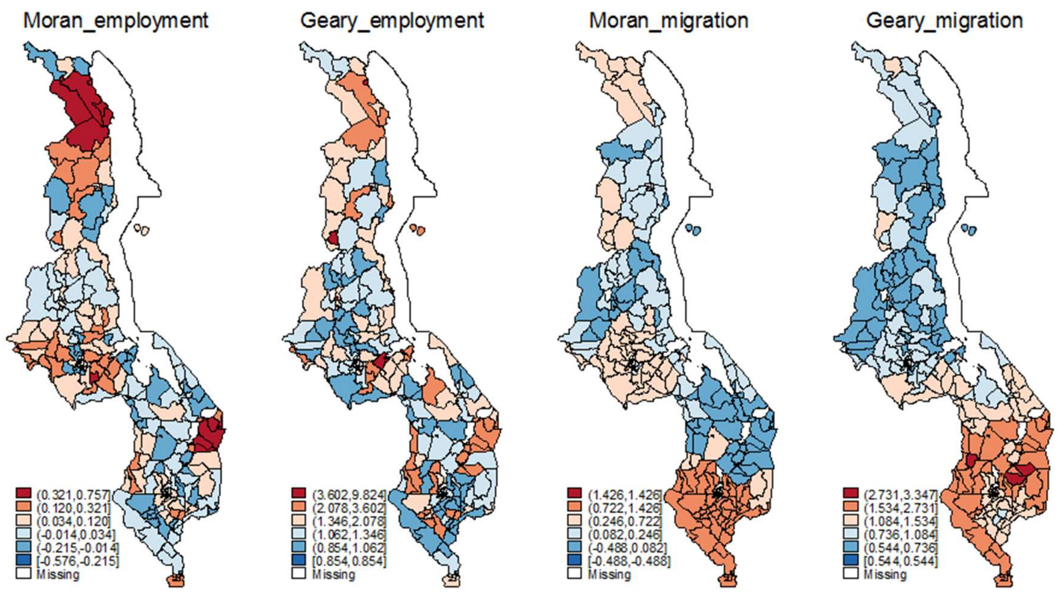


Source: Own computation from census data

The correlograms help us to draw two conclusions. Firstly, spatial dependence is larger for shorter distance bands but diminishes with larger cumulative bands and even becomes negative for employment at large distances. Secondly, spatial dependence is higher for migration than for employment. Employment has lower dependence partly because of the fact that most of Malawi is agriculturally based and there seems to be little variation.

Figure 4.12 shows pairs of choropleth maps of two measures of local spatial autocorrelation (Moran's *I* and Geary's *c*) for employment and migration in Malawi. For each small geographical area, the maps confirm some degree of similarity between neighbouring areas with respect to both employment and migration. Generally, the two measures yield consistent results in spatial patterns.

Figure 4.12: Measures of local spatial autocorrelation for employment and migration



Source: Own computation from census data

4.4 Spatial panel regression results

The spatial descriptive results and spatial autocorrelation statistics presented in Section 4.3.1 indicate that there is some clustering in our variables of interest at different locations. However, according to Anselin (1992b), this does not explain why the clustering occurs. We, therefore, extend our previous analysis through spatial panel regression analysis.

Our analysis is segregated by gender and the 178 small geographical areas form our spatial units. We begin our analysis with the standard non-spatial ordinary least squares and fixed effects regressions. However, as discussed in Section 4.2.3, these are inadequate in providing answers to our question due to the problem of misspecification when spatial locations form our unit of analysis and in the presence of spatial dependence (Pisati, 2001). The rest of the models are spatial fixed effects models whose results we compare with OLS and non-spatial fixed effects (FE).

In this section, we specifically explore the effects of land reform policy, explained earlier, on migration and employment in Malawi through a difference-in-difference (DID) estimation strategy. Considering that migration and employment are inherently spatial, we expect the programme to have spill-over effects affecting surrounding districts.

We now provide a brief discussion on the set-up of the DID. Firstly, we created a dummy variable called time to indicate the time when the treatment started. The land reform started in 2004 and we assign a value of "0" to the period before 2004 and "1" post-2004. Secondly, we create a dummy variable $\mathrm{treatment}$ to identify the group exposed to the treatment. Small geographical areas in Thyolo and Mulanje are assigned a value of "1" and "0" is assigned for the rest of the country. Finally, we create an interaction term between time and treatment as follows ($\mathrm{DID} = \mathrm{time} * \mathrm{treatment}$). The coefficient for DID is the difference-in-difference estimator in which we are interested.

On the basis of the foregoing discussion, the regression implementation of DID for estimating the impact of treatment on the outcome variable (y) is as follows:

$$\mathrm{y}_{it} = \mathrm{time} * treatment_i + \beta treatment_i + \delta time_i + \varepsilon_{it} \tag{4.13}$$

It is possible to combine fixed effects with propensity score matching (PSM) on the baseline data to ensure that the comparison and treatment groups are similar before applying double differences to the matched sample. Alternatively, balancing tests can be used to check whether the mean of the observables are the same in the base period. The tests are necessary for obtaining valid results because the treatment and comparison groups must be balanced using observed characteristics.

4.4.1 Effects of land reform policy on migration

Our econometric results are presented in Table 4.8 for the male and female subsamples, respectively. The key explanatory variables are DID, schooling and the spatially lagged migration (W*migration). We begin with results based on pooled OLS where the results show a positive significant impact of the land reform policy on migration while holding other factors constant.

Next, we use the fixed-effects regression model (without spatial effects), which allows us to control for unobserved and time-invariant characteristics that may influence migration. The results again show a positive significant effect but only for the female subsample. The third regression (SAR) is also a fixed effects regression model but takes into account spatial dependencies between variables. Although the positive effect is retained, the sizes of coefficient reduce after controlling for spatial effects. There is also clear evidence of spatial spill-overs in migration between areas as shown by large and statistically significant coefficients for the spatial lag term, namely $W * migration$. Our results are also robust to the choice of spatial models because the coefficients also remain stable for the SEM and SAC models (see the results in Table A 11 in the appendix).

Overall, three issues are worth noting from the results. Firstly, allowing for spatial effects further reduces the size of coefficients for our variable of interest (DID). Secondly, higher migration in one area is associated with higher migration in neighbouring areas, as indicated by statistically significant coefficients for $W * migration$ and $lambda$ (shown in the appendix). Thirdly, there seems to be a gender dimension to our results. We observe that compared to males, the coefficients for females are not only larger but statistically significant even after controlling for spatial effects.

Table 4.8: Effects of land reform policy on migration

	Male			Female		
	OLS	Non spatial	SAR	OLS	Non spatial	SAR
Time	-0.180***	-0.188***	-0.089***	-0.153***	-0.076**	0.001
	(0.010)	(0.028)	(0.029)	(0.009)	(0.031)	(0.025)
Treated	-0.039***			-0.036***		
	(0.014)			(0.012)		
DID	0.035*	0.029	0.019	0.036**	0.038*	0.027**
	(0.020)	(0.020)	(0.014)	(0.017)	(0.020)	(0.013)
Schooling	-0.004	-0.004	-0.009	0.000	-0.022	-0.023**
	(0.004)	(0.014)	(0.009)	(0.003)	(0.014)	(0.009)
Age	0.056	0.106	0.097*	-0.045	-0.123	-0.130**
	(0.053)	(0.083)	(0.056)	(0.036)	(0.078)	(0.052)
Age squared	-0.001	-0.002	-0.002**	0.001	0.002	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Married	0.229**	0.427**	0.389***	0.083	0.174	0.108
	(0.095)	(0.177)	(0.120)	(0.061)	(0.152)	(0.102)
Employed	-0.036	-0.08	-0.081**	-0.112***	-0.116***	-0.114***
	(0.039)	(0.057)	(0.038)	(0.022)	(0.035)	(0.023)
Assets	0.039***	0.034**	0.034***	0.020***	-0.002	-0.005
	(0.005)	(0.014)	(0.010)	(0.005)	(0.015)	(0.010)
W*migration			0.726***			0.783***
			(0.160)			(0.131)
Observations	356	356	356	356	356	356
Spatial units	178	178	178	178	178	178

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; standard errors in parenthesis

We further conducted balancing tests, which is a t-test of the difference in the means of the covariates between the control and treated groups in the base period. This test was performed Stata's user defined command, namely $diff$. The balancing tests presented in Table 4.9 show that there are no differences in observed mean outcomes between the treatment and control samples. This further confirms that our results are robust. Moreover, the World Bank (2012) has rated the land reform policy as satisfactory and exceeding expectation. According to their findings, 15,142 households were successfully relocated against a target of 15,000. Therefore, our finding that the people successfully migrated is substantiated.

Table 4.9: Balancing tests for base period

Description	Weighted Variable(s)	Mean Control	Mean Treated	Diff.	t	P-value
Male						
	Migration	0.150	0.106	-0.044	3.12	0.0021***
	Schooling	2.565	2.472	-0.093	0.74	0.4600
	Age	20.736	20.622	-0.115	0.72	0.4720
	Married	0.333	0.326	-0.007	1.48	0.1400
	Employed	0.646	0.636	-0.01	1.18	0.2410
	Assets	-0.410	-0.500	-0.09	0.56	0.5770
Female						
	Migration	0.148	0.093	-0.054	3.69	0.0003***
	Schooling	1.275	1.233	-0.042	0.4	0.6888
	Age	21.681	21.758	0.077	0.35	0.7268
	Married	0.364	0.361	-0.003	0.82	0.4132
	Employed	0.623	0.627	0.004	0.22	0.8271
	Assets	-0.463	-0.595	-0.132	0.9	0.3707

Note: *, **, *** denote significance at 10%, 5% and 1% levels

4.4.2 Effects of land reform policy on employment

Regression results based on agricultural and government employment are presented in Table 4.10 and Table 4.11 respectively for both the male and female subsamples. It is important to note that our treatment dummy in this section is for the areas where the people were resettled. Effectively, our results should be understood in terms of the regions where the land reform was implemented.

DID, schooling and the spatially lagged employment (W*employed) are the three explanatory variables of focus. The results show that the land reform had positive significant effects on agricultural employment, more especially for females where the coefficients for DID (non-spatial fixed effects and SAR) are not only significant but also generally larger than for males. Similar findings are established for the SEM and SAC as reported in Table A 12 in the appendix. The programme, therefore, might have improved the opportunities of women in agricultural employment.

The gender dimension of our results has implications for policy and poverty reduction because the participation of females in employment contributes to economic growth. For example, in a recent study, the Food and Agriculture Organisation (2011) found that by closing the resource gap between men and women in developing countries, agricultural yields for females could potentially improve by between 20% and 30%. As a result, total agricultural output of developing countries would grow by between 2.5% and 4.0%. In turn, the levels of malnutrition would decline by between 12% and 17%, which would globally imply that 100 to 150 million people would be lifted out of hunger.

Table 4.10: Effects of land reform policy on agricultural employment

	Male			Female		
	OLS	Non spatial	SAR	OLS	Non spatial	SAR
Time	-0.289***	-0.390***	-0.306***	-0.316***	-0.443***	-0.252***
	(0.035)	(0.045)	(0.106)	(0.028)	(0.056)	(0.087)
Treated	0.080***			0.032		
	(0.027)			(0.021)		
DID	0.036	0.025	0.025	0.042	0.048*	0.045**
	(0.037)	(0.029)	(0.020)	(0.029)	(0.028)	(0.019)
Schooling	-0.064***	0.077**	0.074***	-0.042***	0.060	0.057**
	(0.010)	(0.033)	(0.023)	(0.006)	(0.040)	(0.027)
Age	2.989***	0.225	0.237	1.840***	0.986*	1.020***
	(0.549)	(0.777)	(0.538)	(0.237)	(0.507)	(0.348)
Age squared	-0.048***	-0.004	-0.004	-0.030***	-0.016*	-0.016***
	(0.009)	(0.012)	(0.009)	(0.004)	(0.008)	(0.006)
Married	0.109	-0.071	-0.103	0.217	-0.100	-0.103
	(0.282)	(0.425)	(0.297)	(0.146)	(0.378)	(0.260)
Dependency	0.677***	-0.262	-0.257*	0.199*	0.105	0.111
	(0.188)	(0.202)	(0.140)	(0.118)	(0.155)	(0.106)
W*employed			0.268			0.569**
			(0.324)			(0.233)
Observations	356	356	356	356	356	356
Spatial units	178	178	178	178	178	178

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; standard errors in parenthesis

The results further show that government employment increased for both males and females but the magnitude was larger for males (see Table 4.11 and Table A13). The analysis was repeated for other types of employment such as self-employment and private employment (see results in Table A 14 in the appendix) and we found a negative coefficient for DID, which is significant for both types of employment for males and also in the case of private employment for women. The opposite sign of these coefficients suggests some trade-off between wage sectors, particularly when contrasted with agricultural and government employment where the coefficient is positive for both males and females.

We find positive and significant effects of schooling for all models, except for OLS where the coefficient is negative in agricultural employment. More generally, it is interesting to note that the coefficient for schooling reduces in size after taking into account spatial dependence. The reduction demonstrates the importance of taking into account spatial effects. The statistically significant spatially lagged term W ∗ employed also confirms evidence of spatial spill-overs which implies that the employment status in each small geographical areas affects the likelihood of employment in other neighbouring regions.

Table 4.11: Effects of land reform policy on government employment

	Male			Female		
	OLS	Non spatial	SAR	OLS	Non spatial	SAR
Time	0.042***	0.063***	-0.010	0.017***	-0.006	-0.032**
	(0.007)	(0.015)	(0.011)	(0.006)	(0.014)	(0.013)
Treated	-0.011**			-0.011**		
	(0.005)			(0.004)		
DID	0.029***	0.027***	0.022***	0.015**	0.013*	0.012**
	(0.008)	(0.009)	(0.006)	(0.006)	(0.007)	(0.005)
Schooling	0.015***	-0.005	0.004	0.008***	0.027***	0.029***
	(0.002)	(0.011)	(0.007)	(0.001)	(0.010)	(0.007)
Age	-0.335***	-0.415	-0.344**	-0.116**	0.455***	0.453***
	(0.113)	(0.252)	(0.162)	(0.049)	(0.130)	(0.089)
Age squared	0.005***	0.007*	0.006**	0.002**	-0.007***	-0.007***
	(0.002)	(0.004)	(0.003)	(0.001)	(0.002)	(0.001)
Married	-0.004	-0.008	-0.058	-0.019	0.161*	0.134**
	(0.058)	(0.138)	(0.089)	(0.030)	(0.097)	(0.067)
Dependency	0.039	0.061	0.04	0.006	0.044	0.041
	(0.039)	(0.065)	(0.042)	(0.024)	(0.040)	(0.027)
W*employed			0.913***			0.648***
			(0.060)			(0.195)
Observations	356	356	356	356	356	356
Spatial units	178	178	178	178	178	178

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; standard errors in parenthesis

The World Bank (2012) also cites some positive outcomes of the land reform policy which we briefly discuss. The positive effects include an expansion in the average land holding size from less than 0.5 hectares before to approximately 2.2 hectares after the project. In addition, households that participated reported an improvement in maize production (maize stocks after relocation lasted 10.7 months compared to about 3.6 months before); incomes of relocated households grew by 6 times when compared to the control group; maize and tobacco yields were respectively 4 and 2.6 times higher compared to the base period; yields also reached an average level of 50 to 60% higher as compared to control groups in the surrounding areas.

Due to the positive results and favourable experience, the government and the World Bank are willing to scale up the programme to cover the whole country, where at least 100,000 households would be resettled (World Bank, 2012). However, while the positive benefits of the reform are widely acknowledged, Chinsinga (2008) raises a number of issues that need to be addressed to ensure its sustainability. The main concern surrounds access to functioning services such as health, water, transport and markets. For example, it was reported that due to lack of markets for produce some farmers ended up selling at very low prices. Related to this, an outbreak of cholera almost led to a collapse of one of the settlement trusts. Specifically, about 20 out of 35 households belonging to Kalungu trust immigrated back to Thyolo for lack of access to portable water facilities.

4.5 Conclusions

In this chapter, we created panel data at the level of traditional authorities from national censuses to examine spatial and temporal patterns of employment and migration in Malawi. We make a distinction between males and females because this has implications for development policy.

The results showed that over the long-term changes in the population structure have effects on labour force participation rates. We also found clear evidence of spatial spill-overs in our phenomena of interest. Further, allowing for spatial effects reduces the sizes of coefficients and this confirms arguments in the literature that failure to take into account spatial dependencies tends to overstate the results. The results show that Malawi's community-based land reform policy launched in 2004 influenced migration patterns which were largely rural-rural. Also as a result of the land reform, there was a trade-off in the employment sectors with respect to employment. On the one hand, agricultural and government employment improved, more particularly for the females. On the other hand, self-employment and private employment declined. Finally, by matching GIS codes at the level of traditional authority areas, the work undertaken for this study makes it possible to integrate census data with other Malawian data sets for similar spatial analysis.

In terms of policy, since most indicators are heterogeneous within the country, analysis at low level geographical units as we have done has benefits of allowing for a detailed understanding of phenomena and designing interventions adapted for specific areas. Spatial visualisation of data is also particularly useful for policy makers because it is intuitive.

Our study uses aggregated data at the traditional authority level rather than individual level data. It has been argued in the literature that aggregated data ignores the rich cross-sectional evidence of the individual observations and fails to say anything about the industries, firms and individuals themselves. Moreover, geographical aggregation is arbitrary with respect to the definitions of the spatial observational units. One of the solutions suggested in the literature is to estimate models at a micro level

(Arbia, 2016). However, the main challenge with this approach is the size of the spatial weighting matrix, considering that in censuses we are typically working with millions of observations. The current packages available to the researcher can only handle up to $11{,}000$ x $11{,}000$ for Stata 14.1 SE or MP (e.g., Belotti et al., 2013). This may be an area for further study as we explore other available options.

Chapter 5
Conclusion

5.1 Introduction

This thesis has used multiple data sources to investigate trends in non-monetary dimensions of welfare, labour market outcomes and migration in Malawi. As noted in Chapter 1, the issues are investigated in three separate chapters, namely Chapters 2, 3 and 4. Section 5.2 provides the summary of findings from each of the chapters. In Section 5.3, we discuss the conclusion of the thesis and implications for policy. Research implications are discussed in Section 5.4. Contributions made to the literature are summarised in Section 5.5. Finally, in Section 5.6, we provide suggestions for future research.

5.2 Summary of findings

Chapter 2 was dedicated to the analysis of poverty and inequality over time using two non-monetary measures of welfare, namely child nutritional status and household asset ownership. Data for this study was drawn from nationally representative DHS data sets for 1992, 2000, 2004 and 2010. On the one hand, the temporal aspect of the study provided an understanding of long-term economic well-being in Malawi. On the other hand, the spatial aspect gave the profile of welfare across population groups such as regions and urban-rural areas. The study showed that poverty in Malawi has declined tremendously from as high as 80% in 1992 to about 50% in 2010. Having established that welfare has improved over time, the study set out to investigate who benefited from these gains in welfare. This was investigated through pro-poor growth analysis which showed that all population groups benefited but the greatest gains accrued to the poor as opposed to the rich. This finding implies that inequality in Malawi has declined between 1992 and 2010.

The study also conducted rankings of welfare that are robust to the choices of the poverty line and the dimension of well-being. For example, we found that poverty and inequality are higher in rural areas than urban areas regardless of whether we use assets or child-nutritional status. Our results show that while welfare differs among regions, the regional differences are not as large as the difference between rural and urban areas. For example, using the asset index, only about 7% of the population is poor in urban areas compared to 53% in rural areas. The gap is huge when contrasted with regional poverty rates of 32.4%, 51.3% and 44.5% for the Northern, Central and Southern regions, respectively. Nevertheless, we conclude that poverty in Malawi is both a rural area and regional problem. Therefore, policy interventions should focus more on reducing poverty in rural areas while not neglecting the regional imbalances.

One of the advantages of the measures of poverty and inequality employed in Chapter 2 is that they are decomposable. Poverty decomposition shows that urban areas and households headed by females contribute less to poverty compared to rural areas and households headed by males, respectively. We also decomposed total inequality into between-group (for example, rural and urban) and within-group (for example, rural-urban asset gap) inequality. According to Haughton & Khandker (2009), within-group inequality is typically at least 75% of total inequality in a given country. Our results confirmed that inequality is indeed higher within groups as opposed to between groups; we find magnitudes of between 65% and 100% depending on the measure of welfare used as well as population group pair.

Chapter 3 used the two waves of the IHS3 panel data, namely 2010 and 2013, to analyse returns to education in the wage sector and externalities to education in household enterprises. Three main issues were considered. Firstly, we discussed the various issues that could affect the reliability of the returns and externalities to education. The conclusion was that a robust treatment of outliers and inconsistencies in the data makes it possible to obtain trends in the labour market that are both consistent and reliable.

Secondly, we tested the argument that educational attainment is positively associated with poverty reduction through the labour market. This assertion was investigated through an econometric analysis of the role played by education in the determination of the likelihood of labour force participation and employment. After controlling for other factors, we found that individuals with tertiary education are more likely to participate in the labour market than those without education. However, the results showed that those with primary and secondary are less likely to participate in the labour market than those without education. We reason that those without formal education drop out of schooling to immediately enter the labour market, unlike those with primary and secondary education who choose to continue with their education and, therefore, only enter the labour market at a later stage. Moreover, the largest proportion of people without education is engaged in informal sector jobs where earnings are low. We follow a similar line of thinking with the multivariate analysis of employment where we find that those with JCE and MSCE are less likely to get employed compared to those without education.

Thirdly, the study investigated if higher levels of schooling or education are associated with higher incomes. Our findings, which are robust to sample selection, yielded large positive returns to years of schooling in Malawi for both wage employment and self-employment in household enterprises. The analysis was repeated with education categories instead of years of schooling. The results did not change, implying that the definition of education does not matter. However, using education categories gives us more information in the sense that we are able to show that the returns to education are heterogeneous rather than homogeneous. Not only do the returns increase by the level of education but are also higher amongst females compared to males. Through these findings, the study demonstrates the importance of distinguishing between groups such as economic sectors and gender when analysing labour market outcomes.

In Chapter 4, trends in employment and migration were examined using the 1987, 1998 and 2008 census data which are the only three publicly available census data sets in the country. The study has shown that both geography and time matter in the understanding of long-term patterns of employment and migration. While time is important, we demonstrated that it is not everything we should be controlling for when analysing economic outcomes that are essentially spatial such as employment and migration. The importance of geography is proven by the fact that the magnitude of explanatory coefficients dramatically decline once we take spatial dependencies into account in our regressions.

Our analysis was done in two stages. In the first stage, we confirmed that spatial dependencies exist in our data using the two most commonly used tests for spatial autocorrelation, namely the Moran's I and Geary's c statistics. Spatial distributions on the map of Malawi showed some heterogeneity and clustering in the variables of interest. The detection of spatial dependence only provided the first step in spatial data analysis since it is possible that such clustering could occur randomly. Therefore, as a second step, we conducted some diagnostic tests in order to understand the type of clustering. This further informed the specification of the spatial regression model. The diagnostic tests showed that it is possible to fit both the spatial lag and error models or their variants with our data. Therefore, our regression results are robust to the choice of model specification from a whole range of the available model options commonly used in spatial panel data analysis.

The study also explored the effects of the Malawi Community Based Rural Land Development Project (CBRLDP) on migration and employment. The results show that the land reform policy positively influenced migration patterns and also resulted in increased employment opportunities for the individuals who migrated. These results have implications for policy. By showing that it is possible for people to migrate in the pilot phase, the programme can potentially be scaled up to cover the whole country given the positive benefits of the programme in terms of gains in employment and incomes for the participating households. The World Bank and Government of Malawi already have plans to expand the programme to cover the whole of the country for similar settlements. The results also revealed an interesting gender dimension in the sense that employment effects differ between males and females; not only did agriculture employment improve by larger magnitudes for the female subsample when compared to males, but was also significant, unlike the male subsample where we obtained non-significant results. Therefore, we expect the programme to have improved the incomes of women involved in agricultural employment for a living. This finding is important because the Food and Agriculture Organisation (2011) states that if resource gap in developing countries between men and women were to be closed, the latter could potentially increase yields by between 20% and 30% and lift many out of food poverty. Although government employment increased for both the male and female subsamples, the magnitude was larger for males.

5.3 Conclusions

The poverty and inequality profiles derived in Chapter 2 suggest that although the welfare changes over time have pro-poor, Malawi remains a poor country in terms of both assets and child-nutritional status. Malawi's poverty is, therefore, multidimensional. A robust ranking of welfare shows that poverty and inequality levels are highest in rural areas and amongst households headed by females. Rankings for the three regions of Malawi depend on the welfare measure used. While asset poverty is the highest in the Central region, child-nutritional status is highest in the Southern region. With respect to inequality, while no robust ranking is determined with respect to assets, height-for-age z-scores (HAZ) and weight-for-age z-scores (WAZ), we are able to conclude that inequality is the lowest in the Northern region when weight-for-height z-scores (WHZ) are used as a measure of welfare. Amongst the three measures of child-nutritional status, HAZ yields the highest levels measures of poverty and inequality.

Regardless of the welfare measure used, poverty decompositions seem to suggest that poverty is the highest in groups where the population shares are the largest. For example, rural areas which constitute about 85% of the population, contribute to about 91.4% of the poverty using WHZ as a measure of child-nutritional status. Similarly, the Central region and households headed by males contribute more to poverty compared to other population groups.

Using multivariate analysis, we have identified factors associated with asset poverty and child nutritional status in Malawi. With respect to child-nutritional status, the study has shown that child-individual characteristics such as age, sex and birth order play an important role in addition to household welfare and the levels of education by both the father and mother. It was also established in Chapter 2 that household characteristics such as household size, dependency ratio, levels of education in the household and sex of the household head are important correlates to asset poverty. Furthermore, asset poverty is geographical and depends on the area of residence.

Chapter 3 confirms the human capital theory. First, we show that education indeed improves chances of employment, particularly for those with tertiary education. Secondly, we show that returns to education increase with levels of education. We also find that the returns are higher for females than males. Furthermore, given that wage employment is a small proportion of total employment, we segregate our results by economic sector. Our results show that modelling returns to education based on the formal sector only is misleading as it assumes a homogeneous labour market which is not true for developing countries such as Malawi. Finally, our results show that we should be correcting for sample selection when analysing labour market outcomes in Malawi.

In Chapter 4, we confirmed that employment and migration are spatial economic phenomena. Econometric analysis of these variables of interest should ideally, therefore, control for both time and geography. We also showed that, between 1987 and 2008, there has been an increase in labour force participation of the youth in Malawi, particularly those aged between 15 and 19 years. This might have implications for youth unemployment in the country. The analysis of long-term occupational mobility showed that a larger proportion of women remains trapped in agricultural employment compared to males. The results also show that the land reform policy in Malawi caused people to migrate to other areas in search of agricultural land. In terms of the effects on employment, agricultural employment increased for the female subsample while government employment particularly increased for the males.

5.4 Implications of the research

The main aim of this research was to contribute to the empirical understanding of poverty, inequality and the labour market in Malawi. Using multiple sources of data which stretch over time, the study shows interesting dynamics of the aforesaid economic phenomena in a country which has been largely static in terms of economic growth and urbanisation. One of the main conclusions one can draw from the study is that Malawi's poverty profile is a 'bad picture' given that almost 50% of the population was still poor in 2010, but a 'good movie' in that the incidence of poverty had fallen from as high as 80% in 1992. This holds for both monetary and non-monetary indicators of welfare. A number of implications can be drawn from this research and these may have broader policy applications. Following the structure of the thesis, the implications are discussed separately for each of the three main chapters.

Chapter 2 provided research evidence on poverty and inequality in Malawi using two non-monetary dimensions of welfare, namely an asset index and child-nutritional status. The first implication arising from this chapter emanates from the finding that Malawi has very high levels of poverty regardless of whether assets or child-nutritional status are used as the measure of welfare. Other research based on consumption (e.g. Mussa, 2013) derives similar implications with respect to this finding. As shown in Chapter 2, despite improvements over time, the data shows very low levels of asset ownership in Malawi. There is also poor access to sanitation, and poor quality of dwellings as reflected in the floor material, as well as poor water services. These have either remained the same or declined over time. Therefore, despite gains in poverty reduction, Malawi remains a very poor country; addressing this problem remains extremely important.

A second implication relates to the finding that there is not much regional variation in terms of both poverty and inequality, as shown by the narrow bands in the poverty and inequality mapping shown in Chapter 2. This suggests that the population subgroups are largely undifferentiated. This homogeneous feature of welfare seems to be a feature that is quite unique to Malawi, as few other developing countries show such small differences in poverty measures across geographic space. Similarly, we find that there are no large differences between male and female-headed households, a finding not so common in most developing countries, including most of Malawi's neighbours. Perhaps these findings relate to the largely undiversified nature of the livelihoods of the Malawian population.

The research further established that rural areas dominate urban areas in terms of poverty. This finding is not surprising, considering that 85% of the population in Malawi lives in rural areas and that urban areas usually offer at least some improved formal sector wage earning opportunities. Nevertheless, it is important to note that despite the rural dominance, poverty is also quite high in urban areas. For example, incidences of stunting in urban and rural areas stand at 40.8% and 48.3%, respectively. Thus, while government policy should focus on rural areas, it is important not to neglect urban areas.

In stark contrast with consumption-based inequality estimates normally used in official reports, this study interestingly shows that non-money metric inequality is higher in rural areas compared to urban areas. This is a new finding and responds to the call by some, such as Grosse et al. (2008), who have highlighted the need to extend pro-poor growth analysis to non-monetary dimensions. Clearly, this has implications for how government considers its policy on inequality. It may be necessary to rethink the policy approach to reducing inequality in Malawi. Specifically, policy makers need to explicitly take into account non-income dimensions of welfare when formulating public policy.

Chapter 3 used panel data to estimate externalities and returns to education in Malawi. Although the chapter shows that returns to education are positive in both the formal and informal sectors, they are lower in the latter where the majority of the population is employed. Therefore, unless workers move from the informal sector to the formal sector, the benefits from education remain low for the majority of the workforce. A second feature evident from the analysis in Chapter 3 relates to the structure of employment in Malawi. Casual employment and self-employment activities (household enterprises) continue to make up the largest share of total employment both in and outside of agriculture. They jointly constitute at least 75% of total employment.

The chapter also shows that Malawi has very high labour force participation rates, ranging between 71% and 91%, depending on whether the broad or narrow definition is used. These traditionally high rates of labour participation can be linked to the existence of a large agricultural sector which in turn makes it difficult to clearly identify actual labour force participation rates, as subsistence activities in the agricultural sector are categorised as labour force participation.

The chapter also shows that the labour force in Malawi remains poorly educated; only 17% of individuals in the working age population has more than primary education. The low levels of education imply low productivity levels that partly account for the large earning differentials between the formal and informal sector. Consequently, there is a need to continue investing in education which has been found to yield positive returns for those employed.

The fact that returns to education are higher for females than males with similar skills (more particularly at high levels of education) shows that the continued expansion of education for girls should translate into gains in the fight against poverty and the economic empowerment of women. It is further shown in the study that the returns to education in non-farm enterprises are higher when maximum household education is used (signifying externalities) as opposed to the level of education of the owner of the enterprise. This suggests that provision of free primary education and subsidies in higher education can be justified on these grounds. In this case, government's investment in education is not only the right thing to do but also have much wider benefits for society in terms of poverty reduction and improvements in food security. Furthermore, the presence of education externalities raises issues for further study. For example, motivated by findings emergent from this study, one may want to further investigate how the allocation of education to different household members of different ages or gender involved in various activities would affect household income. Simply put, whose education matters in the determination of household income?

The use of panel data has enabled an examination of changes in occupational status as well as movements into and out of the labour market. This is not only important for monitoring trends but also for identifying the sources of changes over time. For example, from the data, it is possible to identify individuals who are employed or unemployed in both waves. This is not possible with cross-sectional data. Consequently, panel data not only provided transparent ways of dealing with data issues but also allow a better understanding of the economic phenomena of interest, including how earnings have changed between surveys. The data shows that compared to those only employed in either wave, the initial earnings amongst individuals employed in both years were not only higher (almost double as much) but also increased by a greater magnitude. Specifically, the earnings of those employed in both waves increased by 46% between 2010 and 2013. While most of the individuals employed in both waves were educated, the majority of those unemployed had no education. This seems to suggest that education not only improves chances of finding employment but also that higher education is associated with higher earnings. Furthermore, with panel data, it was possible to control for individual heterogeneity since the individuals served as their own controls.

Chapter 4 looked at migration and employment in Malawi using spatial data. The government has generally regarded the movement of people within Malawi in the light of rural-urban migration which is seen as a natural and inevitable process. Policies have, therefore, tended to focus on developing rural centres to prevent massive movements of people into urban areas, particularly the four main urban centres, namely Mzuzu, Lilongwe, Blantyre and Zomba. However, given the low levels of urbanisation, rural-urban migration is not as massive as perceived. On the contrary, the observed patterns of migration in Malawi are largely rural-rural. Specifically, while most people stay in the home regions there are significant movements across districts. This points to the continued dominance of rural livelihoods, even for those in search of better opportunities.

Furthermore, the study has shown that the land reform policy enabled people to move from the rural areas of sending districts into the rural areas of receiving districts. This policy has had large implications for development because of its effects on employment. Consequently, government policy needs to go beyond simply providing access to land because this may potentially create as many problems as it seeks to solve. On a more substantive level, deliberate policies need to be put in place to support agricultural production in addition to the provision of markets and other necessary amenities.

In Chapter 2, it was shown that there are no major differences between regions in terms of economic welfare. However, there exists some heterogeneity in terms of economic indicators when small-level geographical data is used as shown in the spatial maps in Chapter 4. Specifically, while there seems to be broad uniformity across the main regions, a descriptive spatial analysis of employment, schooling, fertility and assets shows some heterogeneity at the smallest geographical units, namely traditional authorities. One implication of this finding is that policy should not be based on the average picture but instead needs to be tailored to the small geographical areas. In addition, visualisation of spatial is a powerful tool and can be used to aid policies that are specifically adapted for small geographical units.

Some studies (e.g. Gaddis & Klasen, 2014) link fertility rates to economic development. Chapter 2 of this thesis has shown that fertility rates in Malawi have dropped between 1987 and 2008. Although this is the broad trend for the population, fertility rates have increased by almost 2% amongst women aged between 20 and 24 years, contrary to what is desired; this may call for policy action. High fertility rates are known to have effects on education; education expenditures per child tend to be lower in large families. Furthermore, high fertility rates may also negatively affect the supply of labour of the parents, especially when the children are still young. This may, in turn, have a negative effect on the ability of families to save for the children.

The analysis of the long-term patterns of labour force and population in Malawi shows that the percentage of the youth participating in the labour force has been on the increase, particularly for females. These observed changes in the age structure of the population and of labour force participation have significant implications for the supply of labour and therefore also the need for economic growth.

There is a need for the corresponding improvement in education and skills development. The youth, particularly those aged between 20 and 24 years, constitutes the largest percentage of the unemployed. Youth unemployment has implications for development. For example, as education continues to expand, many of those educated may fail to find employment due to increased competition for scarce formal sector jobs. Even those who eventually enter into employment may actually end up in informal employment and subsistence agriculture where the earnings are low.

Another important implication of the research is that it provides a platform for further spatial analysis in Malawi. Specifically, the study has demonstrated the feasibility of creating GIS codes that are consistent over time. This is an important step towards linking census data with other national data sets for similar analysis at the lowest geographical unit.

This thesis has thus shown that Malawi, one of the poorest countries in the world, has many of the features of poor countries, but also some features that make it quite unique. These include the relatively undifferentiated economic activity in rural areas, which lies at the heart of the limited economic and spatial differentiation across the country as a whole. This thesis has also shown that there has been considerable pro-poor growth in Malawi, something that is to be commended and holds promises that future growth too may benefit Malawi's many poor people. As the country develops, however, it is likely that differentiation would increase and that heterogeneity would become the norm in the economic field as in the labour market. The data used in this study offer a useful first attempt at evaluating the features and consequences of economic progress. Further work on new datasets that become available over time would be beneficial for following the trajectory of growth and development.

5.5 Summary of contributions

Previous studies on poverty and inequality in Malawi have been static in nature, due to data limited availability. This study was able to conduct more inter-temporal analysis that was made possible by increased data availability. Chapter 2 has contributed to the literature by examining trends in poverty through pro-poor growth analysis which allows us to understand the distribution of the gains from reductions in poverty and inequality over time. More generally, this is the first attempt to apply an asset index to the measurement of poverty in Malawi. While it is generally understood that poverty is multi-dimensional, there has been a dearth of literature dedicated to the empirical understanding of asset poverty and how the results compare with what is already known with respect to other dimensions such as consumption, education and nutritional status.

Until the release of the IHS3 panel data (IHSP), there has been a lack of panel data in Malawi. Therefore, previous evidence on education and earnings has been based on cross-sectional data. In Chapter 3, we take advantage of the recently made available panel survey results set to explore the interesting dynamic aspects of the observations as well as control for unobservable individual heterogeneity. Furthermore, we conduct robust treatment of outliers and bring to light some of the inconsistencies in the data. These are some of the issues that new users of the data will have to look out for in order to ensure quality data before further analysis. This more generally applies to most datasets in less developed countries with a large informal sector.

The main contribution in Chapter 4 is the application of spatial econometric methods to the understanding of employment and migration in Malawi. For example, by comparing results based on spatial effects to those based on non-spatial regression analysis, we demonstrate that studies that do not take into account spatial effects tend to overestimate other coefficients. Potentially, this study provides an exploratory platform for further spatial analysis in Malawi. In order to achieve this, we have matched GIS codes that are consistent over time with the view to link census data with other data sets for similar spatial analysis. This kind of data integration is an area for further study, as noted in the ensuing section.

5.6 Future research

Chapter 2 dealt with the dynamic aspects of poverty and inequality in Malawi by examining trends over time. However, since we are not tracking the same households, our results only show the overall population changes in welfare that have occurred over time. Another important aspect of the dynamic nature of poverty considered in the literature is the identification of the proportions of households or individuals that are either chronically or transiently poor. This kind of analysis, which can only be achieved with panel data, has not been done in this study. Future research can take advantage of the IHSP data which has been recently released by the National Statistical Office.

Furthermore, in Chapter 3, we identified measurement error in the independent variables as one of the possible sources of bias when estimating returns and externalities to education. The most commonly suggested method of dealing with this kind of measurement error in the literature is differencing (short and longer differencing). Unfortunately, this approach only works for panels with at least three waves of data panel, which is not available in Malawi at the moment. However, as more waves are added to the IHSP in the future, there would be the possibility of accounting for measurement error in the independent variables.

Another area for future research relates to the treatment of individuals with zero earnings. Our approach has been to exclude from the analysis observations with zero incomes. One of the criticisms of this approach is that it may affect the reliability of the estimates (Yu, 2012), especially if the observations are many. Therefore, researchers sometimes consider an alternative of adding a very small number to earnings before taking the log, which would allow considering zero incomes. This option is not considered in this study because zero incomes constitute about 10% percentage in our data which we believe is large and may, therefore, eventually end up distorting the distribution of the data. As a further study, one may set out to investigate the alternative treatment of zero earnings not only for the panel data set but also for other data sets in the country including consumption figures.

The spatiotemporal analysis carried out in Chapter 4 can be generally applied to economic phenomena by matching GIS census data with other data in Malawi. This constitutes an area for future research. With the development of geosystems, it can be expected that geo-coded data will become more readily available in Malawi. This will make spatial analysis possible on the new datasets, but the richness of such data could perhaps be further enhanced by adding information from the census for smaller geographical areas. Thus, this study acts as a platform for further spatial analysis. Another extension of this study would be to conduct spatial analysis using individual level data instead of aggregated data as is the case in the present study. However, this approach requires working with very large matrices involving millions of observations, as is typically the case with census data. The current Stata packages only handle smaller matrix sizes as explained in Section 4.5. Consequently, use of individual data for spatial analysis is dependent on the identification of methodologies and tools that overcome this limitation.

List of references

Alkire, S. & Santos, M.E. (2010). *Acute Multidimensional Poverty: A New Index for Developing Countries* (OPHI Working Paper No. 38). Retrieved from https://www.ophi.org.uk/wp-content/uploads/ophi-wp38.pdf.

Anselin, L. (1992b). *Spatial Data Analysis with GIS: An Introduction to Application in the Social Sciences.* National Center for Geographic Information and Analysis, University of California, Santa Barbara.

Anselin, L. (2003). Spatial Econometrics, in Baltagi, B.H. (Ed.), *A Companion to Theoretical Econometrics.* Malden, Massachusetts, USA: Blackwell Publishing Ltd. doi: 10.1002/9780470996249.ch15.

Appleton, S., Bigsten, A. & Manda, D.K. (1999). *Educational Expansion and Economic Decline: Returns to Education in Kenya, 1978-95* (Working Paper WPS/99-6). Centre for the Study of African Economies, University of Oxford.

Araar, A. & Duclos, J.Y. (2013). DASP: Distributive Analysis Stata Package [Computer Software]. Laval University, *World Bank, PEP and CIRPÉE.*

Arbia, G. (2016, June). *Spatial econometrics: a broad view*. Paper presented at the X World Conference Spatial Econometrics Association (SEA), Rome, Italy.

Baltagi, B. H. & Li, Q. (1990). A Lagrange multiplier test for the error components model with incomplete panels. *Econometric Reviews* 9 (1), 103–7.

Baltagi, B.H. (2013). Panel Data Methods. *Handbook of Applied Economic Statistics*. New York.

Basu, K. & Foster, J. E. (1998). On measuring literacy*, Economic Journal*, 108(451), 1733-49.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *The journal of political economy*, 9-49.

Belotti, F., Hughes,G. & Mortari, A.P.(2014). "XSMLE: Stata module for spatial panel data models estimation," Statistical Software Components S457610, Boston College Department of Economics.

Bennell, P. (1996). Rates of Return on Education: Does the Conventional Pattern Prevail in Sub-Saharan Africa. *World Development*, 24 (1), 183-99.

Blundell, R. & Preston, I. (1998). Consumption inequality and income uncertainty. *Quarterly Journal of Economics*. 113(2), 604-40.

Bokosi, F.K. (2006). Household poverty dynamics in Malawi: A bivariate probit analysis. *Journal of applied sciences*, 7 (2), 258-62.

Booysen, F., Van der Berg, S., Burger, R., Von Maltiz, M., & Du Rand, G. (2008). Using an Asset Index to Assess Trends in Poverty in Seven Sub-Saharan African Countries. *World Development,* 36 (6), 1113-1130.

Bourguignon, F. & Chakravarty, S.R. (2003). The measurement of multidimensional poverty. *Journal of Economic Inequality*, 1 (1), 25-49.

Burger, R. & Yu, D. (2007). *Wage trends in post -Apartheid South Africa: Constructing an earnings series from household survey data* (Working Paper 07/117). Development Policy Research Unit, University of Cape Town.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, *3*, 1801-1863.

Casale, D. & Desmond, C. (2015). Recovery from stunting and cognitive outcomes in young children: Evidence from the South African Birth to Twenty Cohort Study. *Journal of developmental origins of health and disease*, available on CJO2015. doi:10.1017/S2040174415007175.

Chen, M. A. (2012). The informal economy: Definitions, theories and policies. *Women in informal economy globalizing and organizing: WIEGO Working Paper, 1.*

Chirwa, E. & Matita, M. (2009). *The Rate of Return on Education in Malawi* (Department of Economics Working Paper No. 2009/01). University of Malawi, Chancellor College.

Chirwa, E.W. & Ngalawa, H.P.E. (2008). Determinants of Child Nutrition in Malawi. *South African Journal of Economics*, 76(4), 628-40.

Chirwa, E.W. & Zgovu, E.K. (2002). *Does the Return to Schooling Depend on the Type of Employment? Evidence from the Rural Labour Market in Malawi*.

Chinsinga, B. (2008). Exploring the politics of land reforms in Malawi: A case study of the Community Based Rural Land Development Programme (CBRLDP). *University of Manchester: Manchester, UK*.

Clark, T.S. & Linzer, D. A. (2015). Should I Use Fixed or Random Effects? *Political Science Research and Methods*, 3(2), 399–408.

Crow, K. (2015). SHP2DTA: Stata module to converts shape boundary files to Stata datasets.

Cooke, T. J. (2013). Internal migration in decline. *The Professional Geographer*, *65*(4), 664-675.

Da Maia, C. (2012). *Understanding Poverty and Inequality in Mozambique. The Role of Education and Labour Market Status* (Doctoral thesis, Stellenbosch University, South Africa). Retrieved from http://scholar.sun.ac.za/handle/10019.1/71857.

Dancer, D., Rammohan, A. & Smith, M. D. (2008). Infant mortality and child nutrition in Bangladesh. *Health Econ*omics, 17(9), 1015–35. doi:10.1002/hec.1379.

Davidson, R., & Duclos, J. Y. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, *68*(6), 1435-1464.

Deaton, A. (1997). *The analysis of household surveys: A microeconomic approach to development policy*. Baltimore: The John Hopkins University Press.

Duclos, J. & Araar, A. (2006). *Poverty and equity: Measurement, policy and estimation with DAD* (1st ed.). Ottawa: Springer.

Duclos, J.Y. & Verdier-Chouchane, A. (2010). *Analyzing Pro-Poor Growth in Southern Africa: Lessons from Mauritius and South Africa* (Working Papers Series No. 115). African Development Bank, Tunis, Tunisia.

Ewoudou, J. & Vencatachellum, D. (2006). *An Empirical Analysis of the Rates of Returns to Education in Cameroon.* Paper presented at the Centre for the Study of African Economies Conference, University of Oxford, Oxford.

Food and Agriculture Organisation. (2011). The state of food and agriculture: Women in agriculture- closing the gender gap for development. Rome, Italy: FAO. Retrieved from http://www.fao.org/docrep/013/i2050e/i2050e.pdf.

Foster, J., Greer, J., & Thorbecke, E. (1984). A Class of Decomposable Poverty Measures. *Econometrica, 52*(3), 761-66. doi:1. Retrieved from http://www.jstor.org/stable/1913475 doi:1

Gaddis, I. & Klasen, S. (2014). "Economic development, structural change, and women's labour force participation:," *Journal of Population Economics*, Springer, 27(3), 639-81.

Geary, R. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician,* 5 (3), 115–45.

Government of Malawi. (2007). Malawi Growth and Development Strategy I: From Poverty to Prosperity (2006-2011). Lilongwe, Ministry of Economic Planning and Development. Retrieved from https://www.imf.org/external/pubs/ft/scr/2007/cr0755.pdf.

Government of Malawi. (2012). Malawi Growth and Development Strategy II (2011-2016). Lilongwe, Ministry of Finance and Development Planning. Retrieved from https://www.imf.org/external/pubs/ft/scr/2012/cr12222.pdf.

Gondwe, A.S. (2011). *Spatial Comparisons of Multidimensional Poverty in Malawi* (Unpublished Master's thesis). Chancellor College, University of Malawi, Malawi.

Greenacre, M. (2007). *Correspondence analysis in practice*. CRC press.

Greenwood, M. J. (2005). Modeling migration. *Encyclopedia of social measurement*, *2*, 725-734.

Grosse, M., Harttgen, K., & Klasen, S. (2008). Measuring pro-poor growth in non-income dimensions. *World Development*, *36*(6), 1021-1047.

Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. *The American economic review*, 126-142.

Haughton, J. & Khandker, S.R. (2009). *Handbook on poverty and inequality*. Washington: The World Bank.

Heckman, J. (1979). Sample Selection as a Specification Error. *Econometrica,* 47 (1), 153-61.

Human Development Report. (2015. Work For Human Development: Retrieved from http://hdr.undp.org/sites/default/files/2015_human_development_report.pdf.

Hussmanns, R. (2004, February). Statistical definition of informal employment: Guidelines endorsed by the Seventeenth International Conference of Labour Statisticians (2003). In *7th Meeting of the Expert Group on Informal Sector Statistics (Delhi Group)* (pp. 2-4).

International Labour Organization. (2011). ILO estimates and projections of the economically active population: 1990–2020 (6th ed.). Retrieved from http://laborsta.ilo.org/applv8/data/EAPEP/eapep_E.html.

Jeanty, P.W. (2010). spmlreg: Stata module to estimate the Spatial lag, the Spatial error, the Spatial Durbin, and the General Spatial models.

Jolliffe, D. (2002). Whose education matters in the determination of household income? Evidence from a developing country. *Economic Development and Cultural Change*, *50*(2), 287-312.

Kahyarara, G. & Teal, F. (2008).The Returns to Vocational Training and Academic Education: Evidence from Tanzania. *World Development*, Elsevier, 36(11), 2223-42.

Kakwani, N. (1980). On a class of poverty measures. *Econometrica: Journal of the Econometric Society*, 437-446.

Kakwani, N., & Pernia, E. M. (2000). What is pro-poor growth? *Asian development review*, *18*(1), 1-16.

Kerr, A. & Teal, F.(2015). The Determinants of Earnings Inequalities: Panel data evidence from KwaZulu-Natal, South Africa. *Journal of African Economies*, 24(4), 530-58.

Khandker, S.R., Koolwal, G.B. & Samad, H.A. (2010) *Handbook on impact evaluation:* quantitative *methods and practices.* Washington, DC: The World Bank.

Kingdon, G. & Knight, J. (2004). Unemployment in South Africa: The Nature of the Beast. *World Development*, 32 (3), 391-408.

Lassibille, G. & Tan, J. (2005).The Returns to Education in Rwanda. *Journal of African Economies*, 14 (1), 92-116.

Linnemayr, S., Alderman, H. & Ka, A. (2008). Determinants of malnutrition in Senegal: Individual, household, community variables, and their interaction. *Economics & Human Biology*, 6 (2), 252-63.

Matita, M. & Chirwa, E. (2009). *The Impact of Education on Self-Employment, Farm Activities and Household Incomes in Malawi.* (Department of Economics Working Paper No. 2009/02). University of Malawi, Chancellor College.

McKay, A. (2000). Should the survey measure total household income? In Grosh, M. and Glewwe, P. (Eds.), *Designing household survey questionnaire for developing countries: Lessons from 15 years of the living standards measurement study (*pp. 83 – 104). Washington: The World Bank.

Mincer, J. (1974). *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research. Retrieved from http://papers.nber.org/books/minc74-1.

Minnesota Population Center. (2015). Integrated Public Use Microdata Series, International: Version 6.4 [Machine-readable database]. Minneapolis: University of Minnesota. Retrieved from https://international.ipums.org/international-action/variables/group.

Moran, P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society*, Series B (Methodological), 10 (2), 243–51.

Murkherjee, S. & Benson, T. (2003). The determinants of poverty in Malawi, 1998. *World Development*, 31(2), 339-358.

Mussa, R. (2013). Spatial Comparisons of Poverty and Inequality in Living Standards in Malawi. *South African Journal of Economics*, 81(2), 192-210.

Mussa, R. (2014). A matching decomposition of the rural–urban difference in malnutrition in Malawi. *Health Economics Review*, Springer, 4(1), 1-10.

Mussa, R. (2014). *Externalities of Education on Efficiency and Production Uncertainty of Maize in Rural Malawi* (MPRA Paper 54628). University Library of Munich, Germany.

National Statistical Office. (2008). *Population and Housing Census Report.* Zomba, Malawi. Retrieved from http://www.nsomalawi.mw/images/stories/data_on_line/demography/census_2008/Main%20Report/ThematicReports/Migration%20Report.pdf.

National Statistical Office. (2012). *Integrated Household Survey 2010-2011: Household Socio-Economic Report.* Zomba, Malawi. Retrieved from http://www.nsomalawi.mw/images/stories/data_on_line/economics/ihs/IHS3/IHS3_Report.pdf.

National Statistical Office. (2012). *Malawi Labour Force Survey Report 2013.* Zomba, Malawi. Retrieved from http://www.nsomalawi.mw/component/content/article/8/209-malawi-labour-force-survey-2013.html.

National Statistical Office. (2014). *Integrated Household panel Survey 2010-13: Household Socio-Economic Characteristics Report.* Zomba, Malawi. Retrieved from http://www.nsomalawi.mw/images/stories/data_on_line/economics/ihs/IHPS%202013/IHPS%20Report.pdf.

National Statistical Office. (2015). Welfare Monitoring Survey 2014. Retried from http://www.nsomalawi.mw/images/stories/data_on_line/agriculture/wms_2014/Welfare%20Monitoring%20Survey%202014.pdf.

National Statistical Office & ICF Macro. (2011). *Malawi Demographic and Health Survey 2010*. Zomba, Malawi, and Calverton, Maryland, USA. Retrieved from http://www.dhsprogram.com/pubs/pdf/FR247/FR247.pdf.

Organisation for Economic Co-operation and Development. (2008). *2008 Survey on Monitoring the Paris Declaration: Making Aid More Effective by 2010.* OECD Publishing, Paris. doi: 10.1787/9789264050839-en.

Pisati, M. (2001). Tools for Spatial Data Analysis. *Stata Technical Bulletin*. March 2001. STB-60:21-36. Retrieved from http://www.stata.com/products/stb/journals/stb60.pdf.

Pisati, M. (2012). *Spatial Data Analysis in Stata: An Overview*. Department of Sociology and Social Research. University of Milano-Bicocca (Italy). Retrieved from http://www.stata.com/meeting/italy12/abstracts/materials/it12_pisati.pdf.

Pischke, J. S., & Schwandt, H. (2014). *Poorly Measured Confounders are More Useful on the Left Than on the Right* (Working Paper). London School of Economics.

Psacharopoulos, G. (1994). Returns to Investment in Education: A Global Update, World *Development*, 22 (9), 1325-43.

Ravallion, M. (2011). On multidimensional indices of poverty. *The Journal of Economic Inequality*, *9*(2), 235-248.

Ravallion, M., & Chen, S. (2003). Measuring pro-poor growth. *Economics letters*, *78*(1), 93-99.

Reserve Bank of Malawi. (2015). Financial and Economic Review, 49 (4). Retrieved from https://www.rbm.mw/Publications/EconomicReviews/.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, *3*(2), 135-146.

Rutstein, S.O. & Johnson, K. (2004). *The DHS Wealth Index* (DHS Comparative Reports No. 6). Calverton, Maryland: ORC Macro.

Sahn, D.E. & Stifel, D.C., 2003. Urban-rural Inequality in Living Standards in Africa. *Journal of African Economies*, 12 (4), 564-97.

Sen, A. (1976). Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, 219-231.

Sen, A. (1985). Commodities and Capabilities. *North Holland, Amsterdam*.

Sen, A. (1987). The Standard of living. In McMurrin, S. (Ed), *Tanner lectures on Human Values*, Cambridge: Cambridge University Press. Retrieved from http://tannerlectures.utah.edu/_documents/a-to-z/s/sen86.pdf.

Singer, H., & Jolly, R. (1972). Employment, Incomes and Equality: A Strategy for Increasing Productive Employment in Kenya. *Geneva: ILO*.

Sokal, R. R., Oden, N. L. & Thomson, B. A. (1998), Local Spatial Autocorrelation in a Biological Model. Geographical Analysis, 30: 331–54. doi: 10.1111/j.1538-4632.1998.tb00406.x

StataCorp. (Version 13.1). Stata Statistical Software [Computer Software]. College Station, Texas: Stata Corporation.

StataCorp. (Version 14.1). Stata Statistical Software [Computer Software]. College Station, Texas: Stata Corporation.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography, 46*, 234-40. doi:1. Retrieved from http://www.jstor.org/stable/143141 doi:1

Verduzco-Gallo, I., Ecker, O., & Pauw, K. (2014). *Changes in Food and Nutrition Security in Malawi*. IFPRI Working Paper 06.

Villa, J.M. (2011). *DIFF: Stata Module to Perform Differences in Differences Estimation*. Statistical Software Components. Boston College Department of Economics.

Willekens, F. (2008). Models of migration: Observations and judgement. *International migration in Europe: Data, models and estimates*, 117-147.

Williams, R. (2015). Panel Data 4: Fixed Effects vs. Random Effects Models. University of Notre Dame. Retrieved from https://www3.nd.edu/~rwilliam/stats3/Panel04-FixedVsRandom.pdf.

Wittenberg, M. (2014). *Wages and wage inequality in South Africa 1994-2011: The evidence from household survey data.* (SALDRU working paper No. 135). Southern Africa Labour and Development Research Unit, University of Cape Town.

Wittenberg, M., Leibbrandt, M. (2015). *Measuring Inequality by Asset Indices: A general approach with application to South Africa*. (SALDRU Working Paper No. 141). Southern Africa Labour and Development Research Unit, University of Cape Town.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: The MIT Press.

World Bank. (2004). *Poverty reduction strategies papers sourcebook.* Washington, D.C.: World Bank.

World Bank. (2012). Malawi - Community-Based Rural Land Development Project. Retrieved from http://documents.worldbank.org/curated/en/2012/03/16215843/malawi-community-based-rural-land-development-project.

World Health Organisation. (2006). WHO Child Growth Standards: Methods and development. Retrieved from http://www.who.int/childgrowth/standards/Technical_report.pdf.

Yu, K.C.D. (2012). *Using household surveys for deriving labour market, poverty and inequality trends in South Africa* (Doctoral thesis, Stellenbosch University, South Africa). Retrieved from http://scholar.sun.ac.za/handle/10019.1/71638.

Appendices

Appendix A1: Coefficient comparison test

The coefficient comparison was achieved by seemingly unrelated regression (SURE) which takes the following code form in Stata 13.1:

```
.svy: regress lwage edu x if Male
.estimates store Male
.svy: regress lwage edu x if! Male
.estimates store Female
.suest Male Female
.test_b [Male:edu]=_b[Female: edu]
.test_b [Male:x]=_b[Female: edu]
```

Since *suest* (and its alternatives) do not work with *xtreg*, we conduct the coefficient comparison tests for pooled Ordinary Least Squares (OLS) only.

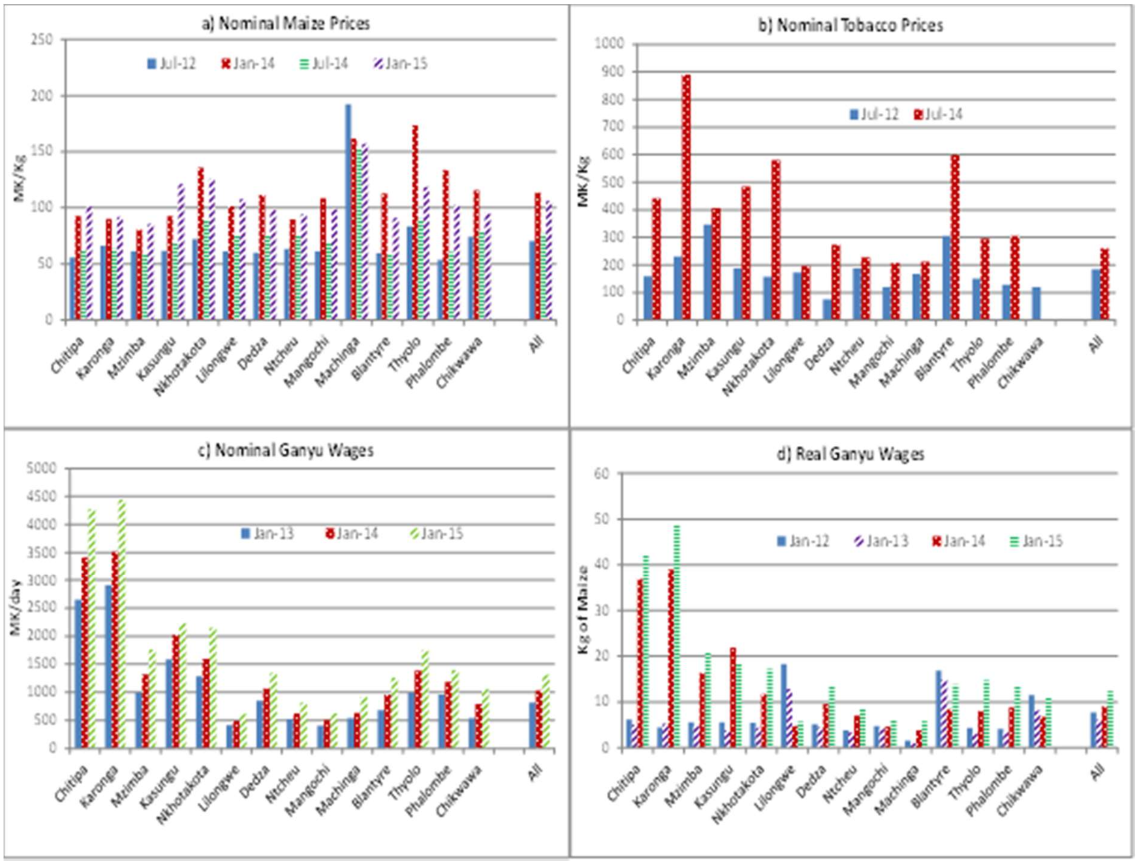
Appendix A2: Dealing with data inconsistencies

In order to ensure data quality, we created a salary payment period that is consistent across the waves using the following Stata code (in Stata 13.1) for the formal and informal sectors, respectively:

```
.bys PID: egen period1=max (salary_period)
.bys PID: egen period2=max (ganyu_dys)
```

It does not matter whether we use maximum (max) or minimum (min). The most important thing is to come up with a period that is consistent across waves. Otherwise, the data will show that wages increased while in actual fact it is as a result of changes in the period of payment.

Figure A 1: FISP evaluation panel data on real ganyu wages



Source: Farm Input Subsidy Programme (FISP) evaluation panel data

Table A 1: Table showing the distribution of households by year and population subgroups

	2010		2004		2000		1992	
Description	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Areas								
Urban	4,116	16.6	2,262	16.6	1,949	13.7	603	11.3
Rural	20,709	83.4	11,402	83.5	12,264	86.3	4,720	88.7
Region								
Northern	2,716	10.9	1,584	11.6	1,496	10.5	589	11.1
Central	10,627	42.8	5,589	40.9	5,744	40.4	2,043	38.4
Southern	11,482	46.3	6,491	47.5	6,973	49.1	2,691	50.6
Residence								
Rural North	2,476	10.0	1,325	9.7	1,215	8.6	529	9.9
Rural Centre	8,876	35.8	4,615	33.8	5,049	35.5	1,822	34.2
Rural South	9,356	37.7	5,462	40.0	6,000	42.2	2,369	44.5
Urban North	240	1.0	259	1.9	281	2.0	60	1.1
Urban Centre	1,751	7.1	974	7.1	695	4.9	221	4.2
Urban South	2,126	8.6	1,029	7.5	973	6.9	322	6.1
Sex								
Male head	17,857	71.9	10,295	75.3	10,431	73.4	4,013	75.4
Female head	6,968	28.1	3,369	24.7	3,782	26.6	1,310	24.6
Total	24,825	100	13,664	100	14,213	100	5,323	100

Table A 2: Table showing the distribution of children by year and population subgroups

	2010		2004		2000		1992	
Description	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Age								
0-23	2,050	42.7	4,083	46.9	4,497	46.11	1,591	47.45
24-59	2,751	57.3	4,624	53.1	5,256	53.89	1,762	52.55
Sex								
Male	2,331	48.6	4,355	50.0	4,821	49.4	1,676	50.0
Female	2,470	51.4	4,352	50.0	4,932	50.6	1,677	50.0
Areas								
Urban	721	15.0	1,129	13.0	1,272	13.0	360	10.7
Rural	4,080	85.0	7,578	87.0	8,481	87.0	2,993	89.3
Region								
Northern	514	10.7	1,174	13.5	1,064	10.9	399	11.9
Central	2,226	46.4	3,417	39.3	4,229	43.4	1,387	41.4
Southern	2,061	42.9	4,116	47.3	4,460	45.7	1,567	46.7
Residence								
Rural North	460	9.6	1,001	11.5	855	8.8	354	10.6
Rural Centre	1,903	39.6	3,024	34.7	3,726	38.2	1,235	36.8
Rural South	1,717	35.8	3,553	40.8	3,899	40.0	1,404	41.9
Urban North	54	1.1	173	2.0	209	2.1	45	1.3
Urban Centre	323	6.7	393	4.5	502	5.2	152	4.5
Urban South	344	7.2	563	6.5	561	5.8	163	4.9
Total	4,801	100	8,707	100	9,753	100	3,353	100

Table A 3: WHZ regression results by age category

	Age<24 months		Age>=24 months	
Variables	Coeff	SE	Coeff	SE
Age in months	-0.080**	(0.029)	0.038	(0.025)
Square of age	0.286*	(0.114)	-0.050	(0.029)
Female child	0.036	(0.087)	-0.162**	(0.054)
Child is twin	-0.417	(0.250)	-0.101	(0.163)
Birth order number	0.045	(0.027)	0.032*	(0.013)
Mother's education				
Incomplete primary	0.200	(0.146)	0.059	(0.093)
Complete primary	-0.009	(0.209)	0.186	(0.126)
Incomplete secondary	0.173	(0.201)	0.144	(0.131)
Complete secondary	0.480	(0.252)	0.257	(0.201)
Post-secondary	-0.089	(0.763)	0.012	(0.370)
Father's education				
Incomplete primary	-0.052	(0.164)	-0.028	(0.099)
Complete primary	0.084	(0.218)	0.096	(0.135)
Incomplete secondary	-0.141	(0.190)	-0.028	(0.113)
Post-secondary	0.476	(0.317)	-0.331	(0.288)
Asset index	-0.002	(0.086)	0.000	(0.060)
Square of asset index	0.007	(0.023)	0.009	(0.016)
Rural area	-0.163	(0.142)	-0.042	(0.109)
Region				
Central	-0.162	(0.131)	-0.044	(0.094)
Southern	-0.215	(0.124)	-0.142	(0.094)
Constant	0.652	(0.358)	-0.127	(0.517)
R-squared	0.026		0.017	
Prob > F	0.004		0.097	
Observations	1,863		2,668	

Notes: *, **, *** denote significance at 10%, 5% and 1% levels

Table A 4: List of regions, districts and traditional areas consistent over time

Region	District	Traditional area
Northern Region	Chitipa	Chitipa Boma, Kameme, Mwabulambya
		Mwenemisuku
		Nyika National Park, Mwenewenya
	Karonga	SC Mwakaboko, Silupula
		Karonga, Kyungu
		Wasambo
		Karonga Boma
		SC Mwirang'ombe
	Nkhata Bay, Likoma	Nkhata Bay Boma, Mkumpa, SC Mkumbira, Mankhambira
		Malenga Mzoma, SC Fukamalaza, SC Malanda
		Kabunduli
		Timbiri
		Fukamapiri, SC Zilakoma
		Boghoyo, SC Mkondowe, SC Nyaluwanga, Musisya
	Rumphi	Rumphi Boma, SC Mwahenga, Katumbi, SC Zolokere, Vwaza Marsh Game, Chikulamayembe
		SC Chapinduka, Mwamlowe, SC Mwankhunikira
		SC Kachulu, SC Mwalweni
	Mzimba	SC Jaravikuba Munthali, Mtwalo
		Chindi
		M'mbelwa
		Mzikubola
		Mabulabo
		SC Kampingo Sibande
		Vwaza Marsh Game Reserve, Mpherembe
		SC Khosolo Gwaza
		Mzukuzuku
		Mzimba Boma, Mzuzu city
Central Region	Kasungu	Wimbe
		Santhe
		Chulu
		SC Chisikwa, Kaluluma
		SC Kawamba

		Kaomba and Kasungu national park Kasungu Township SC Njombwa SC Chilowamatambe Kapelula SC M'nyanja SC Lukwa SC Simlemba
	Nkhota Kota	SC Kafulazira, Kanyenda Mwadzama Nkhota Kota Game Reserve, Malenga Chanzi SC Mphonde Nkhota Kota Boma Mwansambo
	Ntchisi	Ntchisi Boma, Kalumo SC Chilooko Kasakula, SC Nthondo Chikho
	Dowa	SC Chakhaza SC Kayembe Chiwere Dowa Boma, Mkukula Dzoole Msakambewa Mponela Urban, SC Mponela
	Salima	SC Msosa, Khombedza, Kuluunda Chipoka Urban, SC Kambalame, Ndindi Maganga Salima Township SC Kambwiri SC Mwanza Pemba
	Lilongwe	Chiseka Kalolo Kabudula Chadza

		Khongoni Mazengera Malili Chimutu Kalumbu SC Mtema SC Chitekwele SC Njewa Tsabango Chitukula Kalumba Lilongwe city
	Mchinji	Mchinji Boma, Zulu SC Mavwere Mkanda SC Mduwa SC Dambe Mlonyeni
	Dedza	Kaphuka Pemba Kachindamoto Kasumbu Tambala SC Chilikumwendo SC Kamenya Gwaza Dedza Township SC Chauma
	Ntcheu	Goodson Ganya Champiti, Makwangwala Ntcheu Boma, Kwataine Phambala Njolomole Mpando Chakhumbira Masasa
Southern Region	Mangochi	Palm Beach forest, Monkey Bay urban

		Chimwala
		Mponda
		Chowe
		SC Mbwana Nyambi
		Jalasi
		Makanjila
		Katuli
		Mangochi Boma
		SC Namabvi
	Machinga	Kawinga
		Liwonde National Park, Liwonde, Lake Malawi national park
		SC Chikweo
		Nyambi and SC Chiwalo
		SC Mlomba
		Machinga Boma, Sitola, Liwonde Township
		SC Ngokwe
		SC Mposa
		Chamba
	Zomba	SC Mbiza
		Mlumbe
		Mwambo
		Kuntumanji, Mkumbira
		Malemia
		Chikowi
		Zomba municipality
	Chiradzulu	Chiradzulu Boma, Chitera, Mpama
		Kadewere
		Likoswe
		Nkalo
		Mchema
	Blantyre	Kapeni
		Kuntaja
		Somba
		Chigeru
		Makata, Machinjili
		Kunthembwe

		Lundu Blantyre city Mwanza Mwanza Boma, Nthache Kanduku Symon Mlauli Dambe Ngozi
	Thyolo	Luchenza Township, Chimaliro Bvumbwe Thyolo Boma, Nchilamwela Kapichi SC Thukuta, Nsabwe SC Mphuka SC Kwethemule SC Mbawela Changata Thomas
	Mulanje	Mabuka Nkanda, Mulanje Mountain Mulanje Boma, Chikumbu Juma Laston Njema Nthiramanja
	Phalombe	Phalombe Boma, Mkhumba Nazombe
	Chikwawa	 Ngabu Urban, Ngabu Chapananga, Lengwe National Park Mankhwira Lundu Chikwawa Boma, Katunga Kasisi, Majete Game Reserve Maseya
	Nsanje	Mlolo

	Malemia, Tengani, Mwabvi Game Reserve
	SC Mbenje
	SC Makoko, Nyachikadza, Ndamera
	Chimombo, Ngabu
	Nsanje Boma
Balaka	Msamala and Balaka town
	Kalembo

Source: Own compilation from census data

Table A 5: Results showing spatial dependencies in variables for 1987

Variables	Moran's I	SD	p-value	Geary's c	SD	p-value
Migration	0.010	0.005	0.004***	0.998	0.017	0.909
Employed	0.015	0.005	0.000***	0.996	0.015	0.792
Years of schooling	0.100	0.005	0.000***	0.866	0.007	0.000***
Sex	0.000	0.005	0.540	0.997	0.005	0.540
Age	0.060	0.005	0.000***	0.935	0.009	0.000***
Married	0.081	0.005	0.000***	0.899	0.007	0.000***
Dependency ratio	0.039	0.005	0.000***	0.966	0.008	0.000***

Note: *, **, *** denote significance at 10%, 5% and 1% levels

Table A 6: Results showing spatial dependencies in variables for 1998

Variables	Moran's I	SD	p-value	Geary's c	SD	p-value
Migration	.	.	.	.	.	.
Employed	0.059	0.005	0.000***	0.924	0.013	0.000***
Years of schooling	0.126	0.005	0.000***	0.841	0.007	0.000***
Sex	0.000	0.005	0.540	0.997	0.005	0.540
Age	0.043	0.005	0.000***	0.971	0.01	0.003***
Married	0.069	0.005	0.000***	0.913	0.007	0.000***
Dependency ratio	0.033	0.005	0.000***	0.960	0.008	0.000***

Note: *, **, *** denote significance at 10%, 5% and 1% levels

Table A 7: Results showing spatial dependencies in variables for 2008

Variables	Moran's I	SD	p-value	Geary's c	SD	p-value
Migration	0.420	0.005	0.000***	0.591	0.006	0.000***
Employed	0.025	0.005	0.000***	0.983	0.007	0.017***
Years of schooling	0.107	0.005	0.000***	0.870	0.008	0.000***
Sex	0.000	0.005	0.540	0.997	0.005	0.540
Age	0.010	0.005	0.005***	1.002	0.011	0.890
Married	0.043	0.005	0.000***	0.928	0.008	0.000***
Dependency ratio	0.074	0.005	0.000***	0.912	0.009	0.000***

Note: *, **, *** denote significance at 10%, 5% and 1% levels

Table A 8: Population figures for individuals aged (15 years and older)

	1987		1998		2008	
District	Male	Female	Male	Female	Male	Female
Chitipa	20,100	24,450	29,740	35,780	42,900	47,600
Karonga	32,580	37,540	49,130	56,430	66,590	73,190
Nkhata Bay, Likoma	31,780	35,390	45,300	49,980	56,980	61,000
Rumphi	22,100	24,750	33,720	35,840	45,120	46,780
Mzimba	108,440	117,170	161,690	173,590	228,090	244,440
Kasungu	89,990	78,670	139,600	125,510	170,160	164,410
Nkhota Kota	41,230	39,810	63,140	61,800	79,230	79,550
Ntchisi	30,030	31,700	44,960	46,330	63,000	63,740
Dowa	81,260	86,150	111,530	114,430	152,060	158,700
Salima	47,140	52,590	67,400	71,110	89,810	93,730
Lilongwe	246,180	254,090	381,780	371,920	540,870	534,880
Mchinji	63,850	56,860	87,940	86,530	121,150	119,380
Dedza	83,060	108,420	120,660	142,510	159,680	182,870
Ntcheu	64,030	87,060	92,040	110,580	123,020	140,760
Mangochi	103,950	127,900	158,050	179,370	194,000	218,630
Machinga	110,070	140,180	160,250	186,870	196,960	223,400
Zomba	105,430	128,100	151,760	164,380	174,900	192,770
Chiradzulu	48,330	64,420	62,160	76,890	74,070	88,790
Blantyre	168,480	145,910	252,610	229,830	299,600	288,050
Thyolo	102,140	118,680	120,750	141,180	150,130	171,600
Mulanje	132,890	177,440	170,600	210,110	206,610	244,320
Chikwawa	68,390	70,110	95,930	98,280	117,710	117,230
Nsanje	33,870	39,900	49,520	55,770	62,890	64,510
Neno	23,600	29,200	35,320	38,970	50,140	54,810
Total	1,858,920	2,076,490	2,685,580	2,863,990	3,465,670	3,675,140

Source: Own computation from census data

Table A 9: Number of people who migrated to other districts (15 years and older)

	1987			2008		
	Male	Female	Total	Male	Female	Total
Chitipa	206	281	487	61	64	125
Karonga	375	411	786	182	185	367
Nkhata Bay, Likoma	434	418	852	296	213	509
Rumphi	453	460	913	223	199	422
Mzimba	1,914	1,877	3,791	1,153	1,036	2,189
Kasungu	2,040	1,438	3,478	792	664	1,456
Nkhota Kota	955	814	1769	244	212	456
Ntchisi	464	410	874	164	146	310
Dowa	1,198	1,105	2,303	275	286	561
Salima	956	757	1713	306	270	576
Lilongwe	3,751	3,236	6,987	3,201	2,590	5,791
Mchinji	1,203	903	2106	295	297	592
Dedza	1,129	897	2026	250	228	478
Ntcheu	637	668	1305	281	265	546
Mangochi	1,817	1,672	3,489	705	659	1,364
Machinga	2,098	1,888	3,986	713	662	1,375
Zomba	1,800	1,503	3,303	931	836	1,767
Chiradzulu	593	595	1188	259	276	535
Blantyre	4,457	3,101	7,558	4,428	3,805	8,233
Thyolo	1,626	1,522	3,148	733	676	1,409
Mulanje	1,462	1,310	2,772	685	646	1,331
Chikwawa	862	687	1549	563	527	1,090
Nsanje	328	325	653	257	264	521
Neno	252	283	535	346	327	673
Total	31,010	26,561	57,571	17,343	15,333	32,676

Source: Own computation from census data

Table A 10: Proportions of people who migrated to other districts (15 years and older)

	1987			2008		
	Male	Female	Total	Male	Female	Total
Chitipa	10.2%	11.5%	10.9%	1.4%	1.3%	1.4%
Karonga	11.5%	10.9%	11.2%	2.7%	2.5%	2.6%
Nkhata Bay, Likoma	13.7%	11.8%	12.7%	5.2%	3.5%	4.3%
Rumphi	20.5%	18.6%	19.5%	4.9%	4.3%	4.6%
Mzimba	17.7%	16.0%	16.8%	5.1%	4.2%	4.6%
Kasungu	22.7%	18.3%	20.6%	4.7%	4.0%	4.4%
Nkhota Kota	23.2%	20.4%	21.8%	3.1%	2.7%	2.9%
Ntchisi	15.5%	12.9%	14.2%	2.6%	2.3%	2.5%
Dowa	14.7%	12.8%	13.8%	1.8%	1.8%	1.8%
Salima	20.3%	14.4%	17.2%	3.4%	2.9%	3.1%
Lilongwe	15.2%	12.7%	14.0%	5.9%	4.8%	5.4%
Mchinji	18.8%	15.9%	17.5%	2.4%	2.5%	2.5%
Dedza	13.6%	8.3%	10.6%	1.6%	1.3%	1.4%
Ntcheu	9.9%	7.7%	8.6%	2.3%	1.9%	2.1%
Mangochi	17.5%	13.1%	15.1%	3.6%	3.0%	3.3%
Machinga	19.1%	13.5%	15.9%	3.6%	3.0%	3.3%
Zomba	17.1%	11.7%	14.1%	5.3%	4.3%	4.8%
Chiradzulu	12.3%	9.2%	10.5%	3.5%	3.1%	3.3%
Blantyre	26.5%	21.3%	24.0%	14.8%	13.2%	14.0%
Thyolo	15.9%	12.8%	14.3%	4.9%	3.9%	4.4%
Mulanje	11.0%	7.4%	8.9%	3.3%	2.6%	3.0%
Chikwawa	12.6%	9.8%	11.2%	4.8%	4.5%	4.6%
Nsanje	9.7%	8.1%	8.9%	4.1%	4.1%	4.1%
Neno	10.7%	9.7%	10.1%	6.9%	6.0%	6.4%
Total	15.8%	12.9%	14.3%	4.2%	3.7%	3.9%

Source: Own computation from census data

Note: As earlier stated, four of the districts in Malawi consist of cities, namely Mzuzu city (in Mzimba district), Lilongwe city (in Lilongwe district), Zomba municipality or city (in Zomba district) and Blantyre city (in Blantyre district). A movement from the rural areas of the district into the city (e.g. from Blantyre rural to Blantyre city) and vice versa is recorded as migration. As a result, the high migration patterns observed in Blantyre and Lilongwe (which are the two major cities in Malawi) could largely be rural-urban migration involving people moving rural areas to urban areas within the district boundaries.

Table A 11: Effects of land reform policy on migration

	Male		Female	
	SEM	SAC	SEM	SAC
Time	-0.170***	-0.102***	-0.028	-0.016
	(0.025)	(0.038)	(0.023)	(0.039)
DID	0.015	0.015	0.006	0.017
	(0.016)	(0.016)	(0.017)	(0.015)
Schooling	-0.013	-0.013	-0.037***	-0.029***
	(0.010)	(0.010)	(0.010)	(0.010)
Age	0.088	0.088	-0.179***	-0.151***
	(0.056)	(0.056)	(0.051)	(0.052)
Age squared	-0.002**	-0.002**	0.002***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Married	0.384***	0.382***	0.074	0.076
	(0.125)	(0.123)	(0.116)	(0.110)
Employed	-0.084**	-0.085**	-0.103***	-0.109***
	(0.039)	(0.038)	(0.023)	(0.023)
Assets	0.033***	0.033***	-0.012	-0.008
	(0.010)	(0.010)	(0.010)	(0.010)
W*migration		0.564**		0.447
		(0.260)		(0.284)
Lambda	0.755***	0.528*	2.024***	0.789***
	(0.155)	(0.302)	(0.145)	(0.152)
Observations	356	356	356	356
Spatial units	178	178	178	178

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; standard errors in parenthesis

Table A 12: Effects of land reform policy on agricultural employment

	Male		Female	
	SEM	SAC	SEM	SAC
Time	-0.387***	-0.352**	-0.436***	-0.324***
	(0.033)	(0.137)	(0.043)	(0.123)
DID	0.028	0.027	0.053**	0.050**
	(0.021)	(0.021)	(0.021)	(0.021)
Schooling	0.074***	0.073***	0.054*	0.054*
	(0.024)	(0.024)	(0.028)	(0.028)
Age	0.286	0.279	1.018***	1.031***
	(0.542)	(0.543)	(0.341)	(0.343)
Age squared	-0.004	-0.004	-0.016***	-0.017***
	(0.009)	(0.009)	(0.006)	(0.006)
Married	-0.156	-0.153	-0.166	-0.151
	(0.309)	(0.309)	(0.271)	(0.269)
Dependency	-0.268*	-0.265*	0.096	0.101
	(0.142)	(0.142)	(0.113)	(0.111)
W*employed		0.113		0.342
		(0.438)		(0.352)
Lambda	0.334	0.275	0.611***	0.479
	(0.319)	(0.412)	(0.218)	(0.309)
Observations	356	356	356	356
Spatial units	178	178	178	178

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; standard errors in parenthesis

Table A 13: Effects of land reform policy on government employment

	Male		Female	
	SEM	SAC	SEM	SAC
Time	0.079**	-0.002	-0.011	-0.024
	(0.038)	(0.021)	(0.012)	(0.017)
DID	0.028***	0.024***	0.013**	0.012**
	(0.007)	(0.007)	(0.006)	(0.005)
Schooling	0.007	0.011	0.033***	0.033***
	(0.008)	(0.007)	(0.007)	(0.007)
Age	-0.277*	-0.264*	0.432***	0.436***
	(0.162)	(0.159)	(0.087)	(0.088)
Age squared	0.005*	0.004*	-0.007***	-0.007***
	(0.003)	(0.003)	(0.001)	(0.001)
Married	-0.102	-0.113	0.111	0.110
	(0.092)	(0.089)	(0.070)	(0.070)
Dependency	0.021	0.016	0.042	0.041
	(0.044)	(0.043)	(0.029)	(0.029)
W*employed		0.799***		0.324
		(0.130)		(0.369)
Lambda	0.922***	0.830***	0.707***	0.582**
	(0.054)	(0.115)	(0.174)	(0.288)
Observations	356	356	356	356
Spatial units	178	178	178	178

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; standard errors in parenthesis

Table A 14: Effects of land reform policy on self, wage and private employment (OLS only)

	Male		Female	
	Self	Private	Self	Private
Time	0.228***	0.185***	0.213***	0.108***
	(0.020)	(0.018)	(0.019)	(0.013)
Treated	-0.018	-0.007	0.002	-0.014
	(0.015)	(0.014)	(0.015)	(0.010)
DID	-0.043**	-0.040**	-0.021	-0.041***
	(0.021)	(0.019)	(0.020)	(0.014)
Schooling	-0.001	0.020***	0.008*	0.011***
	(0.006)	(0.005)	(0.004)	(0.003)
Age	-1.436***	-0.298	-0.691***	-0.106
	(0.310)	(0.284)	(0.161)	(0.110)
Age squared	0.023***	0.005	0.011***	0.002
	(0.005)	(0.005)	(0.003)	(0.002)
Married	-0.274*	-0.047	-0.241**	0.090
	(0.159)	(0.146)	(0.099)	(0.068)
Dependency	-0.134	-0.324***	-0.030	-0.129**
	(0.106)	(0.097)	(0.080)	(0.055)
Constant	22.764***	4.543	10.835***	1.62
	(4.826)	(4.416)	(2.428)	(1.658)
R-squared	0.580	0.572	0.657	0.428
Observations	356	356	356	356

Notes: *, **, *** denote significance at 10%, 5% and 1% levels; standard errors in parenthesis