The Extension of Existing Data and Methods to Measure Poverty and Mobility in Data-poor, Agrarian Sub-Saharan Africa

Thomas Ferreira

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

Supervisor: Prof Servaas van der Berg
Co-supervisor: Dr Dieter von Fintel

December 2017

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.

Abstract

In sub-Saharan Africa (SSA), poverty rates have declined, but the absolute number of poor has increased (Beegle et al., 2016). The poor are disproportionately found in rural households which practise smallholder agriculture. Poor data availability and quality have no doubt curbed poverty alleviation efforts. Smallholder agriculture is hard to measure and the statistical capacity of many developing countries in SSA is low (Carletto et al., 2015). In this environment, this thesis shows through three separate studies, how existing data and methods can be extended to overcome usual data deficiencies in SSA and enhance knowledge on welfare.

In the first study, econometric techniques used in analysis of the formal labour market are extended to understand the educational returns in agricultural productivity in Malawi. Poverty in Malawi fits the typical pattern in SSA. The majority of the poor live and work in smallholder agriculture. In settings like these, schooling is believed to be a valuable tool in lifting people out of poverty. Yet, little is known about how schooling affects agricultural productivity. The effect of education on smallholder agricultural production has been estimated for countries in SSA but no studies deal with the endogenous nature of education in the production process. This chapter contributes to the literature by estimating for the first time, the causal effects of education on agricultural productivity in SSA. An instrumental variable (IV) approach is used, using the introduction of free primary education (FPE) and the age of paternal orphanhood as IV's for education. The instruments are shown to calculate local average treatment effects for individuals who only entered school due to FPE and only left school due to paternal orphanhood. It is found that there are large differences in returns to education between the subgroups. Returns are low and insignificant when FPE is used as an IV but they are larger and there is a significant effect when age of paternal orphanhood is used. Thus, while education can have large effects on agricultural productivity, this is not so for most of the population, especially individuals specifically targeted by large scale expansions in educational access.

The second study shows how existing satellite data can be used to enhance knowledge of welfare in rural agrarian areas where data on poverty is infrequent and only statistically representative for large geographic areas. Satellite data overcomes both of these limitations as small geographic areas are captured at high frequency. Night lights satellite data have been shown to correlate well with GDP but are not in areas that are not electrified, such as rural, agriculturally based communities in developing countries. This is why this study explores the use of daytime satellite data. Satellite data measuring land use and vegetation quality, have been used to model socio-economic outcomes across regions, but no studies have explored whether daytime satellite data

can be used to track welfare longitudinally. This study argues that indicators of vegetation quality, derived from satellite data, can be used to track welfare over time in agriculturally dominant areas. Such indicators are used extensively to predict agricultural yields and thus should correlate with welfare, as agriculture is an important source of income. This study explores whether this is the case using data from Namibia. Firstly, it is shown using classification of cropland, that daytime satellite data can identify areas of economic activity where night lights cannot. Secondly the relationship between vegetation quality and welfare is studied. Cross-sectionally, increases in vegetation quality correlate negatively with welfare. This is expected as the poor are more likely to live in rural areas. Within rural areas, however, vegetation quality correlates positively with welfare. This shows that vegetation quality can be used to track welfare over time in areas where night lights are not present. It can also do this at higher frequencies and smaller areas than household survey data.

The final study shows how repeated cross-sectional data can be used to enhance knowledge on economic mobility, in a data deficient setting such as Namibia, where panel data does not exist. Namibia has experienced sharp declines in poverty over the last two decades, but in 2009/10 the headcount poverty rate was still 28.7%. Nothing is known, however, about the household dynamics of poverty and poverty reduction over the period as there is no panel data available. To overcome this limitation, this study exploits a method developed by Dang et al. (2014) that uses repeated cross-sections to estimate bounds on the joint probability of staying in or out of poverty, or moving into or out of poverty over two periods. The results showed that Namibia deals with both chronic and transitive poverty. Poverty is found mostly in rural agriculturally dominant areas. It was shown that across these regions, the dominant nature of poverty varies. Particularly, chronic poverty was dominant in the poorest three regions.

In general this thesis contributes by providing new information on the state of welfare and mobility of the poor in sub-Saharan Africa. This thesis also contributes by showing that even when ideal data does not exist, knowledge can be enhanced by the extensions of good methods, used in non-agricultural settings and data used for alternative purposes, such as commercial yield prediction. Finally, this thesis highlights the importance of acknowledging the links between poor households and smallholder agriculture. As is shown, survey data capturing these links is needed, and where it is available, it enhances knowledge on poverty and pathways out of it. Acknowledging the linkages also enables the use of alternative data sources, such as daytime satellites, to improve our understanding on welfare.

Opsomming

Alhoewel die armoedekoers in Sub-Sahara Afrika afgeneem het, het die getal armes toegeneem (Beegle et al., 2016). Die meeste armes word disproporsioneel aangetref in landelike gebiede waar bestaanslandbou beoefen word. Die gebrek aan voldoende data bemoeilik ongetwyfeld pogings tot armoede-verligting. Opbrengste en inkomste in bestaanslandbou is moeilik meetbaar, en heelparty ontwikkelende lande in Sub-Sahara Afrika kort betroubare statistiese inligting (Carletto et al., 2015). Hierdie proefskrif spreek hierdie kwessies aan en toon, aan die hand van drie afsonderlike studies, hoe bestaande data en meetmetodes uitgebrei kan word om die gebrek aan goeie data te oorkom en in die proses ook kennis rakende welvaart uit te brei.

In die eerste studie word ekonometriese tegnieke wat in die formele sekor gebruik word, uitgebrei om 'n beter begrip van die invloed van onderwys op landouprodutiwiteit in Malawi te verkry. Armoede in Malawi volg die tipiese Sub-Sahara patroon, waar die meeste armes in bestaansboerdery is. In sulke omstandighede word onderwys dikwels beskou as instrument om mense uit armoede te lig. Maar min inligting is beskibaar oor hoe skolastiese ontwikkeling produtiwiteit in die landbousektor beïnvloed. Alhoewel die rol wat onderwys in bestaanslandbouproduktiwiteit speel al vantevore ondersoek is, probeer hierdie studie veral die endogene aard van onderwys in landbouproduktiwiteit aanspreek. Die studie dra tot die bestaande literatuur by deur vir die eerste keer die kousale rol van onderwys op landbouproduktiwiteit te probeer meet. 'n Instrumentele-veranderlike (IV) benadering word gevolg, waar die inwerkstelling van gratis primêre onderwys en die ouderdom waarop kinders hulle vader verloor het as IV's vir skolastiese opvoeding geld. Die instrumente meet 'n lokale gemiddelde behandelingseffek en die opbrengste op onderwys onderskeidelik vir individue wat slegs met skool begin het weens die inwerkingstelling van gratis primêre onderwys, en vir individue wat die skool vroeg verlaat het as gevolg van die dood van hul vader. Die resultate waartoe gekom word wys dat daar groot verskille in opbrenste tot onderwys is vir die twee groepe. Uitkomste is laag en onbeduidend waar gratis primêre onderwys as 'n IV geld, maar is groter en meer beduidend waar vader se sterfte as IV geld. Dus, al hoewel opvoeding 'n effek op landbouproduktiwiteit kan hê, geld dit hier nie ten opsigte van die grootste deel van die bevolking nie, veral nie individue wat spesifiek deur die uitbreiding van toegang tot onderwys geraak word nie.

Die tweede studie wys hoe bestaande satelliet-data gebruik kan word om kennis uit te brei oor die stand van welvaart in landelike boerderygebiede, waar gewone statistiese data oor armoede en die stand van welvaart nie geredelik beskikbaar is nie, of net statisties verteenwoordigend is

oor groter geografiese gebiede. Satelliet-data spreek beide hierdie tekortkominge gedeeltelik aan, omdat kleiner geografiese areas teen hoë frekwensies gefotografeer kan word. Nagligte satelliet-data korreleer goed met BBP, maar is nie verteenwoordigend van areas waar daar nie voldoende elektrisiteit is nie, soos landelike, landbou-gebaseerde gemeenskappe in ontwikkelende lande. Daarom ondersoek hierdie studie die gebruik van daglig satelliet-data. Daglig satelliet-data wat landgebruik en plant-gesondheid meet, word gebruik om sosio-ekonomiese uitkomste oor streke te modelleer, maar geen studie het tot nogtoe ondersoek of daglig satelliet-data gebruik kan word om welvaartsverandering oor tyd te meet nie. Hierdie studie voer aan dat aanwysers (indikatore) van plant-gesondheid, afgelei vanaf satelliet-data, gebruik kan word om welvaart oor tyd te meet in areas waar landboubedrywighede dominant is. Sulke aanwysers dien reeds algemeen om verwagte landbouopbrengste te voorspel en behoort dus met welvaart te korreleer. Die studie vors na of dit wel so is, deur data oor Namibië te gebruik. Eerstens is landerye waarop gewasverbouing plaasvind, geklassifiseer. Die bevindinge wys dat daglig satelliet-data areas van ekonomiese bedrywighede in die landbou kan aandui waar nagbeligting-satelliete dit nie kan doen nie. Tweedens is die verwantskap tussen welvaart en aanduiding van plant-kwaliteit ondersoek. In 'n algemeen-ekonomiese deursnit van 'n land of gebied, sal plant-gesondheid nie positief met welvaart korreleer nie, wat verstaanbaar is omdat meer arm mense in landelike gebiede woon. Maar in die landelike gebiede self korreleer plantkwaliteit positief met welvaart oor tyd. Dus kan daglig satelliet-data gebruik work om welvaartsveranderings te monitor in gebiede waar daar min of geen lig in die nag aanwesig is.

Die finale studie toon aan hoe opeenvolgende deursnit-data aangewend kan word om kennis oor ekonomiese mobiliteit uit te brei, veral in 'n data-arm gebied soos Namibië, waar longitudinale of paneeldata nie bestaan nie. Alhoewel daar oor die afgelope twee dekades 'n skerp afname in armoede in Namibië was, was die armoedekoers in 2009/10 steeds 28,7%. Geen inligting bestaan egter oor die dinamiek teenwoordig in huishoudelike armoede asook dit wat armoede oor die tydperk heen laat afneem het nie, omdat geen paneeldata beskikbaar is nie. Om hiedie tekortkoming aan te spreek, ontgin die studie 'n metode ontwikkel deur Dang et al. (2014), wat herhalende deursnit-data gebruik om die betroubaarheidsgrense te bepaal van die proporsie van die bevolking wat oor tyd arm gebly het of buite armoede gebly het, asook van dié wat oor die armoedelyn beweeg het. Die resultate toon aan dat Namibië sowel chroniese as kortstondige armoede beleef en dat armoede meestal veral in landelike landbou-dominante gebiede voorkom. Die studie toon ook dat die aard van armoede in hierdie gebiede varieer, maar dat chroniese armoede veral in die drie armste streke dominant is.

Oorkoepelend beskou, dra hierdie proefskrif by om nuwe inligting te verskaf oor die stand van welvaart en inkomste-mobiliteit onder armes in Sub-Sahara Afrika. Die proefskrif lewer ook 'n bydrae deur aan te dui, ook in gevalle waar voldoende goeie data nie bestaan nie, hoe kennis vermeerder kan word deur bestaande analitiese metodes uit te brei, insluitende metodes van buite landbou-omgewings of data wat vir alternatiewe doelstellings, soos byvoorbeeld kommersiële opbrengs-skattings, gebruik word. Laastens beklemtoon die proefskrif die belangrikheid van die verband tussen arm huishoudings en bestaansboerdery-landbou. Soos getoon is, is daar 'n behoefte aan meer meting en opnames waardeur meer data ingesamel word, om hierdie verband verder te ontgin, want waar sulke inligting beskikbaar is, bied dit sowel inligting oor armoede as moontlikhede om armoede te verminder. Die erkenning van hierdie sterk verband tussen armoede en landbou skep ook moontlikhede vir die gebruik van alternatiewe data-bronne, soos byvoorbeeld daglig-satelliete, om ons begrip van ekonomiese welvaart uit te brei.

Acknowledgements

I want to thank my supervisor Prof. Servaas van der Berg for giving me the opportunity to pursue my PhD and all your support and guidance throughout the process. To my co-supervisor Dieter von Fintel, thank you as well for all your support and for the countless times I was able to pop in to your office on short notice to discuss my work.

I am grateful to all my colleagues at Research on Socio-Economic Policy (ReSEP) and the Department of Economics at Stellenbosch University for support and friendship throughout the process and for providing constructive criticism and suggestions.

I want to thank Dr. Christopher Robert with Evidence for Policy Design (EPoD), who granted me the opportunity to travel to India with the ATAI/KGFS Doctoral Student Research Fellowship. The fellowship led directly to the pursuit of using daytime satellite data to measure outcomes in agricultural areas.

I want to thank colleagues at other faculties and institutions and who have hosted me and granted me opportunities to present my work: Janneke Pieters at Wageningen University and Erwin Ooghe at KU Leuven University. I am also grateful for being able to present at the following conferences: Economic Society of South Africa biennial conference (2015) and the 9th RGS Doctoral Conference in Bochum (2016).

I want to thank researchers at the World Bank who have given me the opportunity to work on projects relating to my PhD: Isis Gaddis, Amparo Palacios-Lopez and Gbemisola Oseni.

To my family and friends, thank you for all your love and support and for keeping me going. To my two cats: I am grateful for your company and comic relief you provided me while working at home. Some special thanks is also reserved for a music album I listened to continuously for inspiration: Pat Metheny - Secret Story.

To my beautiful wife Carla: thank for all your love and friendship. Thank you supporting me throughout my PhD and giving me hope and guidance when needed!

Thomas Ferreira

Contents

List of Tables

List of Figures

List of Abbreviations

2SLS	Two-stage Least Squares
ACR	Average Causal Response
AFD	Age when Father Died
AM	Refers to the authors and method proposed in Antman and McKenzie (2007).
BGK	Refers to the authors and method proposed in Bourguignon et al. (2004).
CDF	Cumulative Distribution Function
CGA	Centre for Geographic Analysis
DHS	Demographic and Health Survey
DLLM	Refers to the authors and method proposed in Dang et al. (2014).
DMSP	Defence Meteorological Satellite Program
EA	Enumeration Area
EMR	Electro Magnetic Radiation
EVI	Enhanced Vegetation Index
FAO	Food and Agricultural Organisation
FFS	Farmer Field Schools
FISP	Farm Input Subsidy Program
FPAR	Fraction of Photosynthetically Active Radiation
FPE	Free Primary Education
GDP	Gross Domestic Product
HDI	Human Development Index
IHS	Integrated Household Survey (Malawi)
INEVI	Integral of Enhanced Vegetation Index
ISP	Input Subsidy Program
IV	Instrumental Variable
LATE	Local Average Treatment Effect
LSMS	Living Standards and Measurement Study
LSMS-ISA	Living Standards and Measurement Study - Integrated Surveys on Agriculture
MDG	Millennium Development Goal
NASA	National Aeronautics and Space Administration
NIDS	National Income Dynamics Study (South Africa)
NDLI	Night Lights Development Index
NDVI	Normalised Differenced Vegetation Index
NHIES	Namibia's Household Income and Expenditure Survey
OLS	Ordinary Least Squares
PQD	Positive Quadrant Dependent
RS	Remotely Sensed
SPF	Stochastic Production Frontier
SSA	Sub-Saharan Africa
SWAPO	South West Africa People's Organization
TFP	Total Factor Productivity
UN	United Nations
USAID	United States Agency for International Development
USGS	United States Geological Survey

Chapter 1

Introduction

Poverty in sub-Saharan Africa (SSA) remains high. Beegle et al. (2016) estimates that in 2012 43% of the population were living on less than US$1.90 per day. The poor are predominantly located in rural regions with their livelihoods strongly linked to their own smallholder agricultural activities. In this context, poverty alleviation efforts have been impeded by a lack of good data which is required to inform policy. Jerven (2016) states that *"we know less about growth and poverty based on numbers in African economies than we would like to think"*. This is due to many reasons which include poor statistical capacity, perverse incentives, and the fact that subsistence agriculture is hard to measure (Carletto et al., 2015; Jerven, 2016; Sandefur and Glassman, 2015).

A recent focus on identifying causal relationships in development economics highlights these data limitations even more. Randomized control trials are ideal to identify causal relationships but these are generally costly. Household panel data which eases identification is also at a premium in SSA. Even when the data is available, the quality needs to be scrutinized. In particular, incomparability over time and measurement error are of concern when measuring smallholder agricultural outcomes. Indeed, the World Bank has recently started undertaking survey experiments with the specific purpose of improving measurement of agricultural activities (see for example Deininger et al. (2012)).

In these instances, and when economic experiments are not viable, researchers need to rely on quasi-experimental econometric methods to utilise the available data to its fullest extent. This thesis contains three applications that show how some of the data deficiencies can be overcome in mainly rural settings in SSA. It applies methods and data from other fields to better understand the extent and nature of poverty and mobility in agricultural settings.

This thesis uses Malawi and Namibia as case studies for analysis. The socio-economic conditions in Malawi fit the typical pattern of low-income income countries in SSA: high absolute poverty rates and high levels of reliance on smallholder agriculture.[1] Malawi also has good household level data covering agricultural activities. This used to highlight the value of such data and how existing labour market techniques can be extended to agriculturally dominant settings. By contrast, Namibia represents an example of a country with relatively poor data availability. While Namibia is defined as an upper-middle income country by the World Bank, it is highly unequal and a large share of the poor are reliant on smallholder agriculture. Namibia, is thus used as an example to present methods for overcoming data deficiencies to measure poverty and highlight how a consideration of links between poverty and agriculture can be utilised to overcome data limitations. This is useful for much of SSA.

In chapter 2, a recent dataset containing data on smallholder productivity in Malawi is used to estimate the causal effects of education on agricultural productivity using an instrumental variable approach. These are the first causal estimates of this relationship in SSA and identify how policy interventions and household events can be utilised to infer causality in cases where randomised control trials are either absent or impractical. In chapter 3, it is shown how satellite data can be used to enhance knowledge on welfare in agricultural settings. It shows, using Namibia as example, that daytime satellite data can be used at small geographic level to track welfare over time in rural areas. Previous studies have only highlighted cross-sectional relationships, which this study shows can be misleading for causal interpretations. Finally, chapter 4 illustrates how repeated cross-sectional surveys can be used, where panel data does not exist, to enhance knowledge on economic mobility in Namibia. No panel data exists and poverty dynamics have never been studied in Namibia.

These three chapters highlight that it is possible to expand knowledge on development outcomes in SSA, even when the ideal data does not exist. The next section discusses the state of poverty and the data landscape in SSA. This is followed by a description of the three chapters and this introductory chapter ends with some concluding remarks.

1.1 Background

A recent book on poverty in SSA by Beegle et al. (2016) found that between 1990 and 2012, the poverty rate decreased from 57% to 43%, but the absolute number of people in poverty increased from 280 million to 330 million.[2] They also noted that poverty is higher in rural areas where the majority of the population lives (65%-70%) and that decreases in headcount poverty were smaller in rural than urban areas. Agriculture remains important for the poor. Most rural households (92%) practice some form of agriculture and it constitutes 64% of income earned (Davis et al., 2017). The link between poverty and agriculture therefore merits further investigation.

[1] For Malawi in 2010, the rural headcount poverty rate was 56.6% and approximately 70% of the employed worked in subsistence agriculture (Own Calculations from Malawi's Third Integrated Household Survey - See Chapter 2).

[2] Beegle et al. (2016) noted that the estimated decrease in poverty rates could be an underestimate.

For poverty alleviation, the importance of growth in the agricultural sector cannot be understated. Ravallion and Chen (2007) found that in China, to account for the reduction in poverty from 53% (1981) to 8% (2001), growth in the primary sector (mostly agriculture) had about a four times larger effect than growth in other sectors. A cross-country study of the relationship between agricultural output and poverty by Irz et al. (2001) found that for a typical low-income country with a headcount poverty rate of 40%, an annual growth rate of 2.17% in agricultural productivity was associated with a decrease in the headcount poverty rate of 10 percentage points in just 10 years. The authors also found that an increase in agricultural productivity of 1% was associated with a 12% increase in a country's Human Development Index (HDI). Christiaensen et al. (2011) note that increasing agricultural productivity is particularly effective at reducing extreme poverty, mainly because this is the sector in which the poor participate the most.[3]

Yet, it is notable that SSA was the only region where the Millennium Development Goal (MDG), which was set in 2000, and aimed to halve poverty by 2015, was not reached (Beegle et al., 2016).[4] It should come as no surprise that growth in agricultural productivity in SSA is low. Fuglie (2008) found that between 2000 and 2006, growth in total factor productivity (TFP) in agriculture in SSA was the lowest of all regions in the world. SSA's average annual growth rate in TFP for the period 2000 - 2006 was estimated at 0.61%. For North and East Africa it was 1.56% while for Brazil and China it was 3.66% and 3.22% respectively. This indicates a lack of understanding on poverty and particularly the agricultural activities of the poor.

An important reason for this lack of understanding stems from what Devarajan (2013) labelled as *"Africa's Statistical Tragedy"*. He notes that much of what researchers know about growth and poverty in SSA is based on fundamentally questionable data. Carletto et al. (2010) noted that a recent food crisis placed focus on the weaknesses of agricultural data systems and highlighted the importance of good data for the generation of knowledge.

Data availability in SSA is poor. Between 1990 and 2012, only 27 of the 48 countries in SSA had conducted at least 2 comparable nationally representative household surveys to track poverty (Beegle et al., 2016). In an assessment of the availability of Demographic and Health Surveys conducted in SSA, Jerven (2016) noted that since their introduction thirty years ago, 43 countries have been surveyed of which only 12 countries have been surveyed 5 or more times. Five countries have only been surveyed once and 5 have never been surveyed.[5] Given that cross-sectional surveys are scarce, it should come as no surprise that the availability of panel data is even worse. By tracking the same households and farmers over time, panel data can enhance knowledge of dynamic relationships and assist with addressing causality by allowing researchers to control for unobservable characteristics.

[3] Extreme poverty is expressed as the squared poverty gap at the $1 per day poverty line.

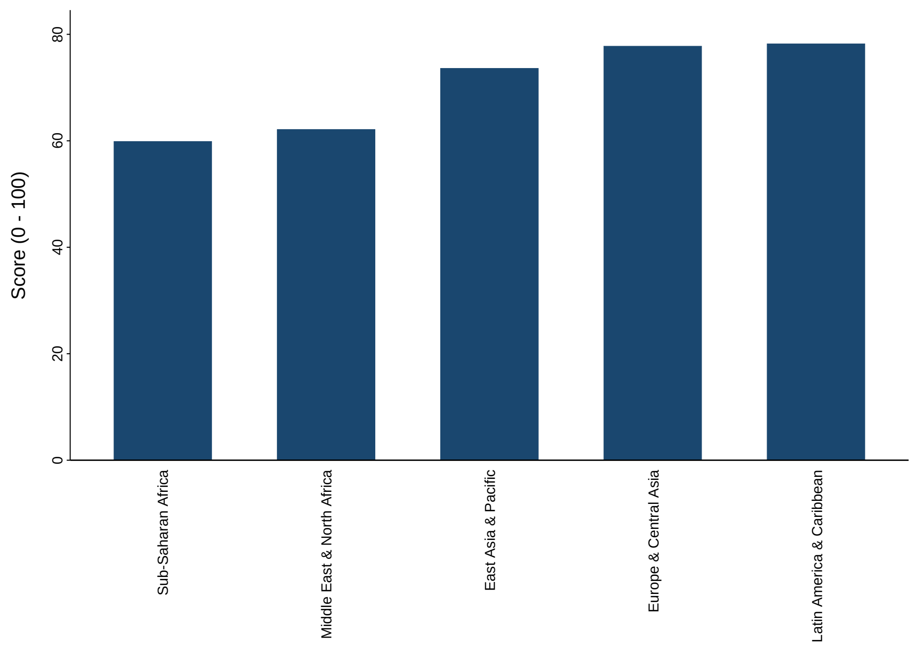
[4] Easterly (2009) notes that the MDG goals were fairly arbitrary and the fact that SSA was not achieving them masks the gains that were made.

[5] See (Beegle et al., 2016) and Jerven (2016) for a comprehensive oversight of data availability in SSA.

Data quality is also a concern. This is highlighted by the recent (dramatic) revisions of gross domestic product(GDP) in Nigeria and Ghana (see Jerven (2016)). Desiere et al. (2016) and Carletto et al. (2015) also show how the same measure can differ by data source. Carr-Hill (2014) also argues that, by design, surveys exclude homeless and nomadic individuals, and under-represent the population in areas that are challenging to survey. He argues that these are some of the poorest individuals in society and estimates that they represent around 11% of the population in SSA.

To further highlight the data constraints in SSA Figure 1.1 shows the World Bank's Statistical Capacity indicator for 2016 for developing regions of the world (excluding high-income countries in those regions). It is a composite measure which evaluates the ability of countries to produce, analyse and distribute quality data in a timely fashion. For 2016, SSA performed the worst of all developing regions (World Bank, 2017a).[6]

Figure 1.1: Statistical Capacity for 2016 in Developing Regions (Excluding High-income Countries)



Source: World Bank (2017b)

A number of reasons have been given for the poor state of data. With regards to agricultural data, Carletto et al. (2015) cite weak demand, poor co-ordination and the fact that smallholder agricultural activities are difficult to measure. Sandefur and Glassman (2015) note that there are incentives for governments and ground-level service providers to over-represent developmental progress. This is also echoed by Desiere et al. (2016), showing that in Rwanda, yield increases due to agricultural reforms were anything between 10% and 60%, but official discourse uses the highest value. They state that differences

[6] For more information on the indicator refer to `http://datatopics.worldbank.org/statisticalcapacity/files/Note.pdf`

in estimates are due to both difficulty in measuring agricultural outcomes and an incentive to show that the reforms work.

This being said, strides have been made in improving data availability and quality. A recent effort by the World Bank's Living Standard Measurement Team - Integrated Surveys on Agriculture, aims to improve knowledge on household agriculture by conducting panel data surveys on households and their agricultural practices (Carletto et al., 2010). Such surveys are, however, currently only conducted in eight SSA countries.[7] Survey experiments are also being conducted to improve the quality of data captured.

Apart from increasing the supply and quality of data, econometric methods have been developed to enhance what can be said with available data. Pseudo-panel techniques, pioneered by Deaton (1985), have made it possible to estimate dynamic relationships from repeated cross-sectional data. It has also been shown how to map poverty to smaller geographic areas using income and expenditure surveys in combination with census data (Elbers et al., 2003). Methods have also been developed to exploit data not initially intended for economic studies such as satellite (see for example Henderson et al. (2012)) and mobile phone meta-data (see Blumenstock et al. (2015)).

Within this context, this thesis contributes to the development literature in SSA by adding new information on the nature and measurement of poverty, which can potentially assist in identifying pathways out of it, highlighting how to exploit existing data to the fullest extent. In chapter 2, good quality household agricultural data is used to present causal estimates of the returns to education in agricultural productivity in Malawi. Chapter 3 shows how indicators of vegetation quality derived from satellite data can be used to track welfare over time in Namibia and chapter 4 uses a new econometric method to exploit repeated cross-sectional surveys to explore economic mobility in Namibia where no panel data exists. Each of these chapters is now briefly described.

1.2 The Causal Effect of Education on Agricultural Productivity in Malawi

The extent of rural poverty and rural links to smallholder agriculture naturally leads to the question of how to increase agricultural productivity of smallholder farmers. Education, often touted as a pathway out of poverty, presents a plausible channel. Welch (1970) noted that education can improve farmers' productivity by making them more productive with a given set of inputs (worker effect) and by making them allocate inputs more effectively (allocative effect). Schultz (1975), however, questioned the value of education in traditional agricultural settings, noting that returns would likely be higher in technologically advanced settings. Large investments in education in SSA, fuelled by the MDG's and the Education for All Initiative, add to the relevance of the topic.

[7] The eight countries are: Burkina Faso, Ethiopia, Malawi, Mali, Niger, Nigeria, Tanzania and Uganda.

There is only weak evidence (in terms of low returns) that education does improve agricultural productivity (Asadullah and Rahman, 2009). Existing evidence is also based on methodology that does not allow for a causal interpretation of the returns to education in agricultural productivity. Existing research has largely used ordinary least squares (OLS) and stochastic production frontier analyses that do not account for endogeneity bias. Not accounting for farmers' ability, and the fact that more productive farmers can afford to invest in education, implies that current estimates are potentially biased.

This chapter presents (to the best of my knowledge) the first causal estimates of the returns to education in agricultural productivity in SSA. This chapter overcomes the limitation of endogeneity by using an instrumental variable (IV) approach applied to Malawi's Third Integrated Household Survey, which combines household and individual level data with data on their agricultural practices. This chapter also shows how econometric techniques, traditionally used in formal sector labour market analysis, can be extended to the smallholder agricultural settings.

Two IV's are used which exploit the introduction of Free Primary Education (FPE) and the age of paternal orphanhood. The IV's used estimate local average treatment effects (LATE) for two distinct groups of individuals. Using the IV based on FPE estimates the returns for individuals who would not have enrolled in primary school, had it not become free. The IV based on age of paternal orphanhood estimates the returns for individuals who would not have left school, had their father not passed away.

The results show that these two subgroups have large differences in returns to education. When FPE is used, returns to education in subsistence agriculture are low and insignificant. Point estimates are similar in magnitude and cannot be statistically differentiated from the OLS estimates which are significant. Estimates when age of paternal orphanhood is used as an IV are larger and statistically significant. This suggests that education can have an effect on agricultural production through a worker and allocative effect, but this is not given for the total population.

For comparison, IV estimates of the returns to education are also estimated on earnings in the non-agricultural formal sector. This also presents first causal estimates of returns to education in the formal sector in Malawi. Point estimates are higher using both IV's, but the estimate using FPE remains insignificant. This indicates that education is more valuable in the formal sector than in smallholder agriculture and this supports the notion that education has greater returns in technologically advanced settings. What is evident, however, is that FPE did not induce increases in agricultural productivity for a large proportion of the population at whom the policy was aimed.

1.3 Using Satellite Data to Measure Welfare in Namibia

Where the previous chapter sheds light on potential interventions to alleviate poverty in agricultural settings, this chapter discusses how satellite data can be used to enhance knowledge on poverty in agricultural settings. As noted in the introduction, only a handful of countries have conducted income and expenditure surveys that are comparable over time. When they are conducted, it is usually infrequent. Another problem is that these surveys are not representative at small geographic levels. Information that can identify poor areas more accurately will improve the efficiency of poverty interventions. Satellites capture data at very fine geographic level, and do so frequently. It thus presents a potential source of data to overcome survey limitations.

One satellite series in particular, the night lights, has been shown to correlate well with economic activity (Henderson et al., 2012). Yet large inhabited areas in SSA are under-electrified, implying that night lights cannot estimate economic activity in such areas. To use the power of satellites in these areas implies that researchers need to use daytime satellite data. Weeks et al. (2012), Watmough et al. (2013) and Watmough et al. (2016) have explored the use of vegetation quality and land use data derived from satellites to predict socio-economic outcome across regions, but no studies have explored correlations between socio-economic outcomes and daytime satellite data over time. In sub-Saharan Africa, there is clearly selection effects present when it comes to welfare across regions. The poor are located in rural agricultural dominant regions. Thus, vegetation quality, for example, would correlate negatively with welfare cross-sectionally. It would, however, be wrong to assume that increases in vegetation quality lead to increases in poverty over time.

Chapter 3 suggests that indicators of vegetation quality, derived from satellite data, can be used to track welfare longitudinally in rural areas dominated by smallholder agriculture. These are generally also areas where night lights are absent. This is based on the premise that indicators of vegetation quality are used to predict agricultural yields, and if livelihoods are dependent on agricultural production, it should be the case that increases in vegetation quality should be associated with increases in welfare.

This is explored in the context of Namibia - a country with large regional inequalities. The first section of the analysis highlights how different satellite measures observed greater areas of economic activity in different settings. A satellite analysis team was hired to identify areas of crop land from daytime satellite imagery. The amount of light and crop land observed were then compared between a wealthy urban region - Erongo - and a poor agrarian region - Kavango. It is shown that night lights are more effective in the urban region. In the agrarian, underdeveloped region, daytime satellite data performs vastly better.

Regression analysis is then used to explore the relationship between welfare and vegetation quality cross-sectionally and longitudinally. Higher vegetation quality is associated with higher poverty in Namibia cross-sectionally. This highlights the selection effect - poor households are likely to reside in rural areas. However, longitudinally within rural regions, increases in vegetation quality are associated with increases in welfare. This finding is robust using a range of vegetation quality indices and two different data sources.

1.4 Using Repeated Cross-sectional Data to measure Economic Mobility in Namibia

As noted earlier, poverty rates across SSA have decreased. Namibia is no exception. Poverty decreased by at least 20 percentage points since political independence in 1991. However, poverty remains high with approximately 29% of the population being poor in 2009/10. To gain deeper insights into poverty trends since independence, chapter 4 explores what can be said about economic mobility in Namibia.

Economic mobility traditionally requires the use of panel data which is absent in Namibia. Thus, chapter 4 explores how repeated cross-sectional surveys can be used to inform economic mobility. Three methods in particular - Bourguignon et al. (2004), Antman and McKenzie (2007) and Dang et al. (2014) - have been proposed to provide insights into economic mobility with repeated cross-sections. The only method truly useful in the case of Namibia is that of Dang et al. (2014). Based on two survey periods, the method estimates bounds on the proportion of the population that remained in or out of poverty; and moved into or out of poverty. The method thus provides insights into whether poverty is mainly chronic or transitive in nature.

Poverty is concentrated in rural areas and the results suggest that Namibia experiences both chronic and transitive poverty. Sub-analysis highlights that the dominant nature of poverty is spatially divided. In three particular regions, home to 50% of the poor in Namibia, it is found that poverty is mainly chronic.

This chapter extends the use of cross-sectional surveys to explore economic mobility. It thereby provides insights in poverty dynamics in Namibia that have never before been provided. The fact that poverty is based in rural, agrarian regions, and that the dominant nature of poverty differs over these regions, highlights how useful datasets of the type used in chapter 2 can be to improve the understanding of the relationship been agriculture and poverty.

1.5 In Conclusion

Poverty in SSA remains stubborn and widespread. While poverty rates have decreased in the last 2 decades, the absolute number of poor has increased. The poor generally live in rural areas, and most poor households practise agricultural activities. Poverty alleviation efforts been limited, however, due to poor data. In this context this thesis provides valuable insights into welfare outcomes in SSA by extending existing data and methods, with a particular focus on links between welfare and agriculture.

In chapter 2 it was shown that expansions in educational access are not enough to increase agricultural productivity for the population targeted by such interventions. In chapter 3 it was shown how poverty can be tracked longitudinally for small agricultural regions and periods where household panel data does not exist and is not representative. Chapter 4 shows, even in the absence of panel-data, how poverty in Namibia is both chronic and transitive in nature and that the nature of poverty differs across rural regions.

This thesis demonstrates that it is possible to expand knowledge on agricultural and welfare outcomes even when it is generally believed that the availability and quality of data is poor.

Chapter 2

Does Education Enhance Productivity in Smallholder Agriculture? Causal Evidence from Malawi

2.1 Introduction

The Malawian economy fits the typical pattern observed in sub-Saharan Africa (SSA). Malawi is a poor country with a small formal sector and large informal sector. In 2010 approximately 70% of the employed worked in subsistence agriculture.[1] Approximately 55% of employed individuals worked in household agriculture while another 17% worked as ganyu labourers, which are casual labourers mostly associated with household agriculture. At the same time, the rural headcount poverty rate was around 56.6% (NSO, 2012a). With the majority of the population living in poverty while working in subsistence agriculture, increasing agricultural productivity is key to addressing poverty.

Chapter 1 presented evidence for the importance of growth in the agricultural sector for poverty alleviation. It was also noted that productivity growth in agriculture in sub-Saharan Africa (SSA) is low. For the period 2000 - 2006, growth in total factor productivity (TFP) in agriculture was 0.61% for SSA - the lowest for any region in the world (Fuglie, 2008). In Malawi around that same period, the growth rate of GDP per capita in smallholder agriculture was negative. For the period 2000-2005 it was -1.78% (Chirwa et al., 2008).

One channel through which agricultural production can be increased is education. Numerous authors have hypothesized how education can affect production. Welch (1970) suggests that education can work through two channels - a worker and an allocative effect. The worker effect refers to how education can make farmers more productive given the resources they have, while the allocative effect is the effect that education can have on farmers' ability to use and allocate resources more efficiently. Many authors have questioned the value of education in traditional agricultural settings and noted that education becomes

[1] Estimates are based on own calculations from Malawi's Third Integrated Household Survey, which was conducted in 2010/2011.

more valuable in technologically advanced settings where the ability to, for example, adopt and learn to use new technologies becomes important (Welch (1970); Schultz (1975); Ram (1980)).

Asadullah and Rahman (2009) note that empirically there is only weak evidence that education really has positive effects on agricultural productivity. A concerning factor about the literature is a lack of causal interpretation. To estimate the effects of education on agricultural productivity, most studies have used Ordinary Least Squares (OLS) and/or Stochastic Production Frontier (SPF) methods which measure the role of education, either in the estimation of the agricultural production frontier or as an explanatory variable of inefficiency. None of the methods allow for causal interpretation, however, as they do not account for the endogeneity of education.

This study fills this gap by estimating the causal effect of education on agricultural productivity using an instrumental variable (IV) approach. It contributes to the literature by presenting the first causal estimates of the effect of education on agricultural productivity in SSA. It also shows how a estimation method traditional used in the formal sector can be applied to smallholder agriculture.

Two instruments are used for education. The first instrument uses the gradual introduction of free primary education (FPE) in the early 1990's in Malawi. The second instrument is the age of paternal orphanhood. The first represents relief from a credit constraint and the second an imposition of one. The IV estimates capture local average treatment effects (LATE). When using FPE as IV, the returns to education are measured for individuals who would not have entered schooling had it not become free. Using age of paternal orphanhood measures the returns for learners who would have stayed in school had their father not died.

Data from Malawi's Third Integrated Household Survey, conducted in 2010/11, is used for the analysis. Production functions for the cash yield of crops are estimated for Malawi's rainy season. The effect of education on formal sector earnings is also estimated using the instruments to gain a deeper understanding of the role that education plays in the country in general. The causal effect of education on earnings has also never been estimated for Malawi.

In agricultural production, the returns to education are low and insignificant when FPE is used as IV.[2] Returns are 1% when controlling for inputs and 3% when not. When age of paternal orphanhood is used as an IV, point estimates are higher and statistically significant with a 7% return when controlling for inputs and 9.3% when not. Returns to education are higher in the formal sector, however. Using FPE, the point estimate is 6.2% but insignificant. Using age of paternal orphanhood, the returns are significant and 15.4% to an extra year of education.

The difference in returns between the two groups captured by the IVs is striking. This study shows that while there can be returns to education in agricultural production, it is not guaranteed. Indeed, it is

[2] When FPE is used as IV, point estimates are statistically similar to ordinary least squares estimates which are statistically significant.

found that there are low returns to individuals, at whom the large expansions in education were especially targeted.

Section 2.2 provides a literature review on the relationship between education and agriculture from both a theoretical and empirical perspective. Section 2.4.1 discusses the data used. Section 2.3 provides a background on agriculture and education in the country. The methodology is discussed in section 2.4.2. Section 2.5 provides the empirical results. This is followed by a discussion of the results in section 2.6 and section 2.7 concludes.

2.2 Literature Review

2.2.1 Model

This review of the literature starts with a short theoretical discussion on how education can affect production in general. Welch (1970) introduced a model whereby the potential effects of education in production can be divided into two categories: the worker and allocative effect. Firstly assume that production, Y, is a function of education, Z, and a set of other inputs, X.

$$Y = f(Z, X) \tag{2.1}$$

The marginal product of education is $\frac{\delta y}{\delta z}$. This captures the "worker effect" of education which is just the amount of extra production due to increases in education if all other inputs are held constant.

In many cases producers produce more than one crop and they need to decide in what combination to produce them with the resources available. In agriculture for example, farmers need to decide what crops to grow. To illustrate this, define a total value function as:

$$Y = p_1 y_1(x_1) + p_2 y_2(x_2), \tag{2.2}$$

where Y is value of all crops produced, p_1 and p_2 refer to the prices of the crops y_1 and y_2. X is a given vector of inputs that needs to be allocated between different uses x_1 and x_2. The maximization of total value given X happens when firms are technically efficient and when:

$$\frac{\delta Y}{\delta x_1} = p_1 \frac{\delta y_1}{\delta x_1} - p_2 \frac{\delta y_2}{\delta x_2} = 0. \tag{2.3}$$

This implies that the maximization of sales happens when the values of the marginal products of y_1 and y_2 are the same. Now, if it is assumed that education has an allocative role, that is $x_1 = x_1(z)$, then the marginal product of education with regards to gross sales becomes:

$$\frac{\delta Y}{\delta z} = (p_1 \frac{\delta y_1}{\delta x_1} - p_2 \frac{\delta y_2}{\delta x_2}) \frac{\delta x_1}{\delta z}. \tag{2.4}$$

Ram (1980) introduced a modification to the above model by highlighting the role of information. He hypothesized that education decreases the marginal cost of acquiring information and also increases the marginal benefit from information. One implication of his theory was that the value of education is linked with the value of information. Thus in economic settings where information is more valuable, it would be worthwhile to get more educated. One example that he notes is that information is more useful to farm operators who make allocative decisions, than to hired workers who perform more mundane tasks. Thus, education would have higher returns for farm operators than labourers.

Ram (1980) also noted that in more dynamic agricultural settings, information and then schooling might have greater value. This point was also highlighted by Schultz (1975) and Welch (1970). In trying to explain the growth in the demand for education in the United States, Welch (1970) postulated that one of the two plausible explanations was "non-neutrality in production", which meant that the marginal product of education increased with advances in production methods and improvements in technology.

Schultz (1975) also highlighted this by comparing traditional and modern agriculture. In traditional settings where technology has not changed for generations, education is of little value. People have learnt from experience and children learn from their parents. In modern settings, farmers need to deal with a flow of changing technologies and economic conditions. In these circumstances there is a demand for the skills to deal with changes in production methods and new technologies.

New research has also highlighted the role that education has in communities. Educated farmers are more likely to be early adopters of new technologies, and this in turn leads to the diffusion of the technology to other less educated farmers in the community (see Foster and Rosenzweig (1995) and Weir and Knight (2000).

It has also been found that the allocation effect of education is present at the household level as well. Education increases the likelihood that farmers allocate their labour away from farming to non-agricultural activities for which there are higher returns (see Huffman (1980) and Fafchamps and Quisumbing (1999)). In this case education could even have negative effects on aggregate agricultural production.

2.2.2 Evidence of the returns to education in agriculture

Researchers have either used Ordinary Least Squares (OLS) or Stochastic Production Frontier (SPF) methods to measure the effects of education on agricultural production. Welch (1970) also showed that how inputs in estimation models are controlled for, could determine whether the worker or allocative effect is measured. When estimating the production of a single crop, only the worker effect will be measured. Total value production functions can measure the worker effect if other inputs are controlled for, and the worker plus allocative effect, if other inputs are not controlled for. What follows is a selective set of evidence from developing countries, firstly from meta-studies and secondly focusing more on individual studies.

Lockheed et al. (1980) did a survey of the evidence regarding farmer education and productivity for low income countries. They reviewed the evidence of 18 studies, done over 13 countries, with 37 datasets. They concluded that four years of education, when compared to zero years, increased farm productivity by 7.4%. The use of four years was motivated by the fact it is often seen as the minimum cycle for a basic education or for the acquisition of basic literacy. They also found that the effects of education were much higher when farmers used modern technologies. In modern agricultural environments they found that the average effect of four years of education was 9.5% while in traditional circumstances they found the average effect to be only 1.3%.

The Lockheed et al. (1980) results should be read with caution, though. In a comment on the study by Phillips (1987) he noted that only 56.4% of the studies reviewed in Lockheed et al. (1980) found a positive and significant effect while there was large regional variation in their findings. For a 0.05 significance level, 17 of 22 studies in Asia had positive significant effects while for Latin America, only 3 of 13 had positive significant effects. Only two studies in Africa, both on Kenya, were reported on. One found an insignificant effect while the other found a negative effect. In a reply Lockheed et al. (1987) noted that their methods were consistent with meta-analytical techniques. They also noted that the positive significant effects for Asia were plausibly due to modernisation in agriculture in the region. This suggests that educational returns may be context specific as returns only become relevant when production techniques are advanced.

Phillips (1994) conducted a meta-analysis of the existing evidence using an alternative method to Lockheed et al. (1980). Instead of averaging over all observations in all the datasets, he used the results of a number of studies as data points in a new cross-section. He used the same data as Lockheed et al. (1980) plus a number of additional studies. He then regressed the coefficient estimates of all the studies on study characteristics. Characteristics included, amongst others, where the study took place, and whether it was a traditional or modern environment. Results supported those of Lockheed et al. (1980). There were higher returns to education in modernizing environments as well as higher returns for Asian countries. There were, however, no new studies added for Africa. Surveying the existing evidence for

Africa, Appleton and Balihuta (1996) found that effects were mostly statistically insignificant, although they were often large.

Asadullah and Rahman (2009) estimated the effect of schooling on rice production in rural Bangladesh. Using OLS they found positive effects for the household head's education, but after adding in the highest education level in the household, the effect of the head's education became statistically insignificant while the effect of the highest level of education became significant. Estimating the effects with SPF it was found that education was positively correlated with potential output if included in the estimation of the frontier. When education was excluded from the first stage and used as an explanatory variable to measure the sources of inefficiency in the second stage, it was found that education was correlated with reductions in inefficiency, again in the same pattern as the other models.

Appleton and Balihuta (1996) estimated the returns to education for farmers in Uganda using production functions and found positive results. Compared to farmers with no education, four years of education was associated with 7% higher output, while full primary education was associated with 13% higher output. For Ethiopia, Weir (1999) found that education did have positive effects on cereal production but only after four years of schooling had been reached. Having four or more years of schooling increased production by more than 10% compared to no schooling when using OLS.

For northern Nigeria, Alene and Manyong (2007) found significant positive effects for schooling, but only when farmers were using new technologies. They found no effect for those working with traditional technologies. The authors used a switching regression model to model a two-stage process where in the first stage, farmers can choose to adopt better cowpea varieties or not, and in the second stage cowpea production is modelled given adoption or not. Whether the household head had four or more years of education had a positive significant effect on the adoption of better technologies and it also had a positive effect on cowpea production given that they adopted new technologies. Having four years of education increased cowpea production by 25.6%, if they used improved cowpea varieties. The proportion of other household members who had completed primary schooling had no significant relationship with adoption of better varieties or production.

A few studies have also distinguished between the allocative and worker effect of education. Appleton and Balihuta (1996) estimated the effect of education on aggregate production without controlling for capital and other purchased inputs and found evidence for the allocative effect of education. Four years of schooling (when compared to no schooling) was associated with 7% higher production when capital and purchased inputs were controlled for, and 10% higher production when those variables were not controlled for. Weir (1999) only found weak evidence that there was a small (1%) allocative effect on Ethiopian cereal crop production for education of non-farming household members. Ram and Singh (1988) studied education's allocative effects on data in Burkina Faso. They found that for household heads, education had a large allocative effect and small worker effect, while for other household members, education had a large worker effect and small allocative effect.

Earlier, the role of education in the adoption and dissemination of new technologies in communities was discussed, but researchers have also tested more directly whether there are community level effects. This has usually been done by adding controls to production functions which capture the effects of the education of farmers in close proximity to the respondents. Asadullah and Rahman (2009) controlled for neighbours' education in their production functions but failed to find positive effects. Appleton and Balihuta (1996) used the average years of education of other farmers in the enumeration area (EA) of the farmer. They controlled separately for the average years of primary and secondary education and found positive results for primary but not for secondary education. An extra year of average primary schooling in the EA was associated on average with a 4% increase in crop production for farmers.

Davis et al. (2010) discuss another form of education which does seem to have positive effects in Africa - Farmer Field Schools (FFS). FFS are aimed at adults and usually work with farmers meeting once per week informally with a facilitator. Through experiential learning, farmers then learn new techniques, how to solve problems, and they are assisted with decision making. The authors used a number of methods to estimate the effects of FFS on farmers in Tanzania, Uganda, and Kenya. FFS were found to increase crop yields by 80% in Kenya, and 23% in Tanzania, while no significant effects were noted in Uganda, which the authors note might be because another government development plan was also running in the same areas as FFS.

The current literature has however not yet dealt with the endogenous nature of education when estimating returns in agricultural productivity. Results can be biased due two factors not discussed in the literature. Firstly, ability is not controlled for. Higher ability can lead to both higher productivity and higher educational attainment and not accounting for it will upwardly bias the estimated effect of education. This implies that the effects could be smaller in reality than what was found in the studies. Secondly, there is also a problem of the direction of causality. Educational attainment can be a function of farm productivity because households with more productive farms have more income to pay for education. Foster and Rosenzweig (2004) showed that in India, during the green revolution, increases in agricultural productivity, due to technical change, were associated with increases in school enrolment.

Apart from the omission of addressing the endogeneity of education, the general picture that has emerged so far is that education does have a positive relationship with production in modernising agricultural settings, though the effects are small. The effects do seem to be heterogeneous though, and increase with use of better technologies. Furthermore, education leads to earlier adoption of new technologies.

2.3 Background on Malawi

To contextualize the analysis, this section provides a brief background of the Malawian economy and education system. This is followed by a review of previous studies on the returns to education in the country.

2.3.1 Economy

Malawi is a landlocked country in Sub-Saharan Africa that shares borders with Mozambique, Zambia, and Tanzania. The country is rated as a low-income country by the World Bank (2015a). Poverty levels are high. Using the \$1.25 and \$2 per day poverty lines, headcount poverty rates for 2010 were 72% and 88% respectively (World Bank, 2015a). At the national poverty line (37 002 Malawian Kwacha per year) the headcount poverty rate in 2010 was 50.7% nationally, 17.3% for urban areas, and 56.6% for rural areas (NSO, 2012a). Evidently, poverty is more concentrated in the rural areas where the majority of the population are subsistence farmers.

Agriculture, forestry, and fishing is the largest sector in the country. It has contributed between 30-35% of GDP since 2011 (Reserve Bank of Malawi, 2014). The agricultural sector can be divided into two sub-sectors - household and estate agriculture. Household agriculture contributes more than 70% to the sector while estate agriculture contributes the rest.

Malawi's staple food is maize and it dominates both consumption and agricultural production. Maize is planted in October/November and harvested in April/May. Rain is of vital importance to the crop because the extent of irrigation is negligible. The cycle coincides with the rainy season in the country which stretches from January to March (Zant, 2012).

The main cash crop in Malawi is tobacco and specifically burley tobacco which has been named "green gold" because of its profitability (Orr, 2000). Other important crops in the country are sugarcane, tea, coffee, wheat, rice, groundnuts, pulses and cotton. The main export crops are tobacco, sugarcane and tea with the latter two being produced mostly (more than 85%) by commercial estates (Chirwa et al., 2008).[3]

Malawi introduced a large-scale farm input subsidy program (FISP) in the 2005/2006 cropping season. The program aimed to supply 50% of farmers with subsidized fertilizer for maize, and also provided vouchers for modern maize seeds and tobacco fertilizer (Dorward and Chirwa, 2011). The official aims of the program were to increase the production of maize, promote food security at household level and to increase incomes in rural areas. Whether the program has been successful is uncertain (Lunduka et al., 2013).[4]

Unemployment in Malawi is low because most individuals desperately needing employment can revert to the informal sector and do casual work. This type of work is commonly referred to as "ganyu" work. Many individuals engage in this type of labour as a coping strategy to deal with food insecurity (Bryceson, 2006). Whiteside (2000) argues that it is the most significant source of income after own agriculture for rural households. Whiteside (2000) further notes that ganyu labour and household production can be in conflict from time to time as ganyu can lock households in cycles where immediate food shortages are

[3] An estate monopoly on burley tobacco was ended in 1990 which allowed subsistence farmers to plant it (Orr, 2000).

[4] See Lunduka et al. (2013) for a review of the evidence of the program.

addressed but long term food supply is not. This then creates the need to do more ganyu work. The problem is exacerbated by low ganyu wage rates.

2.3.2 Education

Educational attainment in Malawi is very low. The government introduced universal FPE in the 1994/95 school year. Some cohorts also started receiving FPE earlier. FPE was introduced in the 1991/92 school year for standard 1 with the idea that every cohort after that would receive FPE as well. Together with that government programme, the United States Agency for International Development (USAID) introduced a programme in 1992/93 to abolish school fees for all girls in Standards 2-8 who had not repeated (Kadzamira and Rose, 2001). Table 2.1 shows which learners received FPE between 1991/92 and 1994/95.

Table 2.1: Cohorts that received Free Primary Education

		Standard															
		Std. 1		Std. 2		Std. 3		Std. 4		Std. 5		Std. 6		Std. 7		Std. 8	
		M	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*
Year	1991/92	Y	Y	N	N	N	N	N	N	N	N	N	N	N	N	N	N
	1992/93	Y	Y	Y	Y	N	Y*	N	Y*	N	Y*	N	Y*	N	Y*	N	Y*
	1993/94	Y	Y	Y	Y	Y	Y	N	Y*	N	Y*	N	Y*	N	Y*	N	Y*
	1994/95	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Source: Kadzamira and Rose (2001)
M and F refers to males and females respectively. Y and N refer to yes and no respectively on whether cohorts received FPE or not. Y* indicates that only non-repeating girls received FPE.

There is no doubt that the programmes helped increase enrolment, which makes the use of FPE as an instrumental variable relevant. It did, however, put large financial strains on Malawi's education system and consequentially education was not supplied at an acceptable level of quality while it also squeezed resources in other parts of the education system (Kadzamira and Rose, 2003).

With agriculture being the dominant sector in the poor parts of the country, measures to improve agricultural productivity are clearly needed. One channel that provides opportunities for this is education. Introducing FPE is an obvious large expense to a poor country and for that reason measuring whether it has had positive effects on agricultural productivity in Malawi is important.

2.3.3 Evidence on the Returns to Education in Malawi

The effect of education on agriculture in Malawi has not been studied extensively. Chirwa (2005) studied factors that were associated with the adoption of improved technologies – inorganic fertiliser and hybrid seeds. This was estimated using bivariate probit regressions to account for the fact that the decisions to adopt either or both of the technologies might be related. Education was found to be a positive and strongly significant factor in the decision to adopt inorganic fertilizer. It was positively but not significantly associated with the adoption of hybrid seeds. Chirwa (2005) noted that the findings were in contrast to those of Zeller et al. (1998) who did not find a significant effect for education in the adoption of new technologies.

Matita and Chirwa (2009) estimated the returns to education for agricultural activities using data from the Second Integrated Household Survey (IHS2). They estimated OLS models of maize output, tobacco sales, business profits, and household income. For maize output they included an explanatory variable which is presumed to be education of household head.[5] Returns were positive and significant and were estimated at 3.91% per year. Households where the head had completed primary schooling produced 9.85% more than households where the head had no schooling. Similarly, heads with junior secondary schooling produced 17.54% more, senior secondary schooling produced 41.56% more and technical and university graduates produced 77% more than the unschooled. For tobacco earnings, a year of education was associated with a 5.63% increase in earnings. Returns to schooling categories (with no education as the base) were as follows: primary schooling (24.45%), junior secondary schooling (32.44%), senior secondary schooling (53.37%), technical education (150%), and a university degree (160%).

Kassie et al. (2014) studied poverty and its determinants for a small sample of rural maize farmers. They randomly sampled 68 households from the Balaka district and 87 households from the Mangochi district. The two districts were randomly selected but had to adhere to the requirement that they only had a 20-40% probability of having a failed season. This was done to make the districts more comparable with each other by limiting the role that climatic factors could have on the harvests of farmers. They used quantile regressions with per adult equivalent expenditure divided by the poverty line (using the $1.25 per day poverty line) as a dependent variable to estimate the determinants of poverty. They included both a variable indicating whether the household head was illiterate and also the average level of education in the household. Both where found not to be significantly correlated with well-being.

There have also not been many studies on the returns to investment in education in the modern wages sector of the economy. Mingat et al. (1985) used data from the 1981 Household Income and Expenditure Survey, supplemented with civil servants earnings data for 1983, and the 1983 Education Finance Study. They found that compared to anything less than primary schooling, the private rates of returns to primary education were between 11.6% and 15.7% depending on their assumptions. Further, the private rates of return were 26.3% for lower secondary schooling, 16.8% for upper secondary schooling, and 46.6% for university graduates. The social returns were marginally lower but had the same pattern as the private returns except for university, for which the social returns were 11.5%.

Chirwa and Zgovu (2001) used survey data on paid employees from four rural districts to study the determinants between choosing to work casually or formally and whether the returns to education differed between the two sectors. Using Mincerian earnings functions they estimated the average return to an extra year of education to be 6.61% with OLS on the full sample. They also modelled the two sectors

[5] It is not totally clear whether household head's education was used, but Matita and Chirwa (2009, 7) state that their main hypothesis is that the education of a household head is positively correlated with maize production.

separately using OLS and also the Heckman (1979) selection model to control for selection bias. Returns to education in the formal sector were estimated to be 9.44% using OLS and 9.41% using the Heckman estimator. For casual employment, returns were estimated to be 4% but insignificant using OLS, and 5.41% and significant at a 90% confidence level using the Heckman method. Yet samples were small. The authors argued that the results showed that when informal sectors are excluded from estimations in rural economies, the returns to education tend to be overstated.

Chirwa and Matita (2009) used data from the Second Integrated Household survey (IHS2) for 2004/05 to estimate the returns to education. They estimated Mincerian type earnings equations for employed individuals using OLS. The authors noted that due to data limitations, corrections for bias were not possible. They found that education had a significant positive effect on earnings. An extra year of education was associated with a 10.12% increase in earnings on average.

2.4 Data and Methodology

2.4.1 Data

This study uses the Third Integrated Household Survey (IHS3) from Malawi. The survey was conducted in 2010/2011 and forms part of a wider range of surveys conducted by the World Bank as part of the Living Standards and Measurement Study (LSMS) as well as the Living Standard and Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA). The IHS3 data contains household, community level, fisheries, and agriculture components. The household and agricultural components of IHS3 are used in this study. The household questionnaire contains detailed information at the individual and household levels, while the agricultural questionnaire was conducted at plot level. IHS3 surveyed 12 271 households and 10 401 households also answered the agricultural questionnaire (NSO, 2012b).

To show the importance of household agriculture in the country, Table 2.2 shows the distribution of employment over different sectors in IHS3 for the working-age population. Some individuals working in other sectors may also own farms. It is also the case that individuals working in the private sector could be employed on commercial agricultural estates. These individuals are not included under household agriculture as that was not the main sector of employment they reported. Around 55% of the employed work in household agriculture. If ganyu workers, who mostly reside in rural areas, are counted with the household agricultural workers, around 71% of the employed work in smallholder agriculture. Given that individuals mainly employed in other sectors could also possibly manage agricultural plots, the extent of employment in agriculture is possibly understated.

Next, the data used for the subsequent analysis is discussed. The variables used loosely follow Kilic et al. (2015), who studied gender productivity differentials in smallholder agriculture in Malawi. The data is captured in a manner that households can farm numerous plots. Plots could be managed by

Table 2.2: Employment Distribution by Sector

	%	Cum. %
Private	11.6	11.6
Public	3.9	15.6
NGO/Church	1.1	16.6
Self Employed	12.0	28.7
Ganyu	16.8	45.5
Household Agriculture	54.5	100.0

Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.
Sample is limited to the working-age population which is age ≥ 15 & age ≤ 64.

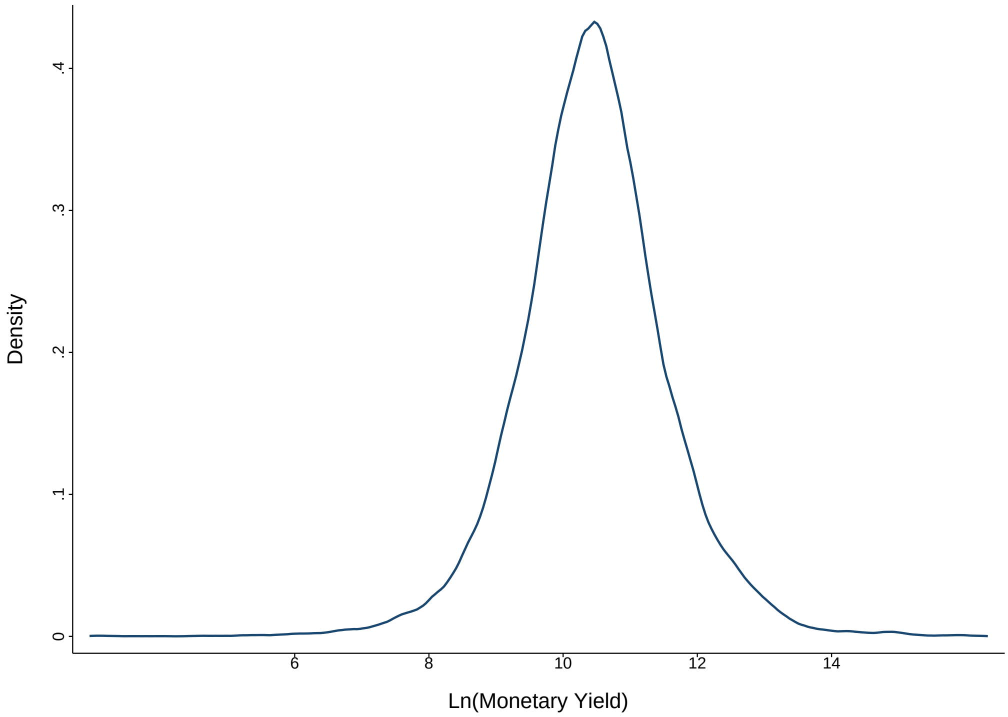
the same or different household members. Plot level data also complicates the estimation procedure. If, for example, the number of plots owned by a household is correlated with education, it will inflate the estimates on education if education has a positive correlation with production. To deal with this, this study takes the probability weight of each individual and divides it by the number of plots he/she manages. Running IV estimates at individual level (instead of plot level) and weighting the estimates using individual probability weights produced similar results.

One concern in the agricultural data was the units of measurement in which harvests were reported. For example, farmers could report harvesting in kilograms, or a number of other non-standard measuring units (such as an ox-cart). The amount of other commodities harvested, such as ground nuts, could also be reported as shelled or unshelled. Harvests were converted back to kilograms using conversion factors which were obtained from researchers for LSMS-ISA. Estimates for the total value of production were then calculated using the median kilogram price in each enumeration area if there were at least 10 observations. If not, the median was taken for the next level of geographical aggregation, which is the traditional authority, followed by district, region and then country. To deal with outliers, OLS regressions on logged yield were estimated using the most exhaustive model specification (discussed later). From this the studentized residuals for the estimates were obtained, and observations with studentized residuals larger than 5 or smaller than -5 were dropped. Figure 2.1 shows the distribution of the logged estimated monetary yields - defined as the total value of all produce divided by land size in hectares. The distribution is fairly smooth and approaches the shape of a normal distribution.

Two indices to control for wealth were also created using the household level data. The first is a farm implements index: because it is created at the household level (as opposed to plot level), it implies potential access to implements. The second is a wealth index that is constructed from whether households own a range of durable goods. Both indices are created using principal components analysis (see Filmer and Pritchett (2001) for example).

Table 2.3 gives a description of the agricultural data used in the regression analysis. This data is based on the sample of working-age population farm managers for which none of the data was missing, which was 14293 plots. The average years of education of plot managers is 5.24 years which represents incomplete

Figure 2.1: Distribution of ln(Agricultural Yields (Kwachas))



Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.

primary education. 27% of plots are managed by females while the average age of plot managers is 38.18 years. Approximately 14% of plot managers are also employed in the formal sector. Furthermore, the average household size for plot managers was 4.89 and the average child dependency ratio in the households was 1.13. The average size of plots was 0.5 hectares while the average monetary yield was 50295.86 Kwachas. In terms of labour, female household members, on average, worked more on plots than male members while hired and unpaid labour was not used much in general.

Apart from inorganic fertilizer, the use of modern technologies was very low. Herbicides and pesticides were used on 1% of plots. To further highlight the low use of technologies, Table 2.4 shows the percentage of plots which had access to different farm implements. Less than 1% of plots had access to tractors, 1.65% had access to an ox-plough and the percentage of plots having access to a ridger, treadle pump and sprayer were 0.66%, 2.69% and 4% respectively. On the contrary, almost all plots had access to a hand hoe. Less than 1% of plots had irrigation systems while 29.29% had access to a watering can.

For the formal sector in Malawi, wages are reported on the main jobs individuals had in the last 12 months. Only the wages of those who were still employed in the formal sector at the time of their interview were considered. Similarly, the sample is limited to the working-age population and only those individuals for which there were no missing values. Outliers are once again determined by running OLS regressions on logged earnings using the most exhaustive model specification (discussed later) and removing observations

Table 2.3: Agriculture Data Description - Plot Level

	Mean	Median	Min	Max
Yield per Ha (Kwachas)	64576.14	30888.13	18.97	12355250
Used Pesticides/Herbicides (Yes=1 No=0)	0.01	0.00	0.00	1
Used Fertilizer - Organic (Yes=1 No=0)	0.12	0.00	0.00	1
Used Fertilizer - Inorganic (Yes=1 No=0)	0.66	1.00	0.00	1
Fertlizer Inorganic (Kg per Ha)	195.68	102.96	0.00	51480
HH Labour - Males (hrs per Ha)	467.16	296.53	0.00	30023
HH Labour - Females (hrs per Ha)	548.87	360.07	0.00	34348
HH Labour - Children (hrs per Ha)	88.05	0.00	0.00	27676
Hired Labour (Days per Ha)	7.01	0.00	0.00	3459
Free Labour (Days per Ha)	0.94	0.00	0.00	1483
Plotsize (Ha)	0.51	0.33	0.00	295
HH Agri implements index	0.11	-0.17	-2.21	29
HH Durable Goods index	-0.54	-1.13	-1.54	20
Season[b]	0.58	1.00	0.00	1
Household size	4.90	5.00	1.00	17
Education (yrs)	5.24	5.00	0.00	15
Female	0.27	0.00	0.00	1
Married	0.79	1.00	0.00	1
Age	38.21	36.00	15.00	64
Household size	4.90	5.00	1.00	17
Child dependency ratio	1.13	1.00	0.00	8
Formally Employed	0.14	0.00	0.00	1

Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010-2011.
Sample is limited to the working-age population which is age $\geq$ 15 & age $\leq$ 64.
[b] Season is a dummy variable where 0 indicates the 2008/2009 rainy season and 1 indicates the 2009/2010 rainy season.

Table 2.4: Percentage of Plots using a given Agricultural Technology

	%
Hand Hoe	99.65
Sprayer	3.92
Treadle Pump	2.70
Watering Can	29.27
Ox Plough	1.66
Tractor	0.04
Ridger	0.66
Irrigation	0.54

Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.

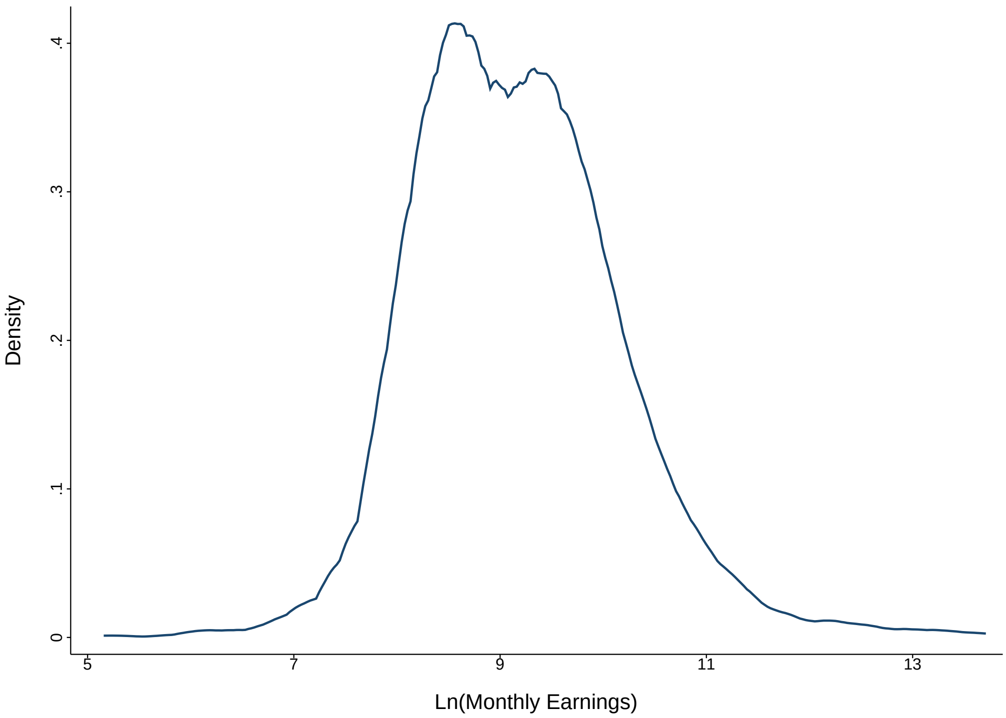
with studentized residuals above 5 or below -5. Table 2.5 provides a description of the data for formal sector employees that is used for the estimates of earnings regressions. Formal sector employees generally live in houses with fewer people and lower child dependency ratios than agricultural plot managers, while they also have higher levels of education with 8.42 years of education on average. Figure 2.2 show the distribution of formal sector sector wages.

Table 2.5: Wage Data Description

	Mean	Median	Min	Max
Education (yrs)	8.46	9.00	0.00	15.00
Earnings+	17822.21	8000.00	173.58	900000.00
Female	0.20	0.00	0.00	1.00
Married	0.77	1.00	0.00	1.00
Age	35.03	33.00	15.00	64.00
Rural	0.59	1.00	0.00	1.00
Household size	4.81	5.00	1.00	18.00
Child dependency ratio	0.85	0.67	0.00	5.00

Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.
+Earnings are displayed as monthly Malawian Kwachas earned.

Figure 2.2: Distribution of ln(Monthly Earnings (Kwachas))



Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.

2.4.2 Methodology

Instrumental Variables

A two-staged least squares (2SLS) approach is used to estimate the causal effect of education on production. The following model is estimated:

$$S_i = X_i\beta + Z_i\gamma + \mu_i \tag{2.5}$$

$$Y_i = X_i\beta + S_i\alpha + \varepsilon_i \tag{2.6}$$

where X_i is a vector of control variables for individual i , Z_i is a vector of instrumental variables, S_i is the years of schooling, Y_i is either the log of earnings or the log of yields, and μ_i and ε_i are the error terms that are assumed to be $N\left(0, \sigma^2\right)$ distributed.

Cobb-Douglas type production functions are used to model agricultural yields. The log of monetary yield is regressed on numerous inputs into the production process. Inputs into the production process include land, household and other labour, the use of fertilizer, pesticides, herbicides and capital. Capital is proxied by the two created indices for durable goods and access to farm implements.

In terms of land, it is expected that there are decreasing economies of scale. This is known as the inverse production hypothesis. The hypothesis states that crop yields - defined as the amount of crop produced per unit of land - decrease as land size increases. Larson et al. (2014) found a significant correlation of -0.349 between the log of land and the log of yields in Malawi using OLS.

To account for rainfall and other climatic factors, districts, elevation, agro-ecological zones, and season of harvest are controlled for using geographic variables from the IHS data. Apart from education, other individual characteristics that are controlled for include age, whether plot managers are married, and, gender.

With regards to gender it is expected that the coefficient on being female would be negative. Female farmers are less productive than male farmers. Female farmers have been shown to have worse access to resources and are less productive than male farmers with the resources at their disposal (Kilic et al., 2015; Oseni et al., 2015).

Depending on whether inputs are included in the model will lead to the estimation of different effects of education. Provision for the worker and allocative effect is made. When the effect of education on the total value of production is estimated while controlling for other inputs, the worker effect is estimated. Not controlling for other inputs will estimate the worker plus allocative effect. A Wald test is also done to estimate whether the coefficient on education differs significantly between the worker and allocative effect models. The null-hypothesis of the test is that the coefficients do not differ significantly from each other.

For comparison, the returns to education on the earnings of the formally employed are also estimated. This will show whether returns to education are heterogeneous across sectors. For the earnings estimates, OLS and IV models will be estimated. In the estimates, the log of monthly earnings (including monthly profits) is the dependent variable. Years of education is the independent variable of interest. The models also control for age and its square, individuals' gender and marital status, where individuals live, and sector of employment.

2.4.3 Instrumental Variables for Education

Two different instrumental variables will be used for education. The first is the introduction of Free Primary Education and the second will be the age of paternal orpanhood. The first is formed by exploiting a change in the cost of primary schooling. Between 1991 and 1994 FPE was introduced in Malawi. This shock is expected to have increased the demand for schooling which would have led to an increase in enrolment in primary schooling. A variable capturing whether individuals benefited from this policy change or not should thus be correlated with the amount of schooling received but not directly with either earnings or agricultural productivity.

The IV created to capture the effects of FPE was designed as follows. Firstly, children could get FPE whether they used it or not. This implies that cohorts of school going age (6 years) in 1991/92 and subsequent years get the treatment. Secondly, students who started school in 1991/92 or later regardless of their age also received the treatment. Finally, students who were already in higher grades in 1991/92 who potentially received some FPE due to the introduction for non-repeating girls in 1992/93 or full introduction in 1994/95, also received the treatment. This was inferred by calculating in which grade they were when FPE was introduced by subtracting their years of education from the year that they stopped school. This relies on the assumption that those students did not repeat grades during their tenure in school. Thus, anyone who could have or did receive FPE, and students who possibly received some years of FPE, were given a 1 and the rest a 0. For the regression analysis, the proportion of individuals who received partial or full FPE was approximately 23% and 30% in the agricultural and formal sectors respectively.

The second instrument that will be used is people's age when their father died. Parental death should thus be correlated with educational attainment and should not be directly correlated with productivity or earnings in the long term. Many studies find negative effects of parental death on schooling. Using Demographic and Health Survey (DHS) data for 10 countries in Southern Africa, Case et al. (2004) found that orphans were significantly less likely to go to school than non-orphans in the same households. The authors also estimated that for Malawi in 1992, 9.2% of children aged 14 and younger had lost either one or both of their parents, while in 2000 that rate was 11.7%. For the same age group, the rate of paternal orphanhood was 4.6% in 1992 and 6.5% in 2000, while for maternal orphanhood the rate was 3.0% in 1992 and 2.9% in 2000. Further, they found regressions for school enrolment in Malawi indicated that orphans were significantly less likely to be enrolled in school compared to non-orphans for the same demographic profile. The negative effects were smaller in 2000 than in 1992, and for double and paternal orphans, effects were not significantly different from 0 in 2000. This possibly indicates that FPE reduced the effects of the shock. For both years the negative effects of maternal orphanhood were greater than paternal orphanhood.

Evans and Miguel (2007), using a 5-year Kenyan panel dataset, also found a significant decrease in school participation after the death of a parent. The effect was larger for the death of mothers and also for children who had low academic performance prior to a parent's death. Both paternal and maternal orphanhood were tested as instruments in this study, but paternal orphanhood was a much stronger instrument than maternal orphanhood. This is not a common finding, however, Alam (2015) also found that paternal illness (as opposed to death) had significant effects on schooling outcomes while maternal illness did not, for both boys and girls. In both the agricultural and formal sector estimation samples, aproximately 33% of individuals lost their father before the age of 25.

Both instruments have potential reasons for being endogenous and these caveats will be taken into account in the discussion of the results. FPE could have reduced the quality of schooling by putting strain on the schooling system. Research from both Kenya (Lucas and Mbiti (2012)) and Tanzania (Valente (2015)) on the introduction of FPE notes negligible declines on test scores though. Similarly it could be argued that government spending on public education could lead to crowding out of expenditure on goods that could affect agricultural productivity. It is questionable though, how the introduction would affect smallholder farmers 15 years later. Previously, Uwaifo Oyelere (2010) has also used the length of exposure to FPE in Nigeria as instrument in measuring the returns to education.

Paternal orphanhood could be causally correlated with production if, for example, poorer fathers are at greater risk of mortality, which could be correlated with their children's production. Yet one can control for own wealth in production functions. This control would be valid if it is assumed that fathers' and their children's wealth are highly correlated. Secondly, it might affect the motivation of individuals in the short and long term. However in which direction the motivation will go is not clear. Individuals could for example become less motivated to farm given grief, or could become more motivated to succeed and provide for the future given that their father has passed away.

The variable for the age at which parents died was truncated at a value of 25. This value was given to all individuals older than 25 whose parents had not yet died. This assisted in dealing with missing values on parental death, as well as classifying those whose parents died at a later age. Death at any later age should not influence educational attainment. Possibly a few candidates would still be in tertiary education at or after the age of 25. They are, however, few and are not expected to influence the results substantially. Rigobon and Stoker (2005) showed that truncated instrumental variables do not bias estimates in general. Specifically, if an uncensored version of the variable is observed and it is a valid instrument, it still produces a valid instrument.

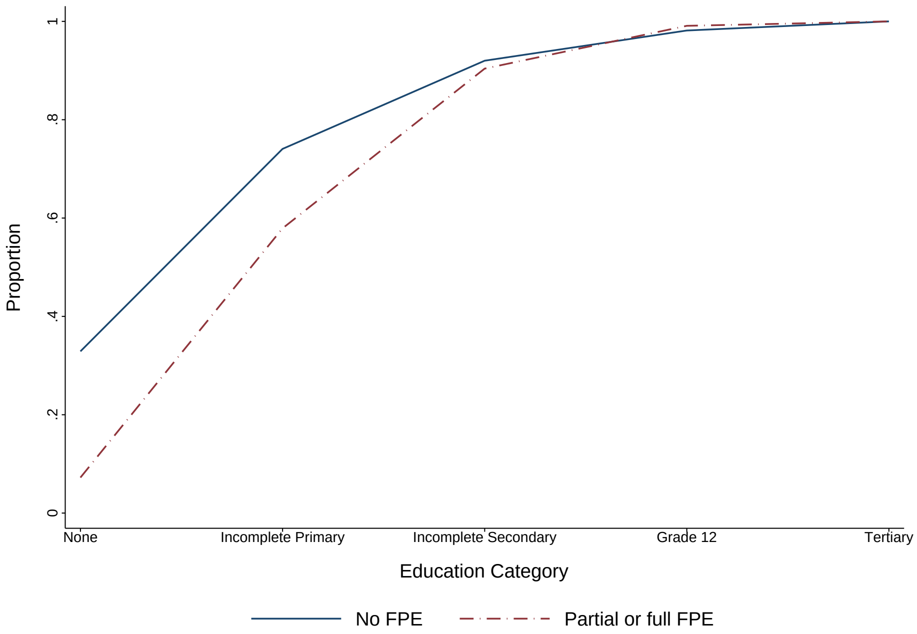
Angrist and Imbens (1995) show that where IV's have variable treatment intensity 2SLS estimates can be interpreted as an Average Causal Response (ACR). This is a weighted average of the effect of the endogenous variable on the outcome for those whose behaviour was altered because of the exogenous treatment effect, if two assumptions are satisfied. This paper will refer to the ACR as a local average treatment effect (LATE). The first is that the instruments only affect the outcome variable through the instrumented variable. The argument for the validity of the instruments was already made earlier. The second assumption is monotonicity, which implies that the different levels of treatment affect individuals in the same direction when compared to not receiving the treatment. In this example it means that receiving any FPE always increases educational attainment compared to individuals who received no FPE. For the age of paternal death it implies that education is always positively correlated with the age of paternal orphanhood.

Angrist and Imbens (1995) note that it is possible to get an idea about whether the monotonicity assumption holds by looking at whether cumulative distribution functions (CDF) of education at some level of treatment and at no treatment cross. To do this the CDF of education is estimated for different treatment groups. CDF in this case will show the proportion of individuals in the sample who have completed a certain level or less of education.

With regards to FPE a person falls either in the treatment or no treatment group. For age of paternal orphanhood individuals are categorised by different stages of school when they became orphans: before school ($age < 6$), during primary school ($6 <= age <= 14$), during secondary school ($15 <= age <= 18$), and after secondary school ($age > 18$). Figure 2.3 shows the CDF of education for those who received no FPE and those who received FPE. Figure 2.4 shows the CDF of education by different age groups when paternal orphanhood started. It is evident that FPE positively influenced the distribution of education and that the CDF lines do not cross until after primary schooling. Given that the effect of the policy should be noticeable on primary schooling, the figure suggests that the monotonicity assumption holds with regards to FPE. For the age of paternal orphanhood, the monotonicity assumption also holds. Individuals whose fathers are still alive or died after school are taken as those who received no treatment because, for them, there is no effect of paternal orphanhood on schooling outcomes. None of the CDFs for groups of individuals whose father died earlier cross with the CDF for those whose father did not die during their school going age. All of the other CDFs also lie to the left which means that paternal orphanhood during schooling appears to have had negative effects on educational attainment.

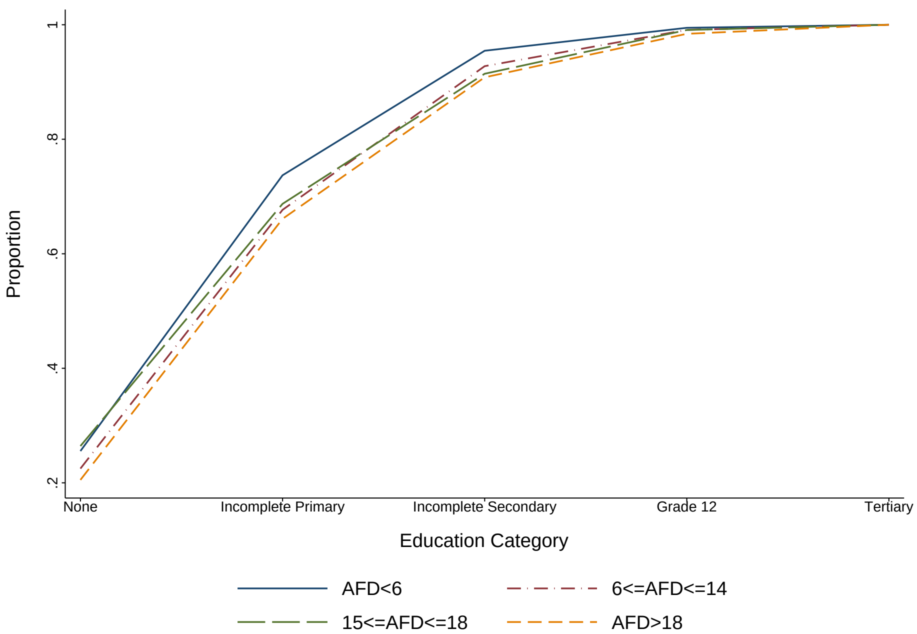
To illustrate the effects of the two instruments in more detail, differences between the CDFs of those that received no treatment and those who received certain levels of treatment are also shown. Firstly, grouping was done between those that received no FPE and those that received partial or full FPE. Within these two groups, sub-groups were identified by age of paternal orphanhood as discussed aboved. For each of these sub-groups the reference group was taken as those whose father did not die while they were at school. For example, the CDF of individuals who received no FPE and whose father died when they were younger than 6 years old is subtracted from the CDF of those who also received no FPE and

Figure 2.3: Cumulative Distribution Functions of Education for Individuals who Received No FPE and Those that Received at Least Partial FPE



Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011. AFD refers to age at which father died.

Figure 2.4: Cumulative Distribution Functions for Education for Different Age Groups of Paternal Orphanhood

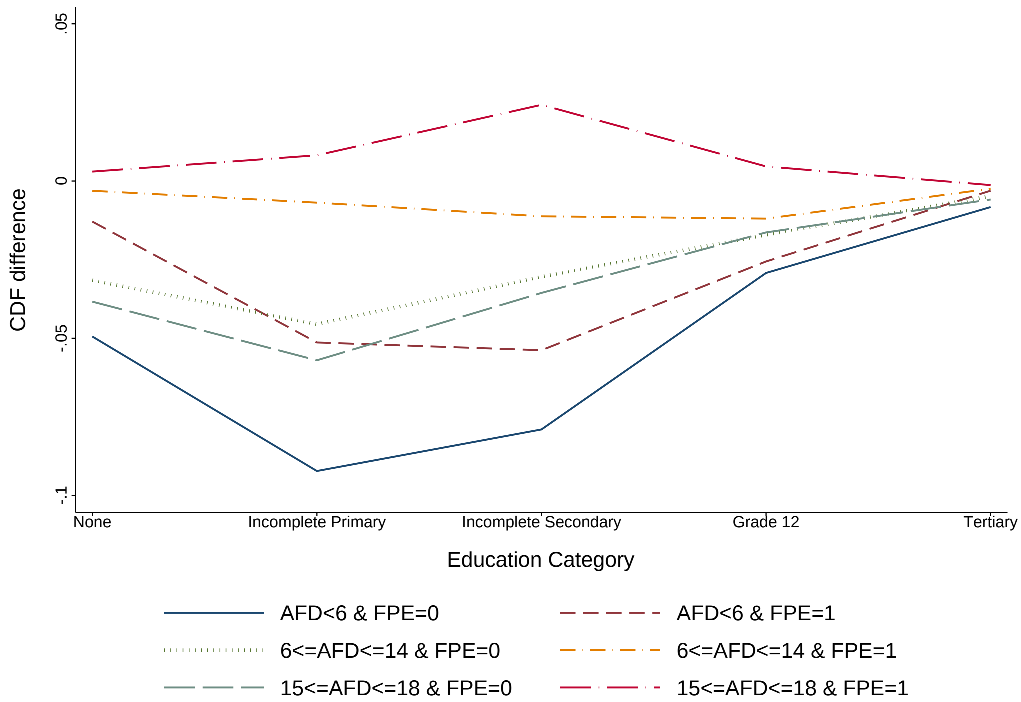


Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011. AFD refers to age at which father died.

whose father was still alive by the time that they were 19 years old. Similarly, the CDF of individuals whose father died when they were younger the 6 years old and who did receive FPE, is subtracted from the CDF of those whose father was still alive by the time that they were 19 and who received FPE. Figure 2.5 shows the differences.

From figure 2.5 it is generally clear that paternal orphanhood during or before school negatively affects schooling outcomes. It is also evident that the negative effects of paternal orphanhood are much less severe when individuals received FPE. This supports the suggestion that paternal orphanhood negatively affects education through an income shock.

Figure 2.5: IV distributional Shocks



Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.

For the respective IV's the LATE measures the effect of education on those individuals whose behaviour was altered because of FPE or paternal orphanhood. With regards to FPE, this implies that the estimate captures the effect of education for those who would never have attended school or who would have dropped out earlier if primary schooling were not free. With regards to the estimates using age of paternal orphanhood as treatment, the estimates are measuring the effects of education on those who - because of paternal orphanhood - could not go to school or had to drop out. Both represent the imposition of, or relief from, a credit constraint. FPE offered a credit relief to individuals who could not have gone to school otherwise. Paternal orphanhood imposed a credit constraint on those who had to stop schooling.

If it is assumed that credit constraints affect both groups equally, an alternative explanation would be that the instruments identify two groups with different expected returns to schooling. Students who only entered school because of FPE, would have been those who would have had low expected returns to education in the counterfactual. If they had to pay, the gains from education would not have been worth the cost. Similarly, students who had to exit the schooling system as a result of an income shock such as paternal orphanhood, would have had high counterfactual returns. They would have been individuals who did see the value of investing in education as they expected that their future returns (due to education) would be greater than the cost of the investment in education. It is expected that if this explanation holds truth, returns to education would be higher when age of paternal orphanhood is used as IV.

This paper does not address the possibility of non-linear returns to education. There could be increasing or decreasing marginal returns to education in agricultural productivity (or the formal sector). If this is the case and the LATE identifies groups that stop schooling at systematically different levels, returns could be inflated for subgroups identified by the respective instruments.

2.5 Results

As it was established in the literature review, there is not strong empirical support for the case that education has a significant influence on agricultural productivity. To establish that the IV's work as expected, the results for the returns to education in the formal sector are discussed first as it is generally accepted that there are significant returns to education in the formal economy.

To test whether the instruments are weak, the value of the first stage F-statistic of the instrumental variables is shown as proposed by Bound et al. (1995). The rule of thumb proposed by Staiger and Stock (1997) is that instruments are not weak if the F-statistic is above 10. The P-value of the Wu-Hausman test is also shown. Under the test, the efficient OLS estimate is compared to the consistent IV estimate. Under the null-hypothesis there are no systematic differences in the coefficients of the two models.

The OLS and IV results for the formal sector are shown in Table 2.6. The OLS estimate of the return to education suggests that an extra year of education is associated with a 11.3% increase in wages on average, controlling for other factors. The FPE IV estimate is insignificant while the paternal orphanhood IV suggests a significant effect of 15.4%.

Table 2.6: OLS and 2SLS Estimates on ln(Monthly Formal Sector Earnings)

			2SLS Estimates			
			FPE		AFD	
	OLS1	OLS2	1st Stage	2nd Stage	1st Stage	2nd Stage
Education (yrs)	0.141***	0.113***		0.062		0.154***
	(0.009)	(0.007)		(0.038)		(0.046)
FPE			1.315***			
			(0.205)			
Age Father Died					0.057***	
					(0.017)	
Age		0.079***	0.352***	0.089***	0.181***	0.071***
		(0.011)	(0.056)	(0.013)	(0.053)	(0.014)
Age^2		-0.001***	-0.005***	-0.001***	-0.003***	-0.001***
		(0.000)	(0.001)	(0.000)	(0.001)	(0.000)
Sector(ref = Private)		ref.	ref.	ref.	ref.	ref.
Public		0.309***	3.265***	0.479***	3.309***	0.173
		(0.045)	(0.206)	(0.131)	(0.205)	(0.159)
NGO/Church		0.276***	2.537***	0.411***	2.566***	0.168
		(0.093)	(0.328)	(0.141)	(0.335)	(0.152)
Rural		-0.037	-1.390***	-0.108	-1.402***	0.020
		(0.066)	(0.263)	(0.081)	(0.261)	(0.091)
Female		-0.014	-0.244	-0.022	-0.157	-0.007
		(0.045)	(0.248)	(0.047)	(0.261)	(0.046)
Married		0.090**	0.786***	0.133**	0.825***	0.056
		(0.045)	(0.236)	(0.058)	(0.230)	(0.059)
Household size		0.021**	-0.009	0.021**	-0.014	0.022**
		(0.009)	(0.039)	(0.009)	(0.039)	(0.010)
Child dependency ratio		-0.098***	-0.784***	-0.141***	-0.805***	-0.065
		(0.030)	(0.123)	(0.045)	(0.126)	(0.045)
District	No	Yes	Yes	Yes	Yes	Yes
Observations	2985	2985	2985	2985	2985	2985
R^2	0.334	0.454	0.292	0.422	0.286	0.434
F-stat (first stage)			33.596		14.520	
Wu-Hausman p-val				0.084		0.322

Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.
FPE uses Free Primary education as IV. AFD uses age father died as IV. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Having established that the IV's work and that education can have a significant effect in the formal sector, the agricultural production functions are presented. As a start, OLS regressions are shown in Table 2.7. The first model shows the correlation between yields and years of education when no other inputs are controlled for. The second model shows the OLS result of the combined allocative and worker effects of education. In this model personal, household, and plot characteristics are controlled for, but inputs are not. The final model also controls for inputs and gives an estimate of the worker effect of education. The difference between model 2 and 3 gives an approximation of the allocative effect of education.

Table 2.7: OLS Estimates on ln(Agricultural Yields)

	OLS1	OLS2	OLS3
Education (yrs)	0.045***	0.030***	0.010***
	(0.003)	(0.004)	(0.003)
Female		-0.122***	-0.161***
		(0.039)	(0.038)
Age		0.025***	0.024***
		(0.008)	(0.007)
Age^2		-0.000***	-0.000***
		(0.000)	(0.000)
Formally Employed		-0.012	-0.064*
		(0.042)	(0.035)
Season		0.002	0.085***
		(0.031)	(0.028)
Elevation		0.000**	0.000
		(0.000)	(0.000)
Pesticides/Herbicides			0.522***
			(0.097)
Fertilizer - Organic			0.116***
			(0.036)
Fertlizer Inorganic (ln(Kg per Ha))			0.056***
			(0.003)
HH Labour - Males (ln(hrs per Ha))			0.012***
			(0.003)
HH Labour - Females (ln(hrs per Ha))			0.020***
			(0.004)
HH Labour - Children (ln(hrs per Ha))			0.002
			(0.002)
Hired Labour (ln(Days per Ha))			0.024***
			(0.003)
Free Labour (ln(Days per Ha))			0.002
			(0.005)
Land (ln(Ha))			-0.657***
			(0.071)
Land^2 (ln(Ha))			-0.023
			(0.021)
Agri implements (access index)			0.083***
			(0.008)
HH Durable Goods index)			0.034***
			(0.007)
District	No	Yes	Yes
Agroecological Zones	No	Yes	Yes
Observations	14567	14567	14567
R^2	0.025	0.090	0.307
Allocative Effect (Wald Test):			
F-Stat			68.318***

Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The OLS estimates suggest that there is a significant and positive correlation between educational attainment and agricultural productivity. As is expected, the correlation between education and agricultural yields is larger when inputs are not controlled for. This points to an allocative effect of education: it is statistically significant as indicated by the Wald Test. Results suggest that an extra year of education is associated with approximately a 1.0% and 3.0% increase in yields, on average, when inputs are controlled for and when they or not, respectively.[6] In terms of the control variables, the results are generally consistent with what would be expected. Apart from child and free labour which are insignificant, all inputs have a positive and significant correlation with yields, while land size has a significantly negative coefficient, which is consistent with the inverse productivity hypothesis.

The 2SLS estimates are shown in Tables 2.8 and 2.9. The instruments all seem to be strong with the first-stage F-statistic well above 10. When FPE is used as IV, point estimates of the returns to education are low and insignificant. The Wu-Hausman test suggests, however, that the estimates cannot be distinguished from the more efficient OLS estimates which are still low, but significant.

When age of paternal orphanhood is used as IV, point estimates are larger that when FPE is used and they are significant. Estimates are also statistically different from the OLS estimates. The worker effect is estimated at 7.0%. When inputs are not controlled for, the coefficient is 9.3%. This indicates that education can have a significant and fairly large effect on agricultural production.

[6] It is not possible to make a direct comparison to cited studies that reported the returns to four years of education (see Appleton and Balihuta (1996) and Weir (1999)). However, as a proximation the 1 year marginal return estimated here can be multplied by 4 years. This leads to a worker effect of approximately 4.08% and an allocative effect of approximately 12.7% to four years of schooling. The calculation is: $100 \times (e^{educ*4} - 1)$

Table 2.8: 2SLS Estimates on ln(Agricultural Yields) using FPE as IV

	Worker + Allocative		Worker	
	1st Stage	2nd Stage	1st Stage	2nd Stage
Education (yrs)		0.014		-0.001
		(0.016)		(0.016)
FPE	2.120***		1.973***	
	(0.141)		(0.135)	
Female	-1.211***	-0.139***	-1.472***	-0.176***
	(0.131)	(0.044)	(0.136)	(0.044)
Age	0.180***	0.024***	0.140***	0.023***
	(0.034)	(0.008)	(0.031)	(0.007)
Age^2	-0.003***	-0.000***	-0.002***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Formally Employed	2.159***	0.022	1.421***	-0.049
	(0.170)	(0.055)	(0.137)	(0.042)
Season	0.070	0.004	0.121	0.087***
	(0.114)	(0.031)	(0.105)	(0.028)
Elevation	-0.000	0.000**	-0.001	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Used Pesticides/Herbicides			0.769***	0.530***
			(0.231)	(0.098)
Ised Fertilizer - Organic			0.135	0.118***
			(0.132)	(0.036)
Fertlizer Inorganic (ln(Kg per Ha))			0.043***	0.056***
			(0.007)	(0.003)
HH Labour - Males (ln(hrs per Ha))			-0.019*	0.012***
			(0.011)	(0.003)
HH Labour - Females (ln(hrs per Ha))			0.005	0.020***
			(0.014)	(0.004)
HH Labour - Children (ln(hrs per Ha))			-0.007	0.002
			(0.009)	(0.002)
Hired Labour (ln(Days per Ha))			0.088***	0.025***
			(0.010)	(0.003)
Free Labour (ln(Days per Ha))			0.001	0.002
			(0.025)	(0.005)
Land (ln(Ha))			-0.123	-0.658***
			(0.090)	(0.071)
$Land^2$ (ln(Ha))			0.006	-0.022
			(0.026)	(0.021)
Agri implements (access index)			-0.038	0.082***
			(0.029)	(0.008)
HH Durable Goods index)			0.560***	0.041***
			(0.026)	(0.012)
District	Yes	Yes	Yes	Yes
Agroecological Zones	Yes	Yes	Yes	Yes
Observations	14567	14567	14567	14567
R^2	0.270	0.088	0.351	0.306
F-stat (first-stage)	222.092		211.449	
Hausman test (p-val)		0.239		0.392

Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.
FPE uses Free Primary education as IV. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.9: 2SLS Estimates on ln(Agricultural Yields) using Age of Paternal Orphanhood as IV

	Worker + Allocative		Worker	
	1st Stage	2nd Stage	1st Stage	2nd Stage
Education (yrs)		0.093**		0.070*
		(0.046)		(0.042)
Age Father Died	0.043***		0.041***	
	(0.007)		(0.006)	
Female	-1.097***	-0.051	-1.377***	-0.077
	(0.135)	(0.063)	(0.140)	(0.069)
Age	-0.083***	0.030***	-0.104***	0.029***
	(0.029)	(0.009)	(0.026)	(0.008)
Age^2	0.000	-0.000***	0.000	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Formally Employed	2.168***	-0.150	1.413***	-0.150**
	(0.174)	(0.107)	(0.139)	(0.071)
Season	0.067	-0.003	0.122	0.077***
	(0.118)	(0.032)	(0.107)	(0.029)
Elevation	-0.000	0.000**	-0.001	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Pesticides/Herbicides			0.797***	0.475***
			(0.239)	(0.104)
Fertilizer - Organic			0.147	0.106***
			(0.131)	(0.036)
Fertlizer Inorganic (ln(Kg per Ha))			0.044***	0.053***
			(0.007)	(0.003)
HH Labour - Males (ln(hrs per Ha))			-0.021**	0.013***
			(0.011)	(0.003)
HH Labour - Females (ln(hrs per Ha))			0.005	0.020***
			(0.015)	(0.004)
HH Labour - Children (ln(hrs per Ha))			-0.009	0.002
			(0.009)	(0.002)
Hired Labour (ln(Days per Ha))			0.087***	0.019***
			(0.010)	(0.005)
Free Labour (ln(Days per Ha))			0.010	0.001
			(0.025)	(0.005)
Land (ln(Ha))			-0.125	-0.649***
			(0.091)	(0.067)
$Land^2$ (ln(Ha))			0.006	-0.023
			(0.025)	(0.020)
Agri implements (access index)			-0.039	0.085***
			(0.030)	(0.008)
HH Durable Goods index)			0.572***	0.000
			(0.025)	(0.024)
District	Yes	Yes	Yes	Yes
Agroecological Zones	Yes	Yes	Yes	Yes
Observations	14567	14567	14567	14567
R^2	0.248	0.052	0.333	0.277
F-stat (first stage)	41.050		42.865	
Hausman test (p-val)		0.049		0.041

Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The results show that returns to education are substantially lower in agriculture than in the formal sector. One possible reason for the low and inefficient coefficients is attenuation bias due to measurement error. There is no way of assessing the extent of possible measurement error however. In an attempt to gain more efficient estimators the IV models were also estimated using bootstrapped (with 500 repetitions) standard errors. Secondly, a sub-sample analysis was done by limiting the sample to farmers younger than 40. This was done to capture the educational shock of FPE better, by limiting confounding effects of earlier educational shocks. The results for these robustness checks are shown in Table 2.10. None of the estimates gained efficiency. Furthermore, the worker effect when using age of paternal orphanhood as IV becomes insignificant with bootstrapping. The sub-sample estimates did not differ significantly from the full-sample either.

Table 2.10: Robustness Checks

	IV	Check	Education (yrs)	Standard Error	Observations	F-Stat (first stage)
Worker + Allocative	FPE	Bootstrap	0.014	0.016	14567	310.541
		Age $\leq$ 40	0.012	0.016	8662	259.621
		Bootstrap + Age $\leq$ 40	0.012	0.016	8662	388.996
	AFD	Bootstrap	0.093**	0.046	14567	57.035
		Age $\leq$ 40	0.087	0.067	8662	17.844
		Bootstrap + Age $\leq$ 40	0.087	0.078	8662	24.134
Worker	FPE	Bootstrap	-0.001	0.016	14567	302.178
		Age $\leq$ 40	-0.005	0.015	8662	246.538
		Bootstrap + Age $\leq$ 40	-0.005	0.015	8662	367.429
	AFD	Bootstrap	0.07	0.044	14567	60.891
		Age $\leq$ 40	0.078	0.06	8662	18.05
		Bootstrap + Age $\leq$ 40	0.078	0.073	8662	25.24

Source: Own Calculations using Malawi's Third Integrated Household Survey, 2010/2011.
FPE uses Free Primary education as IV. AFD uses age of paternal orphanhood as IV. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.6 Discussion

In general, returns to education in agriculture are low but it was shown (when age of paternal orphanhood was used as IV) that education can have a significant effect on agricultural production. The results, however, also provide some further insights. To start, reasons for the lower returns in agriculture and higher returns in the formal sector are discussed. This is followed by a discussion on what the insignificant returns for the individuals entering school due to FPE imply.

2.6.1 Sectoral Differences in Returns to Education

If statistical significance is disregarded and focus is solely placed on point estimates, returns to education for both groups (represented by the two IVs) are lower in agriculture than in the formal sector. This supports the theoretical predictions by Welch (1970), Schultz (1975) and Ram (1980) that education has more value in modernizing environments. As noted earlier, the use of modern agricultural technologies, apart from inorganic fertilizer (much of it subsidized), was negligible.

The difference in returns between the two sectors can also be explained by two other factors. The first is that education acts in part as a screening mechanism in the formal sector. This is supported by the fact that returns in agricultural production, a more pure measure of productivity than wages, is low. Secondly, it is plausible that the education curriculum equips learners with skills more useful in the formal sector than in agriculture. Both these possibilities deserve further study.

2.6.2 The Quantity-quality Trade-off in Education

In both the formal sector and agricultural production, point estimates were lower when FPE was used as IV than when age of paternal orphanhood was used. Point estimates are also insignificant when FPE is used as IV. Thus, the findings suggest another explanation for why economic growth has not accompanied large improvements in access to education (see Pritchett (2001)). While quality has been stressed as one explanation, this study's results suggests two other possibilities.

Firstly, returns to education in agricultural production, where most of the employed in developing countries work, is low. Thus, increased access would not lead to high economic growth if students end up working in smallholder agriculture. Secondly, if the LATE explanation that the subgroup identified by FPE had low expected returns is to be believed the effect of increased enrolment on economic growth would also be diminished. This narrative is supported by evidence from Lucas and Mbiti (2012) and Valente (2015) who studied the effects of FPE on test scores.

Lucas and Mbiti (2012) studied the effect of FPE on test scores in Kenya of children who were already in school for three to seven years by the time it was introduced. FPE put large constraints on the

Kenyan schooling system because of large increases in enrolments without a similar expansion in schools and teachers. Average test scores for the final primary schooling exam in Kenya did decline after the introduction of FPE, but the authors found that FPE only slightly decreased the test scores of learners that would have written the test in any case. The decline in average scores was driven by those who entered the schooling system after FPE was introduced. Valente (2015) analysed the effects on test scores of new entrants in the schooling system after introduction in Tanzania and found similar results. They did however find declines in test scores for students (already in school) in urban schools with high baseline test scores where there were large increases in enrolment.

Indeed, Taylor and Spaull (2013) suggested that looking at enrolment rates and average test scores in isolation are not good indicators of school quality. As a better measure of school quality, they suggested looking at the proportion of learners within a specific age group that reached certain levels of numeracy and literacy. With this measure, they found evidence that the generally accepted trade-off between quality and access in Africa is not as evident as has been believed.

2.7 Conclusion

Malawi is a country with high levels of poverty where a large proportion of the employed work in subsistence agriculture. Productivity levels in the sector are low and increasing it would aid in decreasing poverty in the country. This chapter has studied the effects of one possible channel to achieve this - education. In the literature it was found that the empirical evidence generally only finds small positive effects for education in agricultural production. Some evidence shows that education makes individuals more likely to use technologies and also that education has larger effects in technologically advanced settings. The one problem with much of the evidence, especially in production estimates, is the endogeneity of education in the production function.

This chapter fills this gap by causally estimating the effect of education on agricultural production in Malawi. Causality is established by using an instrumental variable approach. The introduction of free primary education in the early 1990's is used as one instrument and the age of paternal orphanhood is used as a second. The instruments are shown to estimate LATE effects for two distinct groups. The FPE IV estimates returns to education for individuals who only entered school because it became free. Similarly, age of paternal orphanhood, estimates the returns for individuals who had to exit school due to paternal orphanhood.

When FPE is used as IV, point estimates are low and insignificant but similar to the significant OLS estimates. When age of paternal orphanhood is used as IV, point estimates are higher and statistically significant. This suggests that education can have an effect on agricultural production but this is not a given. Estimates suggest higher returns to education in the formal sector than in agricultural production, however. This, combined with the fact that the use of modern technologies in Malawi is negligible, supports suggestions that education is indeed more valuable in technologically advanced settings.

Finally, the low returns to education for individuals affected by FPE, suggests that large expansions in education did not have the desired effect on a subgroup of individuals at whom the expansions were aimed. Access to education was not sufficient to increase their agricultural productivity and, thus, lift them out of poverty. Access to better technologies is also needed and will compliment education as a poverty alleviation strategy. Furthermore, expansions of the formal sector also hold promise in giving educated individuals opportunities to move out of poverty.

Chapter 3

Using Satellite Data to Measure Socio-Economic Outcomes: A Case Study of Namibia

3.1 Introduction

As highlighted in chapter 1, a lack of household survey data limits what can be said about economic development in much of sub-Saharan Africa (SSA). Two ways in which household survey data limits analysis, is the frequency of its release and the geographic scale at which it is representative. For example, surveys used to measure poverty and inequality, generally called Income and Expenditure Surveys, are only undertaken around every 5 years and are not longitudinal. These surveys are also only representative at the first sub-level of geographic administration in most countries.

These factors limit what researchers can say about poverty changes over shorter time periods, and also about the geographic location of poor households. Small-area estimation techniques, such as the method proposed by Elbers et al. (2003) do allow researchers to map poverty at smaller geographic areas, but the method can only map poverty to smaller areas for the same time periods as household surveys, with the added requirement that census data (which is more spatially disaggregated) be released around the same time.

Realising these limitations, researchers have started to investigate alternative data sources, such as mobile phone data, as additional sources to support analysis. Blumenstock et al. (2015) have shown, for example, how people's record of cellphone data usage can be used to predict poverty.

Another alternative is to use satellite data. Satellite data is recorded at high frequency and at spatial resolutions usually much smaller than the smallest statistically representative area in a household survey or census. Currently there are many satellite series available free of charge that go back for decades. These properties make them highly desirable. The process of capturing data with satellites falls under the broader science of remote sensing and this chapter will refer to such data as remotely sensed (RS) data.

In terms of RS data, the night lights series is well-known in economics, and changes in the amount of light emitted in an area have been shown to correlate well with changes in GDP per capita (Henderson et al., 2012). It has been noted, however, that the night lights data adds more value as a proxy in countries with low levels of statistical capacity (Chen and Nordhaus, 2011; Henderson et al., 2012). It is not unreasonable to assume that countries with low levels of statistical capacity are also under-electrified. Thus, while at national level, night lights might correlate with GDP, the ability of night lights to proxy any economic activity at smaller geographic levels within these countries needs to be questioned.

The question that arises then, is whether there is a RS data source that can be used in underdeveloped areas where no light is observed? Research has shown that daytime RS data does hold promise. Weeks et al. (2012), Watmough et al. (2013) and Watmough et al. (2016) have all noted that measures of land cover and vegetation quality correlate with socio-economic outcomes between areas. Combinations of daytime RS data and night lights have also been used to map poverty in a number of African countries (Jean et al., 2016) and GDP in Guangdong, China (Cao et al., 2016).

There is, however, a fundamental shortcoming in the literature which this chapter addresses. No research has focussed on the relationship between daytime RS data and socio-economic outcomes over time. This is an important omission as cross-sectional correlations can be misleading in making causal claims. Selection effects, for example, need to be considered. In SSA, the poor are more likely to be located in rural areas where vegetation quality is higher. However, it cannot be assumed that increases in vegetation quality lead to increases in poverty over time. Differences in agricultural practices, agro-ecological zones and general use of land also implies that relationships between the remotely sensed environment and socio-economic outcomes would be heterogeneous across regions. Henderson et al. (2012) note regional heterogeneity as the reason for only using growth formulations in examining the relationship between night lights and GDP. The establishment of plausible relationships between daytime RS data and welfare over time, would imply that welfare can be tracked within smaller regions and over higher frequencies than household surveys allow and can do this where night lights are not present.

This chapter proposes that measures of vegetation quality, derived from RS data, can be used to track changes in socio-economic outcomes over time in rural regions in SSA. The reasoning is that most under-developed, inhabited areas are dominated by smallholder agriculture, and RS based vegetation indices are widely used to predict agricultural yields (Lobell, 2013). This implies that RS data can proxy for an important source of economic activity in such regions. Thus, welfare should increase in regions as vegetation quality improves.

This chapter shows how existing data sources can be extended to inform on welfare. It presents results from a small-scale study that explores these ideas. Namibia was selected as the country of study because it has large regional inequalities. It has well-developed urban regions, as well as underdeveloped agricultural regions where poverty is prevalent and electrification rates are low.

The chapter firstly shows, using classification of cropland, that in rural agricultural areas, daytime satellite data can indeed identify economic activity where night lights cannot. Secondly, using regression analysis it explores the correlation across time and space between vegetation quality and welfare. The results show that welfare correlates negatively with vegetation quality across geographic areas. Within rural areas, however, vegetation quality correlates positively with welfare over time. This supports the idea that RS based vegetation indices can track welfare over time and map welfare where night lights cannot.

The next section gives a review of the use of satellite data in economic studies. Section 3.3 discusses the data and methodology used. Section 3.4 presents the results. Section 3.5 concludes.

3.2 Economic Studies and Satellite data

The appeal of RS data was noted in the introduction - broad coverage at high spatial resolution and high frequency. Satellites orbit the earth and capture light reflected to it from the earth's surface. This process of capturing light from a distance is known as remote sensing. The light captured is stored and digitized to create an image with geo-referenced points that contain data of some sort.

Daytime RS data captures electro-magnetic radiation (EMR) from the sun, reflected back off the earth's surface. The EMR is captured by sensors on the satellites which store them in different spectral bands depending on their wavelengths. These sensors generally capture a broader spectrum of EMR than what the human eye can capture.[1]

Each object on the earth's surface absorbs and reflects different wavelengths and these properties can allow researchers to identify different objects based on their spectral signatures - how they reflect EMR. Vegetation for example, absorbs EMR in the red band and reflects EMR in the infra-red band. Ratios of the EMR reflected in these two bands can also give indications of the health of vegetation. One popular measure of vegetative health, the Normalised Differenced Vegetation Index (NDVI), is discussed in Section 3.3.

Satellites capturing images of the earth at night work in the same manner, with the exception that much of the light captured is artificially created by humans and not reflections from the sun. The night lights series made available by the Defence Meteorological Satellite Program (DMSP) is a free, user-friendly source and is widely used in economic studies.

[1] The human eye captures the visible spectrum which includes the blue, green and red bands of EMR.

3.2.1 The Use of Night Lights Data

Elvidge et al. (1997) showed that the area lit up in countries was positively correlated with gross domestic product (GDP) and electricity consumption, suggesting that the data could be used to model GDP. The correlation between lights and GDP at national level led researchers to conclusions that the data could be used to obtain estimates of GDP at sub-national levels. Sutton and Costanza (2002) modelled the relationship between total light emitted in a country and GDP at the national level, and created a 1km^2 grid of GDP for the world. Ebener et al. (2005) and Sutton et al. (2007) showed that night lights can be used to model GDP per capita at sub-national levels.

Chen and Nordhaus (2011) compared sub-national GDP estimates based on night lights with estimates based on the G-Econ data which is a data product with small-area GDP estimates.[2] They found that night lights data added valuable information in countries with low levels of statistical capacity, but not in countries with good capacity. This was due to the fact that measurement error in the night lights data is larger than standard economic data.

Henderson et al. (2012) showed how *growth* in luminosity was a good proxy of GDP *growth* in the long- and short-run. They then provided a revised measure of income growth as a combination of national accounts data and predicted income growth from night lights. They calculated the optimal weights to be placed on each measure. Similar to Chen and Nordhaus (2011), they found that in countries with good statistical capacity, night lights did not add value to the national accounts data, but in countries with low capacity, the optimal weighting was roughly an equal combination of national accounts and night lights.

One example of a study exploiting the correlations between GDP and night lights is Villa (2016). He used the night lights as a sub-national proxy of economic activity to measure the effect of a cash transfer programme on economic growth in Colombia. The program was phased in over time at the municipal level allowing for the implementation of a difference-in-difference identification strategy. Night lights data was used to provide estimates of GDP at municipal level.

Pinkovskiy and Sala-i-Martin (2016) used night lights as an independent source of data to study whether national accounts or survey data provide more reliable estimates of poverty. They estimated a proxy for income by calculating the optimal weights to be placed on national accounts and survey estimates, by exploiting the assumption that the measurement error in night lights is unrelated to the errors of national accounts and survey data. They concluded that national accounts provide more reliable estimates of true income than survey data.

Night lights have also been used to create and map global poverty indices. Elvidge et al. (2009) estimated a global poverty map using night lights data. They created a wealth index based on gridded population data and night lights. Elvidge et al. (2012) developed a Night lights Development Index (NDLI) combining

[2] For more information on the G-Econ data visit http://gecon.yale.edu/.

night lights and gridded population data. Anthony (2015) also calculated regional inequality measures by combining night lights data, with gridded population data and the G-Econ data.[3]

3.2.2 The Use of Daytime Remotely Sensed Data

Remotely sensed data has been used substantially in the studies on the trade-off between development and the environment, commonly referred to as the environmental Kuznets curve. These studies combine data on economic growth or changes in welfare with satellite data measuring environmental outcomes such as the area of forest cover. Pfaff (1999) has studied this for the Brazilian Amazon and Foster and Rosenzweig (2003) as well as Bhattacharya and Innes (2012) have done such studies in India. For this chapter, the findings are not of concern, but the measures from satellite data, in combination with survey data, are.

Pfaff (1999) used Landsat images classified by Skole and Tucker (1993). The images were classified into different categories of forest using visual interpretation. This was combined with Brazilian county level data obtained from a number of sources to estimate relationships between deforestation and socio-economic variables.

Foster and Rosenzweig (2003) combined census, household survey and RS data spanning 29 years to study the relationship between income and forest cover around villages in India. They used the NDVI to measure forest cover in a 10km radius from villages. The proportion of pixels with NDVI values above 0.2 multiplied by the average NDVI of those pixels was used as their measure of forest cover. Bhattacharya and Innes (2012) studied the relationship using district level data. For environmental quality they also used NDVI. They used average NDVI over the districts as a measure of total biomass. They also constructed a proxy for forest cover based on the number of pixels that are above a certain NDVI threshold on average over a month.

Klemens et al. (2015) tested whether RS data can help explain poverty at small-area level. They used local-level census and survey data to model poverty rates at local administrative levels. They then added variables derived from satellite data to explore whether they help explain local poverty rates. Night lights were found to be significant and add explanatory power in explaining variation in rural but (surprisingly) not urban poverty rates. Measures of local vegetation were insignificant, however. For their measures of vegetation, they used the Leaf Area Index - a measure of leaf coverage - and the fraction of photosynthetically active radiation (FPAR) - which measures how much incoming radiation from the sun is absorbed by plants.

Imran et al. (2014) use RS data to map poverty in Burkina Faso. They started at the observation that poor households in rural areas produce too little to meet their consumption needs. Using household level

[3] For gridded population data the LandScan series was used in all three studies. The data is created by combining census population numbers with spatial data to create population counts which reflect where people are located during the day. This is referred to as the ambient population.

data, representative at community level, they created a communal level asset index from variables related to food production. They then identified variables from available remote sensing data that represent agro-ecological stressors which affect food production. Variables included NDVI, rainfall data, topographical data, length of the growing season, livestock data, distance to markets and population density. At community level, these variables were regressed on a communal asset index to model welfare. This model was used to predict pixel level welfare in the country.

Weeks et al. (2012) explore the use of RS data to map health outcomes within cities. They argue that health outcomes are correlated within neighbourhoods, and that by combining data on health outcomes with neighbourhood characteristics from remotely sensed data, health outcomes can be mapped for cities. They used Ghana census data for Accra, with geo-references to neighbourhoods in the city. They created a housing index for each neighbourhood based on the census data. They then used RS data to map land cover in the city. They found that neighbourhoods with the lowest quality housing generally also had the least vegetation cover and the worst health outcomes while neighbourhoods with high quality housing had more vegetation cover and better health outcomes. This, they argue, implies that RS data can be used to identify areas of cities with bad health outcomes.

Watmough et al. (2013) study whether RS data could be used to predict female literacy levels in Assam, India. They classified RS images into nine different classes. Census data was used to estimate the level of female literacy in villages. They found that female literacy was positively correlated with the amount of woodland in an area and negatively correlated with the amount of winter cropland. Watmough et al. (2016) expanded on the previous analysis by estimating the relationship between remotely sensed data and a general welfare index derived from variables in census data. Land cover classes and NDVI were regressed on welfare. Woodland cover was again positively correlated with welfare. Surprisingly, NDVI was not an important factor but the authors note that this was possibly due to a high correlation with woodland cover.

Morikawa (2014) used daytime RS data to explore the effects that a poverty alleviation project had on communities and the environment. Based on the idea that poverty and land degradation reinforce each other, he used NDVI to evaluate the environmental effects of the project.

Researchers have also started to combine daytime and night lights imagery. Jean et al. (2016) used a machine learning technique to combine daytime RS data and night lights data to explain a large proportion of regional variation in poverty for a number of countries in SSA. Cao et al. (2016) produced a GDP map of Guangdong, China, using night lights to proxy the secondary and tertiary sectors, and land use data for the primary sector. Additionally they used NDVI data to limit the effect of light overflow in the night lights data. Without addressing light overflow, the extent of urban areas, based on night lights, is overestimated. (Small et al., 2011).

3.3 Data and Methodology

The reviewed literature has shown that daytime RS data can be used to predict socio-economic outcomes. However, the research fails to address relationships between socio-economic outcomes and daytime RS data over time. Thus, while the research can help to show where poverty is located, it cannot say how welfare will change as the RS variables change. This is an important omission as regional heterogeneity and selection effects could lead to misleading correlations and false causal interpretations. This argument was also made with regards to night lights by Henderson et al. (2012). To estimate the relationship between night lights and GDP, they argued that cross-country heterogeneity in, for example, the nature of production and lighting technology made them stick to growth formulations in their models.

This chapter addresses this limitation by exploring whether RS data can be used to proxy welfare, with the added focus of variation over time. For this, it is argued that there is potential for the use of remotely sensed indicators of vegetation quality in agriculturally dominant regions. This argument is based on the fact that indicators of vegetation quality are used substantially to predict agricultural harvests (Lobell, 2013). Welfare should thus increase with increases in vegetation quality over time because livelihoods are dependant on it. In SSA, a large proportion of the population lives in rural areas, and thus RS based vegetation indicators can hold potential to track welfare for a large proportion of the population. It can also do so in under-electrified regions.

The added benefit of vegetation indices is that there are many user-friendly, easily obtainable sources available. If indeed daytime remotely sensed data correlates well, it holds potential for tracking outcomes at small-area level over time. This small scale-study explores whether RS based vegetation indices hold promise. If the results hold, a larger scale study should be undertaken.

3.3.1 Methodology

The analysis proceeds in two steps. Firstly, visual exploration is used, to identify whether daytime RS data does capture economic activity in an under-developed, agriculturally dominant area better than night lights. This is done by comparing the amount of cropland observed from daytime RS data with the amount of night light observed in both an industrialised developed region, and a under-developed agriculturally dominant region.

After establishing that daytime RS data does indeed capture economic activity in under-developed areas better than luminosity data, it is explored whether correlations exist between indicators of vegetation quality and welfare. Regional poverty rates and average expenditure levels are regressed on vegetation quality both cross-sectionally and over time using regional fixed effects. The use of fixed effects implies that the estimates will measure the relationship using deviations over time. It is expected that, cross-sectionally, higher levels of vegetation quality should be negatively correlated with welfare. This is because poverty is higher in rural areas. However, using regional fixed effects, which implies only using variation in vegetation quality over time within regions, the correlation between welfare and vegetative health should be positive. Thus, as vegetation quality improves, which implies higher levels of agricultural production, welfare should increase.

It would be ideal to also explore the vegetation health of land specifically identified as cropland. The process of land classification is tedious, however. Raw RS images need to be processed in a number of ways before they can be classified, and classification itself is also a timely proses. For these reasons vegetation quality for all land in an area is used.

The focus of the study is to establish whether indicators of vegetation quality from RS data correlate with welfare and not to determine the model with the highest explanatory power to predict welfare with vegetation quality. As noted, the relationship will also be heterogeneous over different ecological zones which implies that the optimal measure would differ across regions. The discussion proceeds by describing the study area, followed by the data sources and vegetation indicators used.

3.3.2 Study Area

Namibia was selected as a country of study as it has large regional differences in terms of socio-economic outcomes.[4] This presents the perfect opportunity to highlight that cross-sectional and time-series relationships can differ. Namibia has well-developed urban regions and under-developed rural regions. This should show that vegetation quality correlates negatively with welfare across regions. This should not be the case within regions over time.

Namibia is one of the least densely populated countries in the world and is the driest country in sub-Saharan Africa (Devereux and Naeraa, 1996; Fara, 2001). The north of Namibia is more densely populated than the rest of the country, with 6 regions covering an area of less than 15% (Devereux and Naeraa, 1996) that held more than 50% of the population in 2011 (Own calculations using Census 2011). The areas are Omusati, Oshana, Ohangwena, Oshikoto, Okavango and Caprivi (see Figure 3.1). The country has a diverse population of roughly 2.5 million people and is defined as an upper middle income country by the World Bank (2015b). This hides the fact that it is one of the most unequal countries in the world, with a large share of the population impoverished due to a history of colonialism and racial discrimination.

[4] See Section 3.4.1 for descriptive statistics of regional differences.

Figure 3.1: Regions of Namibia

Colonial Rule

After nearly 90 years of colonial rule, Namibia gained independence in 1990. The period of colonial rule was violent. After the Herero and Nama wars between 1904 and 1908 Germany gained greater control of Namibia. German rule ended during World War 1 after occupation by South Africa in 1915. South Africa gained administrative control in 1920 (World Bank, 2008). In 1966 South Africa rejected a United Nations (UN) mandate to place the country under a trusteeship arrangement (thus continuing to rule the country) and as a result the UN cancelled South Africa's mandate.

In that same year the South West Africa People's Organization (SWAPO) declared war to free Namibia from colonial rule. SWAPO established bases in southern Angola after Angola gained independence in 1975. The organisation used guerilla tactics and northern Namibia became the centre of a war between South Africa and SWAPO. A UN-sponsored peace deal was finally reached in 1989 when Cuban troops who were supporting SWAPO left Angola and South African troops left Namibia. Elections were held in 1989, which SWAPO won easily. A constitution was adopted in February 1990 and in March independence was granted to the country (World Bank, 2008).

Under South African rule and just as in South Africa, apartheid was also enforced in Namibia. Under German administration the country was divided into two areas - the "Police Zone" and the rest. The "Police Zone" was the area of land that was vacated by white farmers. Its northern border stretched across the north-centre of the country from east to west. Essentially only areas south of Namibia's northern border were excluded from the "Police Zone". Under South African rule these policies were continued. In the “Police Zone” the government created “native reserves” or “homelands” where many natives were sent after being dispossessed of their land in favour of white farmers. The government greatly supported white farmers by, for example, granting loans, drilling boreholes, and assisting with drought relief. On the other hand, the government did little to support farmers in the “homelands”. In 1968 more segregation laws were instated with the government establishing 10 “homelands”. These areas were also granted self-rule. These included areas north of the police zone (Odendaal, 2011). Werner (1993) points out that the reserves were essentially created to provide labour to the colonial economy.

Current Economic Climate

The policies led to a highly unequal society in Namibia. The north of the country, which contains a large share of the population, is dominated by smallholder agriculture on communal lands while urban centres such as Windhoek (in Khomas), Walvisbay and Swakopmund (both in Erongo) are well developed and industrialised. Poverty is lower in the urban centres than in the communal lands. In 2009/2010 the rural poverty rate was 37.4% while the urban poverty rate was 14.6%. Since independence, poverty has declined in the country. Between 2003/04 and 2009/10, poverty decreased by 9 percentage points to 28.7% (NSA, 2012).

3.3.3 Data

3.3.4 Socio-Economic Data

Two sources of socio-economic data are used in the analysis. The first is Namibia’s Household Income and Expenditure Surveys (NHIES). NHIES data for 2003/04 and 2009/10 is used. The surveys serve as the official source of poverty estimates in the country. In Namibia, poverty is officially defined using per adult equivalent household expenditure and the upper poverty line is equal to N$4535.52 (±US$533.59) per year in 2009 prices (NSA, 2012). Both poverty rates and average regional per adult equivalent expenditure is used as welfare measures in the analysis.

One problem with the NHIES data is that it is only representative for large administrative regions. At this level of aggregation, the benefits of daytime RS data over night lights will not necessarily be highlighted as these regions all contain urban centres which will be electrified. However, daytime RS data will provide benefits at area level where night lights are not present. Regional level regression analysis would also only be based on 13 observations (see Figure 3.1) which also leaves the statistical analysis with little power.

Thus, the second source of data is constituency level poverty rates for 2001 and 2011 which were estimated using the small-area estimation (NPC, 2015). These rates were estimated using the method proposed by Elbers et al. (2003). Consumption was modelled in the NHIES surveys using variables also found in Namibia’s census data and using these models, consumption was predicted in the census data which is geographically more refined. NHIES 2003/04 was used to impute consumption levels in the 2001 Census and NHIES 2009/10 was used to predict consumption levels in the 2011 Census. The 2011 Census is also used to provide descriptive statistics in this study. With regards to regions of Namibia, note that after the release of the 2011 Census, Kavango was divided into two regions - Kavango East and Kavango West. The name of Caprivi was also changed to Zambezi. The analysis presented here uses regions as defined in the 2011 Census and surveys prior to that.

3.3.5 Remotely Sensed Data

Indicators of Vegetation Quality

Various measures of vegetation quality exist, some of which were briefly referred to in Section 3.2 when discussing other studies. This study uses the Normalised Differenced Vegetation Index (NDVI) for the main results. NDVI correlates well with a number of vegetation properties which makes crop yield forecasting possible (Huang and Han, 2014). It is also the index used most extensively to forecast yields (Mkhabela et al., 2011). It is calculated as a ratio of two bands of EMR reflectance: the red (R) and near-infrared (NIR) bands. It is based on the fact that healthy vegetation absorbs EMR in the red band and reflects EMR in the infrared band. It is calculated (Myneni and Hall, 1995) as follows:

$$NDVI = \frac{NIR - R}{NIR + R}. \tag{3.1}$$

The index ranges between -1 and 1 and it is positively correlated with vegetative health. NDVI for water bodies is negative while rocks and bare soil have values close to zero (Mkhabela et al., 2011).

In the prediction of crop yields, NDVI has been incorporated in various different ways. Research varies on the period prior to harvest for which to include NDVI, and also, whether averages or sums (integrals) over an identified period should be used. Mkhabela et al. (2005) tested various specifications in Swaziland for the prediction of maize yields, and was guided by the explanatory power of different models to identify the best specification of NDVI. Explanatory power of models also differed across regions in the country. Huang and Han (2014) provide an overview of how studies have used NDVI for prediction of yields for a range of crops in different settings. The fact that this study only attempts to establish plausible relationships between vegetation and welfare implies that there is no attempt to find the measure with the highest explanatory power.

As a robustness check results are also shown for series based on the Enhanced Vegetation Index. The EVI is an alternative to NDVI. It was developed to improve vegetation monitoring by removing atmospheric influences and canopy background signals and it is more sensitive in areas with high biomass (Huete et al., 2002).[5]

For this study we use the 16-day composite product from the MODIS Terra Vegetation Indices (MOD13A1), version 5, dataset which includes NDVI and EVI. The dataset contains the highest daily value for NDVI/EVI within a 16 day period which also satisfies a set of data quality measures (NASA LP DAAC, 2017a).

The 16 day period selected was 6 - 22 March for each of the study years. This is based on the fact that the harvest period for millet, the most abundant crop grown in Namibia (NSA, 2015b), begins, on average, in April or May each year, depending on the agro-ecological zone (FAO, 2010). As a robustness check the results for NDVI are also shown when the 16-day period 18 February - 6 March is used.

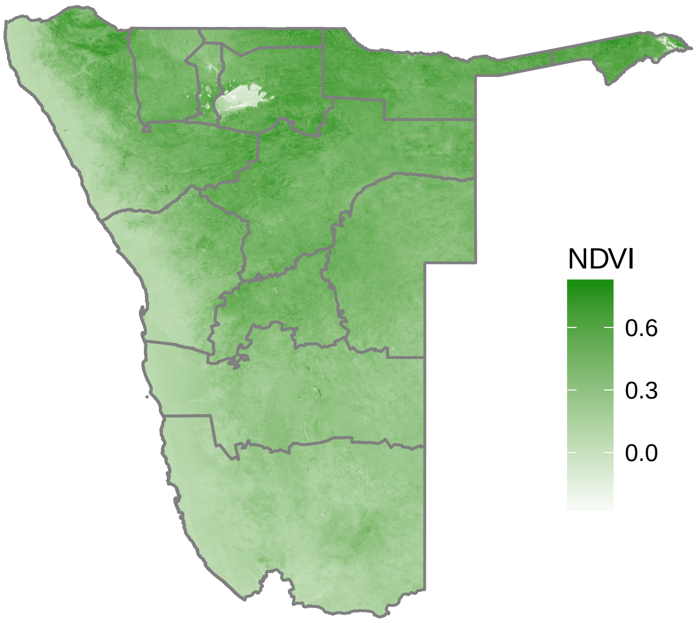
The last source of vegetation is a series based on EVI which calculates the integral of EVI (INEVI) for the primary growing season in a particular year. Zhang et al. (2006) shows that INEVI, based on the RS series NBAR MODIS, correlates well with another RS measure of vegetation productivity - Net Primary Productivity. The INEVI series used, comes from the V005 MODIS Land Cover Dynamics (MCD12Q2) product covering a resolution 500m^2 per pixel. The measure is based on algorithms developed in Zhang et al. (2003) to identify four key phenological phases of vegetation in a year. These phases can be used to identify the primary growing season in a year over which the integral of EVI is taken. Ganguly et al. (2010) provides an overview of the product. The product contains a large number of missing values across Namibia, however. This is due to the products algorithm excluding pixels that do not meet all the quality requirements (NASA LP DAAC, 2012b).[6] Thus, the results demonstrate the use of alternative specification of vegetation quality but should be interpreted with caution.

Data for 2001, 2003, 2009 and 2011 is used. The timing of the NHIES surveys (which stretch over two years) does complicate the choice of which year of data to use. Ideally the vegetation index should capture the season of production which provides for consumption and income that is captured in the surveys. In the NHIES data, the date of interviews of households does not vary systematically by region, and thus it is not possible to identify the reference agricultural season. As noted earlier, harvest of millet starts in April/May. For NHIES 2003/04 data collection started in September 2003 and ended in August 2004 (CBS, 2006). For NHIES 2009/10 data collection started in June 2009 and ended in July 2010 (NSA, 2013). It is thus clear that using 2003 and 2009 RS data should capture the periods of primary production that served consumption for most survey respondents. Figure 3.2 shows the NDVI for Namibia for 6-22 March 2009.

[5] See Huete et al. (2002) for the formula for EVI

[6] See NASA LP DAAC (2012a) for other know issues.

Figure 3.2: NDVI (6-22 March 2009) in Namibia



NDVI Source: MODIS Terra Vegetation Indices (MOD13A1), Version 5. Published by NASA LP DAAC.

Crop Land Data

For the visual exploration of economic activity identified by RS data in under-electrified settings, Erongo and Kavango were selected for study. Erongo, a previously advantaged region, has high levels of wealth and electrification. Kavango was initially north of the Police Zone, which implies it was subject to apartheid era policies. It has high levels of crop-farming and is under-electrified. Farming in the region happens mostly on communal lands.

For the classification of cropland, Landsat data was used. The Landsat program is a joint program by the U.S Geological Survey (USGS) and National Aeronautics and Space Administration (NASA). There have been 8 Landsat satellites launched. This study uses data from Landsat 5 and Landsat 7. Landsat 7 captures images of the same point every 16 days, while Landsat 5 does it every 18 days, both at resolution of 30m^2 (USGS, 2015).

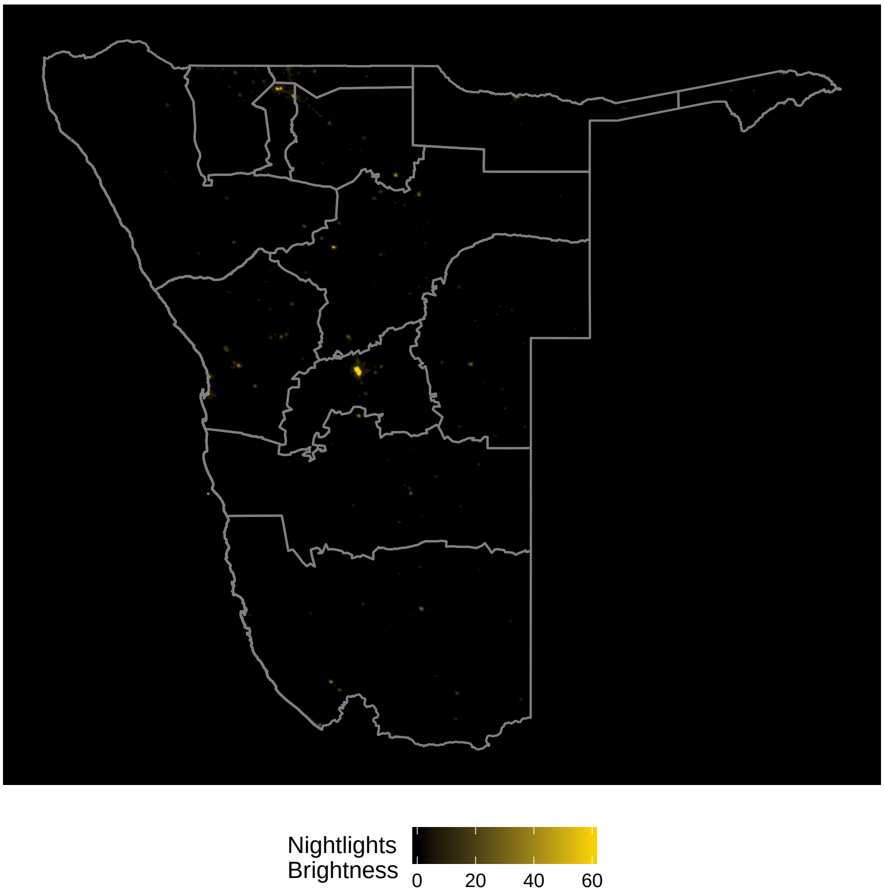
Images for 2003 were obtained from Landsat 7 imagery while Landsat 5 was used for 2009. The reversion back to Landsat 5 was due to a fault on the scanline corrector on Landsat 7, which caused a loss of data in images occurring after May 2003. Given that cloud-free images are required for the classification, the date of each image is not the same. Land use is slow to change, so this is not a problem in identifying cropland within a season. A satellite analysis team at Stellenbosch University - the Centre for Geographic Analysis (CGA) - was hired to process and classify the Landsat images into cropland or not. The CGA's report on how classification was conducted can be found in appendix A.

Night Lights Data

The night lights data is captured by the Defence Meteorological Satellite Program (DMSP) Operational Linescan System. The data is collected by the US Air Force Weather Agency and processed at the DMSP data centre. Raw data includes a range of radiance emitted at night, including moonlit objects, gas flares and man-made light. The DMSP releases a stable lights set for each year which excludes events such as fires and moonlit objects but includes gas flares. However, based on data available from DMSP, no gas flares are present in Namibia. The data is released in the form of 30 arc-second grids which is approximately 1km^2. Values range from 0 (which means that no stable light was detected there) in the year to 63 (which is a top censored value).

There are comparability problems over time as new satellites are launched and censors become older. A number of methods have been proposed to make the series comparable over time. This report uses calibration parameters proposed by Elvidge et al. (2014) which was calculated by identifying areas where the brightness of lights has not changed over time. Using this as a base, the brightness of lights of the different satellite series are calibrated to each other. Night lights data was used for the years corresponding to the survey and poverty mapping data. For 2001 and 2003, the F14 series is used, for 2009 the F16 series is used and for 2011 the F18 series is used. Figure 3.3 presents the brightness of light at night for Namibia for 2009.

Figure 3.3: Brightness of Light at Night in Namibia - 2009



Night lights Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency.

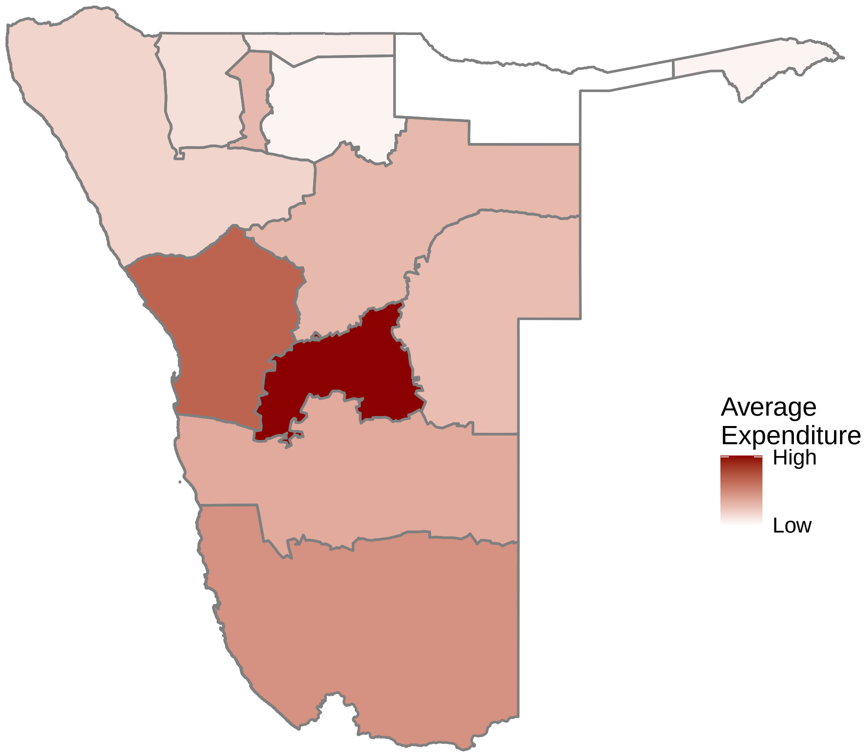
3.4 Results

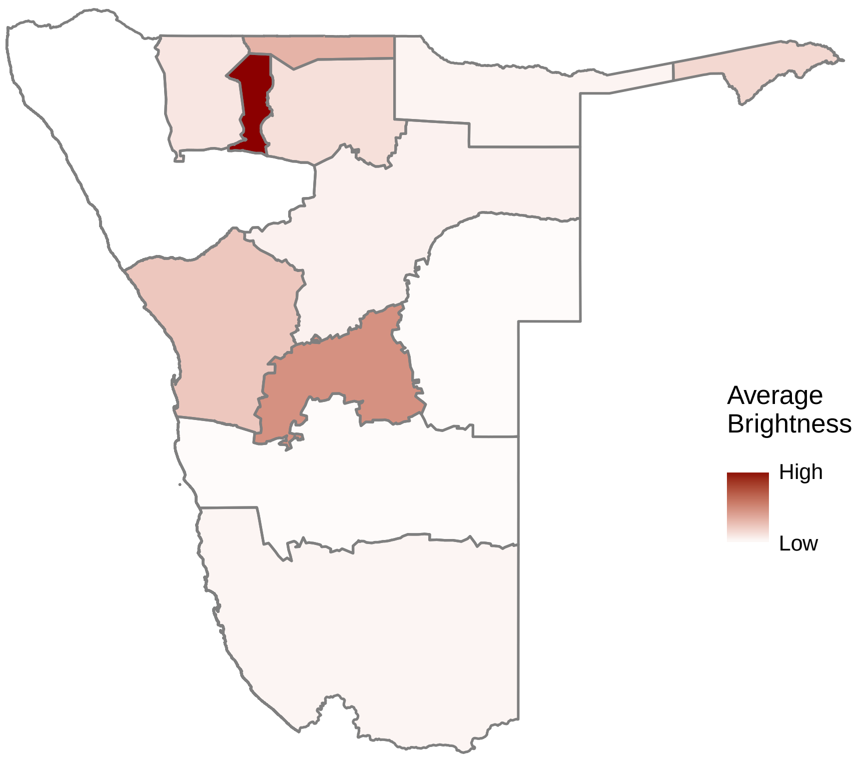
As a start to this section, descriptives highlighting the regional inequalities in Namibia are shown. This is followed by a section which explores the extent of economic activity observed, and, finally the regression results. For all analyses that follow, the brightness of light and vegetation quality are expressed as regional averages.

3.4.1 Descriptives

Figure 3.4 shows average expenditure and average brightness of lights, and Figure 3.5 shows average NDVI and headcount poverty by region for Namibia in 2009. It is evident that the highest poverty rates, and also the lowest average expenditure, is located in the north east of Namibia which are communal agricultural areas. Kavango, specifically, had headcount poverty higher than 50% in 2009. The regions with the main urban centres - Khomas and Erongo - had the highest levels of welfare.

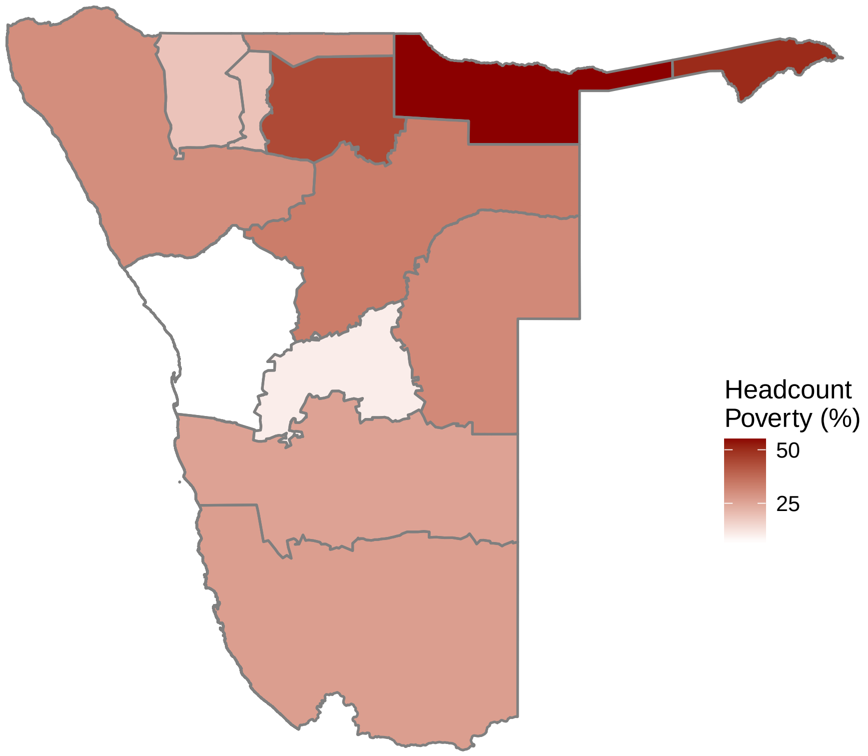
Figure 3.4: Regional Differences in Average Expenditure and Night lights - 2009

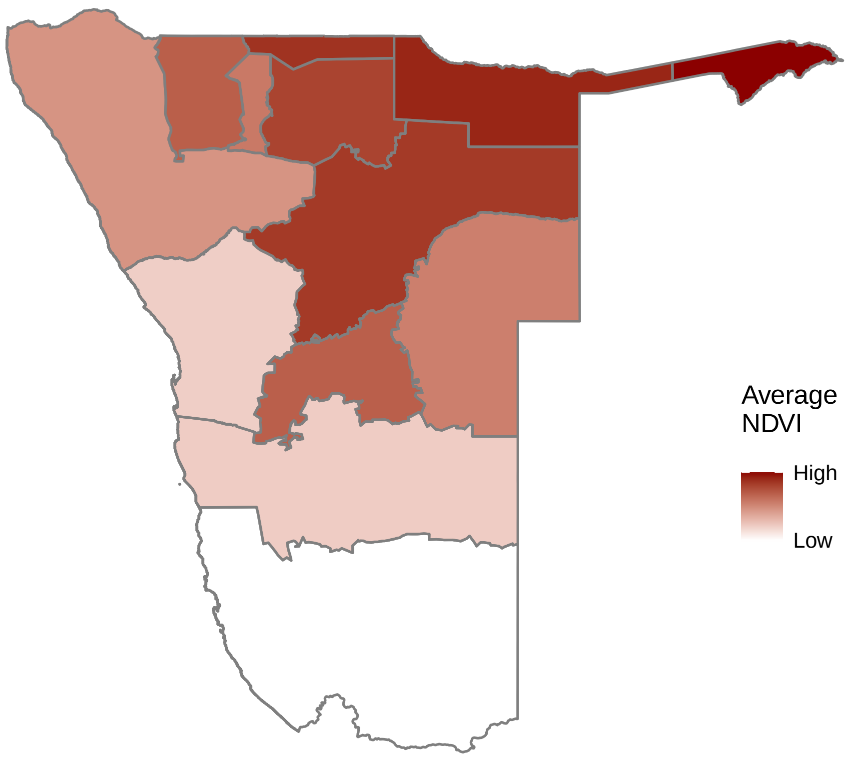




Night lights Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency. Average Expenditure Source: Own Calculations using NHIES 2009/10.

Figure 3.5: Regional Differences in Headcount Poverty and NDVI - 2009





NDVI Source: MODIS Terra Vegetation Indices (MOD13A1), Version 5. Published by NASA LP DAAC. Poverty Rate Source: Own Calculations using NHIES 2009/10.

The regions with higher levels of poverty, also tended to have higher average NDVI levels, and these regions are once again located in the the north-east of Namibia. It needs to be pointed out, however, that Khomas has a relatively high level of average NDVI as well. Windhoek, the urban centre, covers a small proportion of the total area, which implies that the rural areas increase the average NDVI observed. Average brightness of light is low in the poor north-east regions and higher in Erongo and Khomas, which have the lowest levels of poverty. Oshana, in the north-centre of the country had the highest levels of average brightness. Oshana contains a number of large towns and the region is much smaller than both Erongo and Khomas. This highlights the importance of controlling for the size of a region.

To quantify regional differences in agricultural activities and electrification, table 3.1 presents results from Namibia's 2011 Census. The percentage of households in a region that use electricity for lighting varies from 12% in Omusati to 83% in Erongo. The percentage of households that farm also varies from 11% in Erongo and Khomas to 81% in Ohangwena. This indicates, that night lights would likely identify greater areas of economic activity in urbanised regions such as Erongo and Khomas. Daytime RS should identify greater areas in under-electrified regions with high levels of agriculture.

Table 3.1: Household Electrification and Farming for Across Regions in Namibia

	Use electricity for lighting	Farming	Farming is Main Source of Income	Farms Crops
Caprivi	0.35	0.48	0.21	0.44
Erongo	0.83	0.11	0.02	0.04
Hardap	0.69	0.22	0.07	0.04
Karas	0.71	0.14	0.05	0.03
Kavango	0.25	0.56	0.43	0.54
Khomas	0.70	0.11	0.01	0.05
Kunene	0.35	0.48	0.32	0.34
Ohangwena	0.13	0.81	0.26	0.81
Omaheke	0.38	0.30	0.21	0.10
Omusati	0.12	0.74	0.22	0.73
Oshana	0.34	0.49	0.13	0.47
Oshikoto	0.22	0.67	0.33	0.66
Otjozondjupa	0.58	0.24	0.10	0.11

Source: Own Calculations using Namibia Census 2011

3.4.2 Observed Economic Activity in Erongo and Kavango

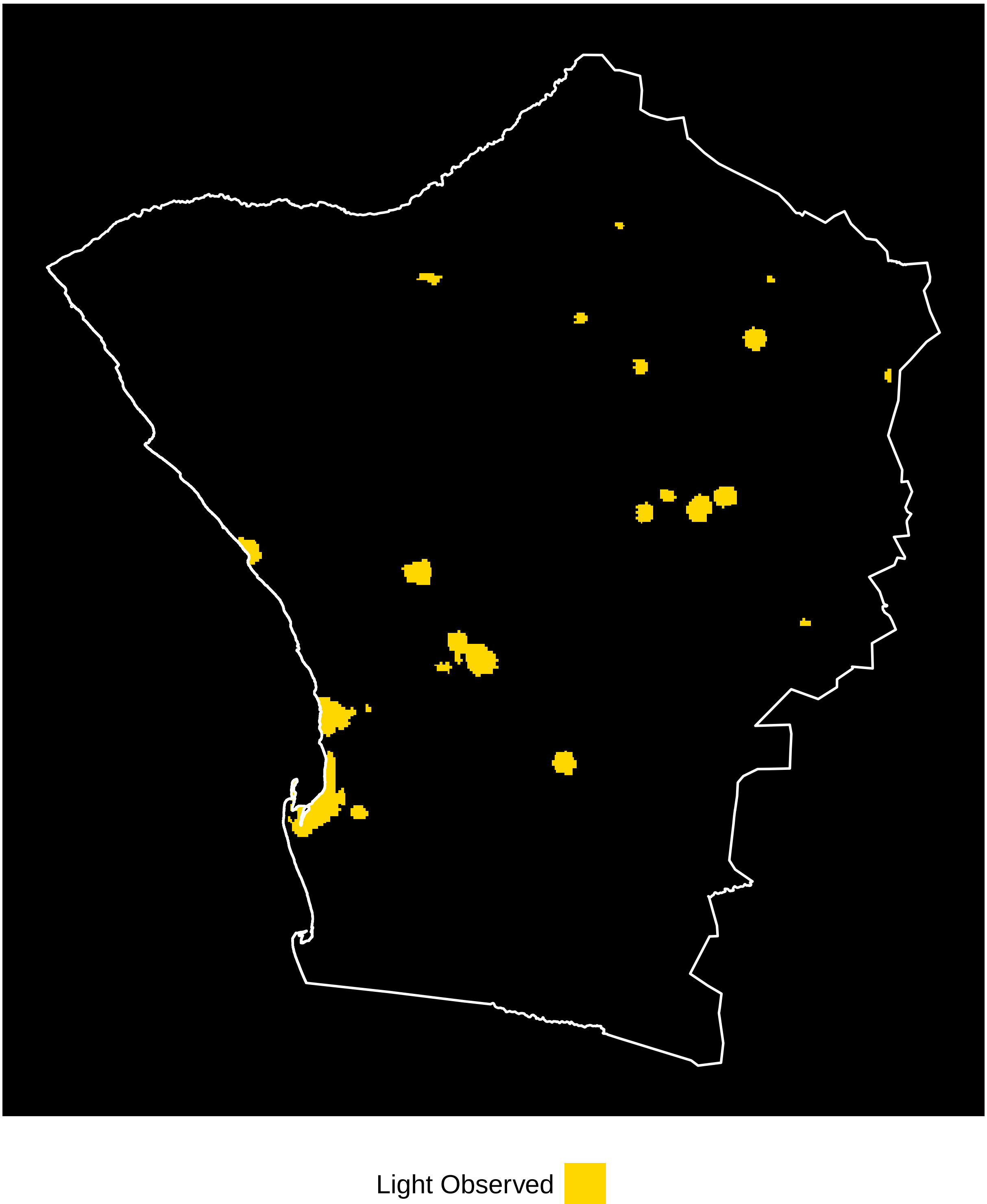
This section explores visually how observed lighting and crop farming differ across Erongo and Kavango. In Erongo, 83% of households use electricity for lighting while in Kavango only 25% do. On the other hand, only 4% of households farm crops in Erongo compared to 54% in Kavango.

For the figures depicting cropland, pixels expressed the percentage of the area covered in cropland.[7] Figures 3.6 and 3.7 show areas observed to have nighlights and cropland in Erongo respectively. A large, and spread out area of Erongo is observed under lighting while limited crop farming is observed in the north east of the region. Lights are observed for 1.6% of the total area while crop farming is observed for 0.046% of the total area. In Erongo, it is thus evident that night lights data has greater ability to locate areas of economic activity.

Figures 3.8 and 3.9 show observed lighting and cropland for Kavango. The figures confirm the expectation that daytime RS data observes a greater area of economic activity than night lights in under-electrified areas. Only urban areas, mostly the town of Rundu, are observed using night lights. Quantitatively, 4.6% of the total area was observed to be cropland. Night lights was only observed for 0.5% of the area.

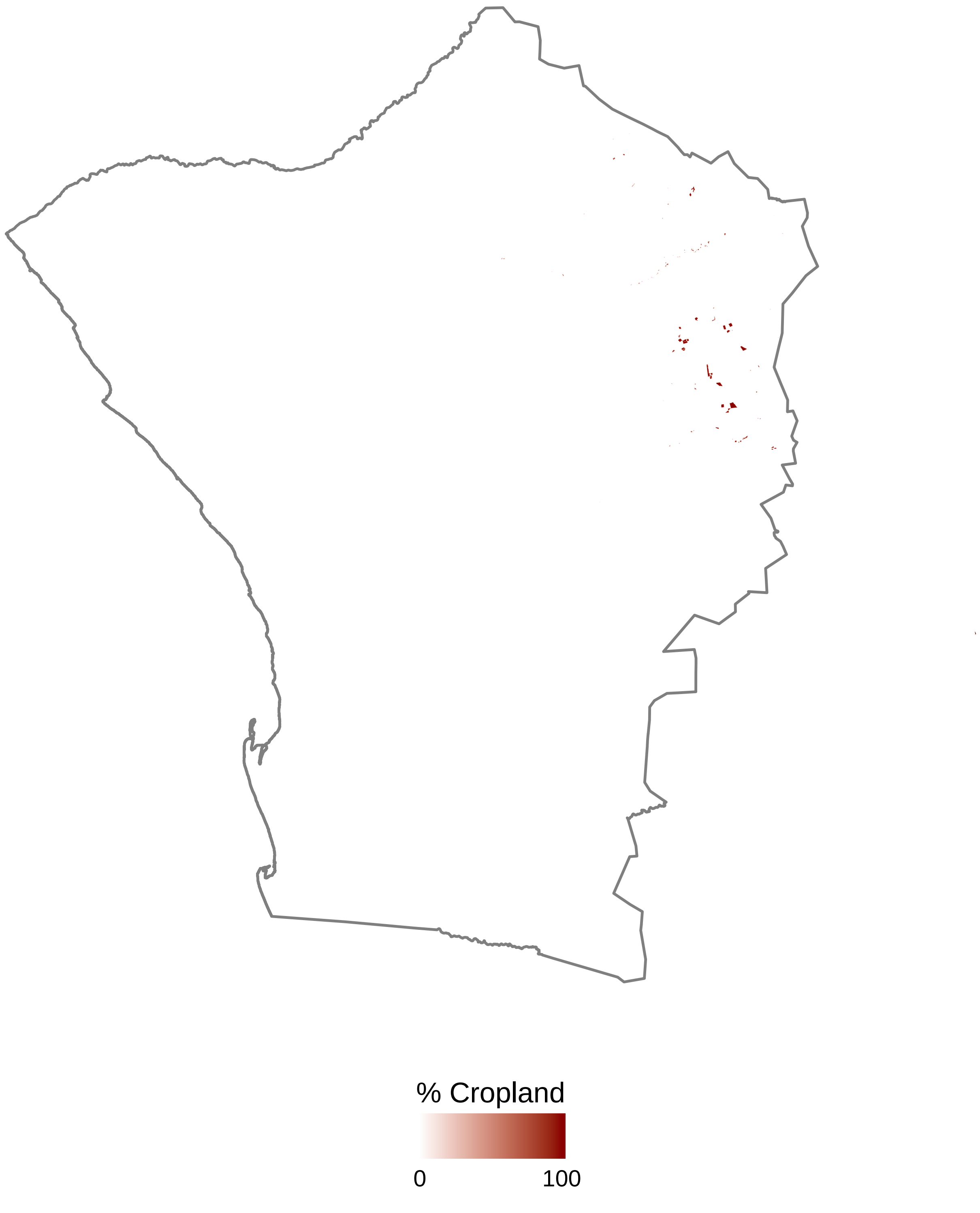
[7] This was done because there are too many data points in Landsat images to plot each individually. Thus, images were aggregated to a smaller number of pixels where each represents the percentage of cropland for the area it covers.

Figure 3.6: Area Observed by Night Lights - Erongo



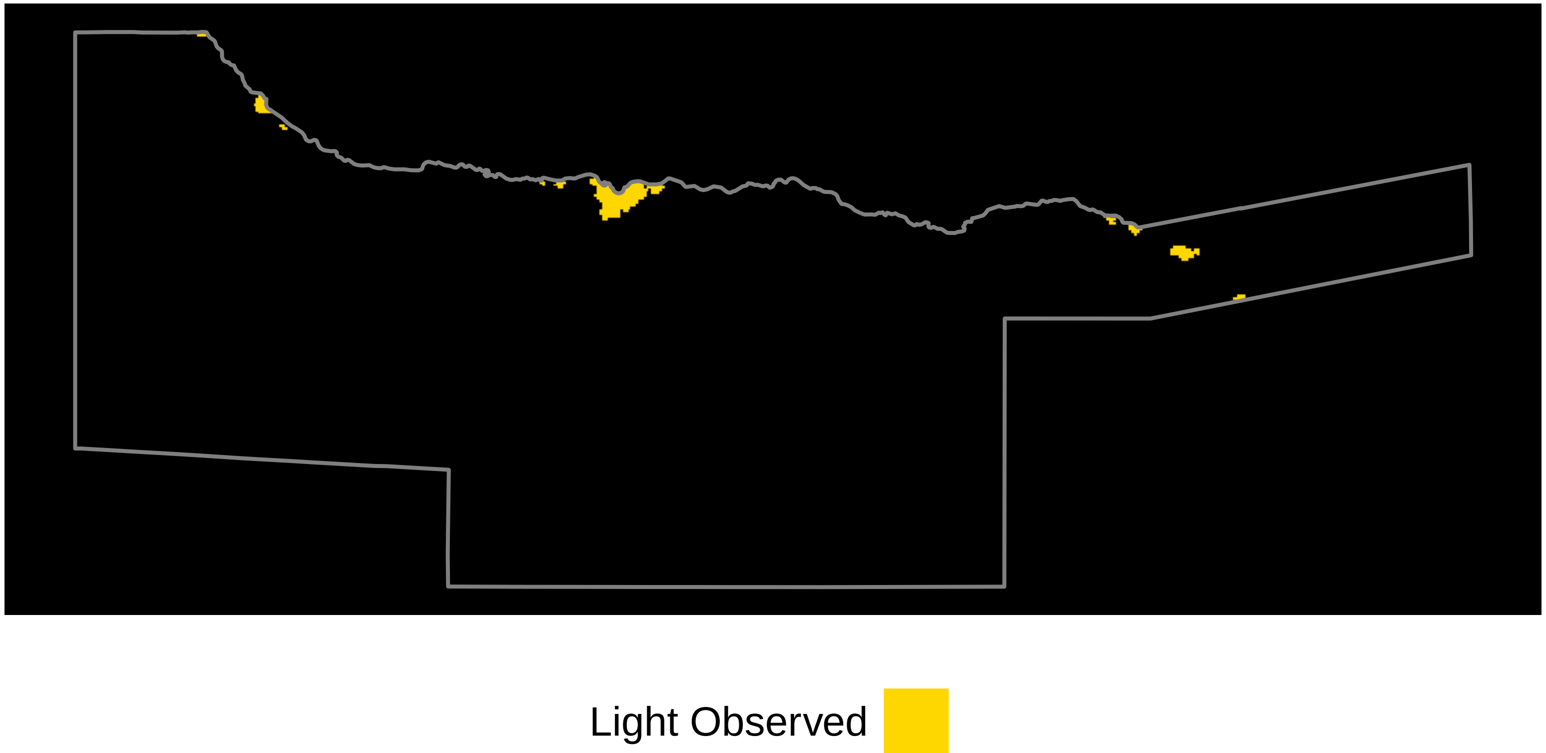
Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency.

Figure 3.7: Area Observed under Cropland - Erongo



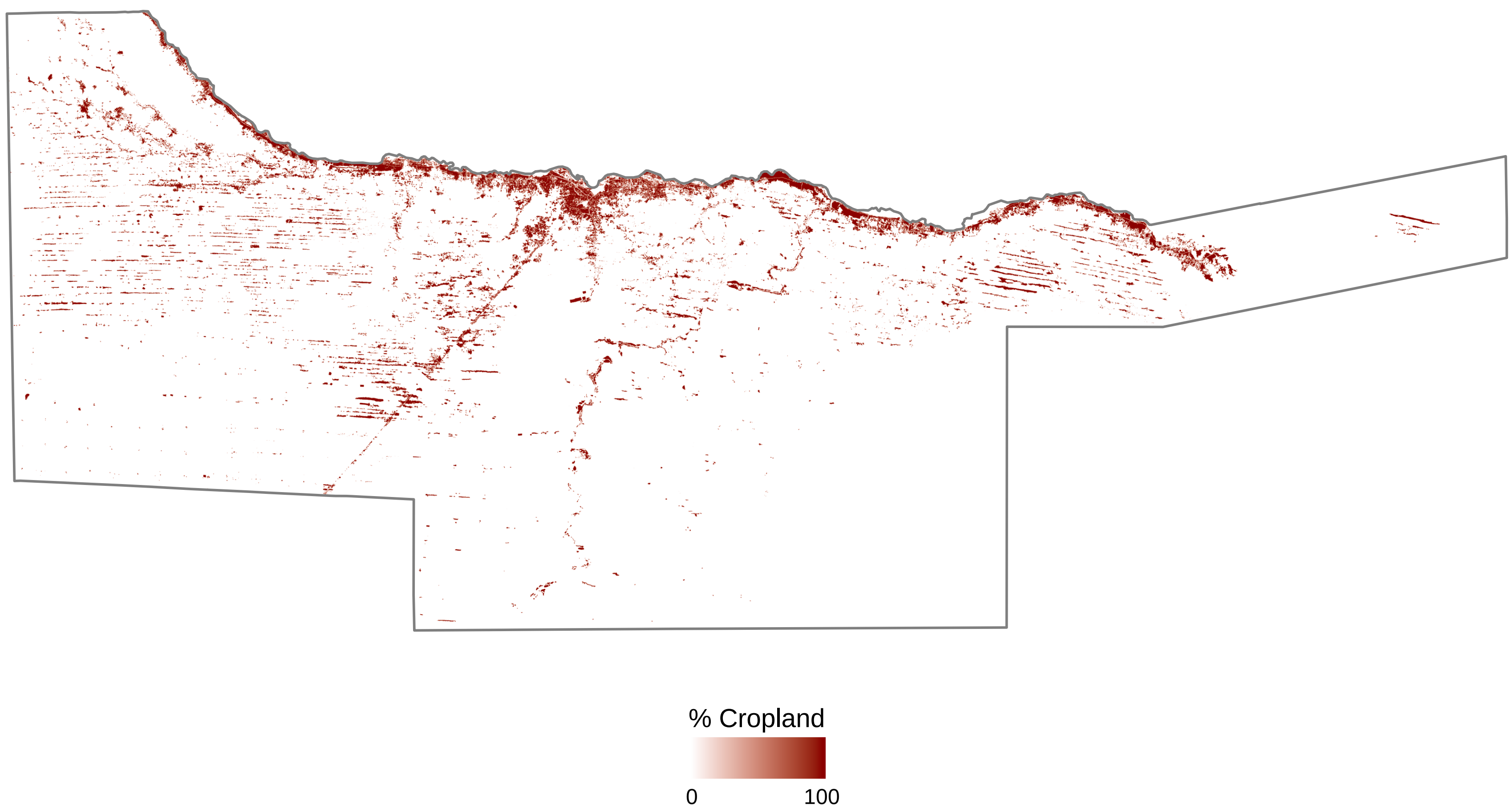
Classification using Landsat 5. Data available from the U.S. Geological Survey.

Figure 3.8: Area Observed by Night Lights - Kavango



Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency.

Figure 3.9: Area Observed under Cropland - Kavango



Classification using Landsat 5. Data available from the U.S. Geological Survey.

Figure 3.10: Location of Kahenge in Kavango

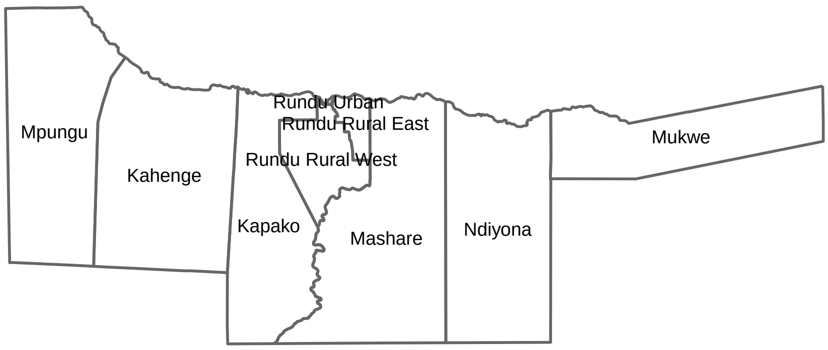


Table 3.2: Household Electrification and Farming for Across Constituencies in Kavango

	Use electricity for lighting	Farming	Farming is Main Source of Income	Farms Crops
Kahenge	0.08	0.68	0.61	0.67
Kapako	0.14	0.64	0.59	0.63
Mashare	0.08	0.68	0.53	0.67
Mpungu	0.10	0.82	0.61	0.82
Mukwe	0.16	0.74	0.60	0.73
Ndiyona	0.13	0.64	0.48	0.63
Rundu Rural West	0.38	0.37	0.22	0.36
Rundu Urban	0.81	0.17	0.04	0.15
Rundu Rural East	0.21	0.41	0.32	0.39

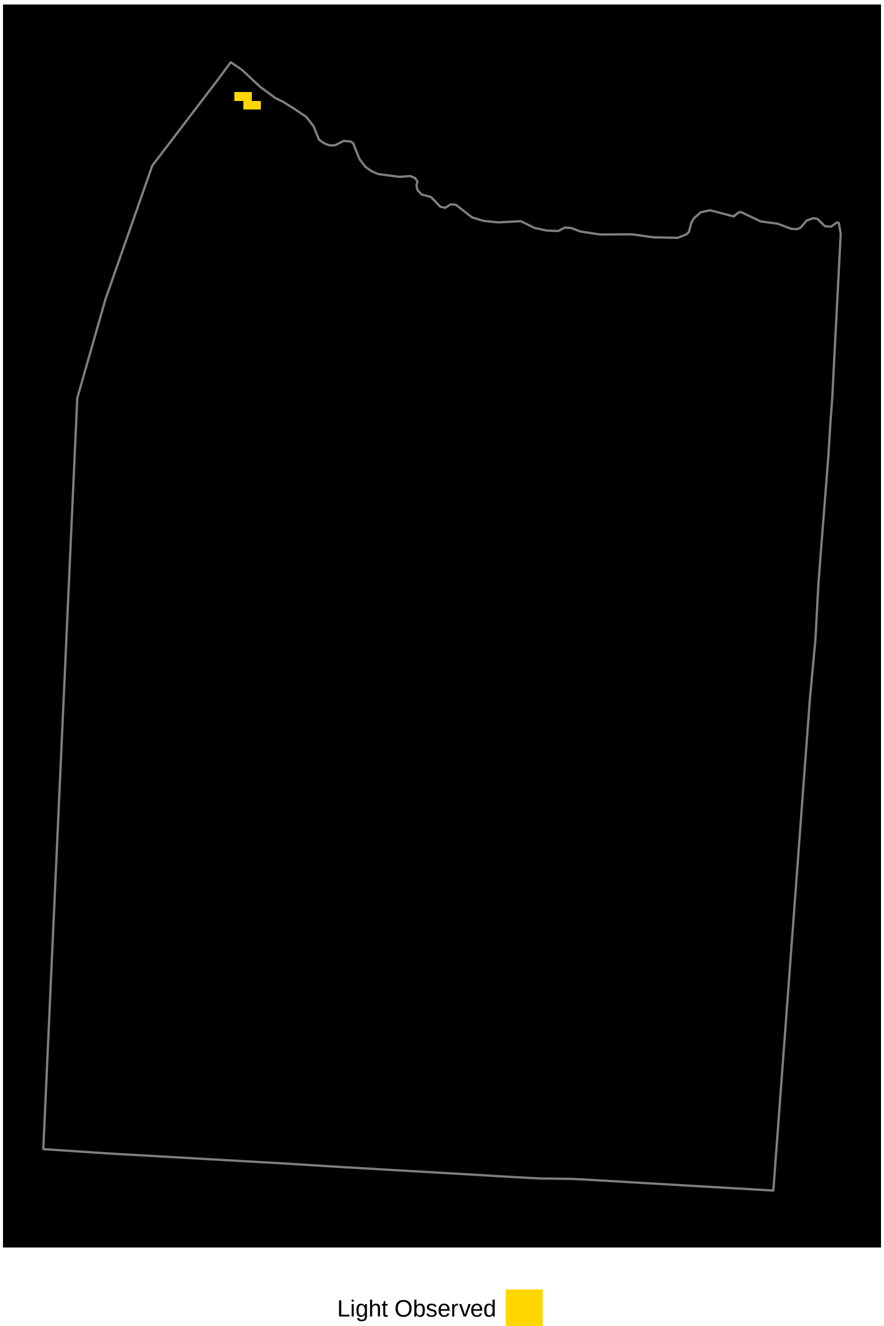
Source: Own Calculations using Namibia Census 2011

The figures for Kavango, at the scale which they can be presented here, fail to show the extent to which daytime RS data can pick up activity in rural areas. To illustrate this further, a sub-region in Kavango - Kahenge - is shown. Table 3.2 show the rates of farming and electrification for constituencies in Kavango and it shows that in Kahenge only 8% of households use electricity for lighting, 68% of households farm and 67% of households farm crops. See figure 3.10 for the location of constituencies in Kavango. Figures 3.11 and 3.12 show observed night lights and cropland for Kahenge. Night lights were observed in only 0.04% of the region while cropland was observed for 5.7% of the region.

Given that this chapter is concerned with tracking socio-economic changes, the area that changed status between 2003 and 2009 was also compared for night lights and cropland in Kahenge. It is to be noted that cropland is not the measure used to track socio-economic outcomes in the next section, but the observed changes drive home the point that daytime RS data does indeed capture changes at much smaller and more disaggregated levels than the night lights data in rural under-developed areas. The areas where night lights changed are shown in figure 3.13 and areas where cropland changed are shown in figure 3.14.

The fact that night lights are observed for a greater area than cropland in Erongo highlights that different satellite measures will be more effective in different settings. This analysis has indeed shown, however, that daytime RS data observes more economic activity and changes in under-developed agricultural areas. This does not imply that night lights data should not be used. At aggregate regional level it is still plausible that it correlates well with welfare. Daytime RS data has the potential, however, to track welfare changes at much smaller levels of aggregation within such regions. This has potential to, for example, accurately identify areas that suffered poor agricultural harvests and thus might be in need of economic relief in a given year.

Figure 3.11: Area observed by Night Lights in Kahenge, Kavango, 2009

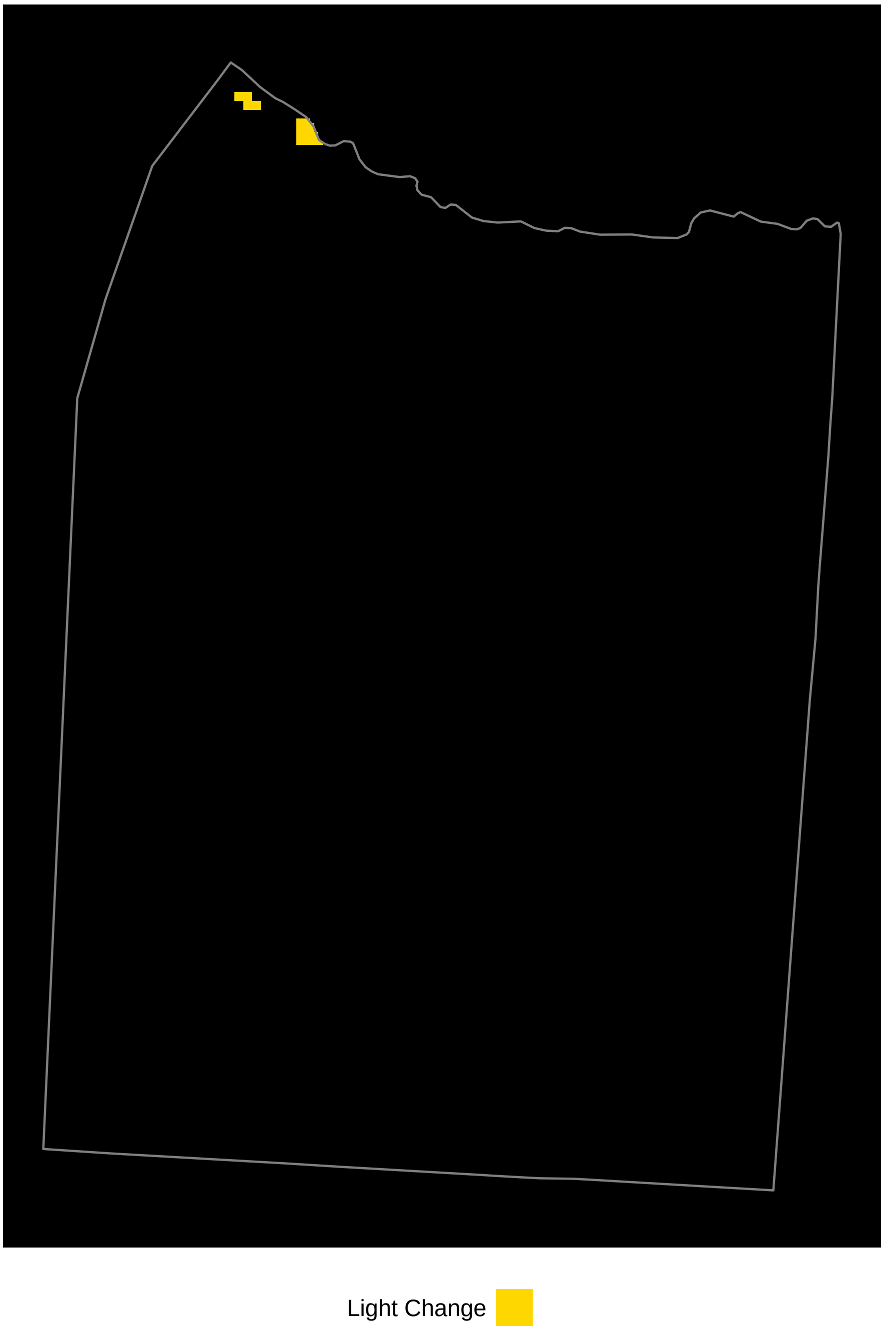


Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency.

Figure 3.12: Area observed by Night Lights in Kahenge, Kavango, 2009

Source: Classification using Landsat 5. Data available from the U.S. Geological Survey.

Figure 3.13: Area Night Lights Observed that Changed Between 2003 - 2009 in Kahenge, Kavango



Source: Classification using Landsat 5. Data available from the U.S. Geological Survey.

Figure 3.14: Area Cropland Observed that Changed Between 2003 - 2009 in Kahenge, Kavango

Cropland Source: Classification using Landsat 5. Data available from the U.S. Geological Survey.

3.4.3 Correlation between Remotely Sensed Data and Welfare

The regression results are presented next. The average brightness of light is logged and vegetation indices are standardized. All standard errors are clustered at the geographic level of observation - either regions or constituencies.

Table 3.3 shows results for regressions on logged average regional expenditure. Model 1 shows the results when no controls are introduced. As is consistent with our expectation, NDVI is negatively correlated with average expenditure cross-sectionally, but once regional fixed effects are introduced, NDVI has a positive effect on average expenditure. Coefficients are statistically significant even with only 26 observations. The fixed effects point estimate suggests that as average NDVI increases by a standard deviation, average expenditure increases, on average, by 15% in a region.

In model 2, population density is included. This is included to control for the number of people that are dependant on a specific area of land for production. After controlling for it, NDVI coefficients do not change by much and remain statistically significant.

Model 3 shows the relationship between average expenditure and the log of average lights when no controls are added and in model 4 population density is once again added. The results show that average expenditure is higher in regions with higher levels of brightness and that within regions, increases in average brightness are associated with increases in average expenditure. This is consistent with the literature. A 1% increase in average brightness is associated, on average, with a 42% increase in average expenditure. For the pooled OLS estimates, the point estimate is initially 0.12 and insignificant, but once population density is controlled for, it increases to 0.43 and becomes significant.

In model 5 both NDVI and night lights are included with population density. As the focus of the study is mainly on the use of indicators of vegetation quality, it is useful to consider night lights as a proxy for urbanization in a region. Thus, even after controlling for the extent of urbanisation, the negative relationship between NDVI and expenditure remains significant. Also, the positive relationship over time remains significant, but the coefficient decreases slightly suggesting that expenditure increases by 11% with a one standard deviation increase in NDVI.

Finally, year dummies are added to control for year specific effects. For the pooled OLS specification this increases the adjusted R^2 to 0.69. For the fixed effects estimates this decreases the coefficients on both the average brightness of lights and NDVI. This is consistent with the finding that NDVI and expenditure are positively correlated. The intra-class correlation is 0.96 suggesting that almost all of the variation is contained in the fixed effects of the model. The fact that the linear time trend causes the coefficient on NDVI to become insignificant suggests that NDVI was capturing the long-term trend in expenditure over time.

The initial results suggest that indicators of vegetation quality do indeed correlate with expenditure as expected. The fact that the sign on NDVI changes between the pooled OLS and fixed effects regressions,

Table 3.3: Regional Regressions on ln(Average Regional Expenditure)

	1		2		3		4		5		6	
	POLS	FE	POLS	FE	POLS	FE	POLS	FE	POLS	FE	POLS	FE
NDVI	-0.25***	0.15***	-0.23***	0.18***					-0.21***	0.11**	-0.33***	0.05
	(0.06)	(0.04)	(0.07)	(0.04)					(0.07)	(0.05)	(0.07)	(0.14)
ln(Light)					0.12	0.42***	0.43***	0.46***	0.41***	0.25	0.35**	0.19
					(0.11)	(0.12)	(0.13)	(0.12)	(0.11)	(0.18)	(0.13)	(0.19)
Pop Density			-0.00	-0.07			-0.07***	-0.04	-0.05***	-0.08	-0.04**	-0.07
			(0.01)	(0.05)			(0.01)	(0.04)	(0.01)	(0.05)	(0.02)	(0.05)
Year	No	No	No	No	No	No	No	No	No	No	Yes	Yes
adj.R^2	0.21	0.50	0.18	0.53	0.03	0.50	0.44	0.50	0.58	0.60	0.69	0.59
R^2: Within		0.52		0.57		0.52		0.54		0.65		0.66
R^2: Between		0.48		0.00		0.06		0.32		0.16		0.17
R^2: Overall		0.24		0.01		0.07		0.33		0.17		0.18
ICC		0.97		0.98		0.97		0.94		0.96		0.96
Observations	26	26	26	26	26	26	26	26	26	26	26	26
Groups		13		13		13		13		13		13

Socio Economic Data Source: Own Calculations using NHIES 2003/04 and 2009/10. NDVI Source: MODIS Terra Vegetation Indices (MOD13A1), Version 5. Published by NASA LP DAAC. Night lights Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency.
Regression also controls for year fixed effects.
POLS refers to Pooled OLS and FE refers to Fixed Effects regression.
ICC refers to the Intra-class correlation.

highlight that there is selection amongst differing regions. The poor are more likely to be located in rural regions with higher levels of vegetation quality. Robustness checks are discussed next.

Robustness checks

Table 3.4 shows the estimates when regional average expenditure is regressed on alternative vegetation indices. The model specification used, is similar to specification 2 in Table 3.3. In the first specification the NDVI from the 16-day composites for 18 February - 6 March (instead of 6 March - 22 March) is used. The second model uses EVI from the 16 day composite for 6 March - 22 March. The final model uses INEVI. In all cases the results are consistent with previous results - vegetation is negatively correlated with welfare across regions and positively correlated with welfare over time within regions. Note that model coefficients do not change dramatically when different indicators are used. The explanatory power of the fixed effects estimator when INEVI is used, is lower though. This can be due to the large amount of missing data.

Table 3.4: Regional Regressions on ln(Average Expenditure) using Alternative Vegetation Indices

	POLS	FE	POLS	FE	POLS	FE
NDVI(earlier date)	-0.26**	0.22***				
	(0.10)	(0.06)				
EVI			-0.27**	0.19***		
			(0.10)	(0.05)		
INEVI					-0.26**	0.15*
					(0.10)	(0.07)
Pop. Density	-0.00	-0.09	0.00	-0.07	-0.00	-0.01
	(0.01)	(0.06)	(0.01)	(0.06)	(0.01)	(0.07)
$adj.R^2$	0.24	0.03	0.21	0.10	0.22	-0.52
R^2: Within		0.57		0.60		0.33
R^2: Between		0.00		0.00		0.31
R^2: Overall		0.00		0.00		0.18
ICC		0.98		0.98		0.95
Observations	26	26	26	26	26	26
Groups		13		13		13

Socio Economic Data Source: Own Calculations using NHIES 2003/04 and 2009/10. NDVI Source: MODIS Terra Vegetation Indices (MOD13A1), Version 5. Published by NASA LP DAAC. Night lights Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency.

Regression also controls for year fixed effects.

POLS refers to Pooled OLS and FE refers to Fixed Effects regression.

ICC refers to the Intra-class correlation.

For comparison with based results in Table 3.3 the model specification used is model 2.

Table 3.5 show the results when regional headcount poverty rates (instead of average expenditure) are regressed on average NDVI and average brightness. The estimates show, as expected, that NDVI is positively correlated with poverty across regions, but negatively correlated over time within regions.

Average brightness is negatively correlated with poverty across regions and within regions over time. Based on the R^2 values, the RS data performs better in explaining variations in average expenditure levels than poverty rates.

Finally, Table 3.6 shows the results when constituency level poverty rates are regressed on NDVI and lights. As the rates are based on imputed income, the standard errors need to be bootstrapped for correct inference.[8] For nine constituencies light was not observed at all in either 2001 or 2011.[9] This also highlights the limitations of night lights for tracking economic activity in under-developed areas. Point estimates for NDVI are consistent with estimates based on regional level poverty rates. They are similar in magnitude and sign and are statistically significant. The relationship between night lights and poverty rates over time is not as clear at constituency as regional level. The point estimate, after controlling for population density, was -24.69 in regional level regressions but -1.45 in constituency level regressions.

On concern is the model specification 6 in Table 3.6. The coefficient on NDVI when fixed effects and year dummies are introduced is positive, suggesting that NDVI and poverty are positively correlated over time within small regions. This result is driven by urban constituencies as is shown in Table 3.7. Table 3.7 introduces the percentage of the population that reside in rural areas within regions or constituencies in regressions on the headcount poverty rate. In the first specification, the percentage of the population residing in rural areas is interacted with NDVI at constituency level and regional level. This shows that over time, within more rural constituencies, poverty decreases with increases in NDVI. This interaction is also statistically significant when the sample of constituencies is used. In the second specification, the sample of constituencies is limited to those where more than 50% of the population reside in rural areas. In this specification, the NDVI is insignificant - a finding that mirrors specification 6 in Table 3.5.

The initial positive correlation between NDVI and poverty over time within constituencies was thus driven by predominantly urban constituencies. The results at regional level were clearly not as sensitive to urban areas because of the relative size of urban areas within regions. Applying the environmental Kuznets curve hypothesis to the results for urban areas suggests that a large number of urban constituencies are at initial stages of development where increases in development are associated with environmental deterioration.

The results support this chapter's premise that RS based vegetation quality indices can be used to track socio-economic welfare over time in rural areas. The coefficients are large and have the correct sign based on the idea that livelihoods are dependant on nature and agriculture. The results have also highlighted why cross-sectional results can be misleading. The night lights results are also consistent with previous findings which is that night lights are positively correlated with welfare. However, as pointed out in the previous section, daytime RS data presents opportunities to track changes at much smaller geographic levels as night lights are completely absent in many rural areas.

---

[8] Pooled OLS results are clustered at constituency level, but the fixed effects estimates are not.

[9] For these constituencies, a value of 0.0001 was assigned before the average level of brightness was logged.

Table 3.5: Regional Regressions on Headcount Poverty

	1		2		3		4		5		6	
	POLS	FE	POLS	FE	POLS	FE	POLS	FE	POLS	FE	POLS	FE
NDVI	5.30*	-6.54**	6.29**	-9.13***					5.54**	-5.08	10.16***	1.20
	(2.54)	(2.54)	(2.73)	(2.71)					(2.20)	(4.38)	(2.78)	(10.17)
ln(Light)					-5.90*	-20.08**	-12.14**	-	-	-15.41	-9.37**	-8.86
								24.69***	11.61***			
					(2.74)	(8.19)	(4.20)	(7.56)	(3.46)	(14.00)	(3.32)	(13.49)
Pop Density			-0.37	5.57			1.34***	4.17	0.98**	5.93*	0.52	5.49
			(0.36)	(3.72)			(0.43)	(2.44)	(0.34)	(3.26)	(0.37)	(3.65)
Year	No	No	No	No	No	No	No	No	No	No	Yes	Yes
$adj.R^2$	0.09	0.29	0.08	0.37	0.16	0.35	0.33	0.39	0.43	0.44	0.61	0.44
R^2: Within		0.32		0.42		0.38		0.44		0.50		0.53
R^2: Between		0.38		0.02		0.19		0.30		0.02		0.02
R^2: Overall		0.13		0.01		0.19		0.29		0.03		0.02
ICC		0.86		0.97		0.90		0.87		0.96		0.96
Observations	26	26	26	26	26	26	26	26	26	26	26	26
Groups		13		13		13		13		13		13

Socio Economic Data Source: Own Calculations using NHIES 2003/04 and 2009/10. NDVI Source: MODIS Terra Vegetation Indices (MOD13A1), Version 5. Published by NASA LP DAAC. Night lights Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency.
Regression also controls for year fixed effects.
POLS refers to Pooled OLS and FE refers to Fixed Effects regression.
ICC refers to the Intra-class correlation.

Table 3.6: Constituency Level Regressions on Headcount Poverty

	1		2		3		4		5		6	
	POLS	FE	POLS	FE	POLS	FE	POLS	FE	POLS	FE	POLS	FE
NDVI	7.14***	-9.99***	6.52***	-					4.99***	-	6.20***	6.65**
				10.28***						10.61***		
	(1.46)	(2.13)	(1.41)	(2.15)					(1.26)	(2.16)	(1.26)	(3.22)
ln(Light)					-3.80***	-1.45	-3.57***	-1.45	-3.29***	-1.83	-3.13***	1.88
					(0.40)	(1.83)	(0.45)	(1.72)	(0.47)	(1.40)	(0.51)	(1.31)
Pop Density			-0.00	0.00			-0.00	0.00	-0.00	0.00	-0.00	0.01
			(0.00)	(0.00)			(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
Year	No	No	No	No	No	No	No	No	No	No	Yes	Yes
$adj.R^2$	0.12	-0.74	0.20	-0.75	0.34	-0.98	0.34	-1.00	0.40	-0.72	0.47	-0.06
R^2: Within		0.13		0.14		0.01		0.01		0.16		0.49
R^2: Between		0.19		0.28		0.40		0.39		0.07		0.12
R^2: Overall		0.12		0.20		0.34		0.33		0.04		0.06
ICC		0.86		0.88		0.70		0.70		0.86		0.95
Observations	214	214	214	214	214	214	214	214	214	214	214	214
Groups		107		107		107		107		107		107

Constituency Level Poverty Rates: NPC (2015). NDVI Source: MODIS Terra Vegetation Indices (MOD13A1), Version 5. Published by NASA LP DAAC. Night lights Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency.
Regression also controls for year fixed effects.
POLS refers to Pooled OLS and FE refers to Fixed Effects regression.
ICC refers to the Intra-class correlation.

Table 3.7: Regressions on Poverty Rates Accounting for the Rural Population

	Rural Interaction				>50% Rural	
	Regional Estimates		Constituency Estimates		Constituency Estimates	
	POLS	FE	POLS	FE	POLS	FE
NDVI	7.16**	4.38	3.33***	14.91***	7.21***	3.50
	(2.69)	(11.78)	(0.93)	(2.86)	(1.39)	(4.36)
NDVI*% Rural	1.03	-4.25	3.08*	-14.45***		
	(4.29)	(9.22)	(1.72)	(3.37)		
% Rural	22.42	28.97	33.09***	12.45		
	(13.62)	(36.88)	(2.23)	(12.43)		
ln(Light)	-4.36	-5.19	-0.64	1.49	-1.34*	2.02
	(3.51)	(14.10)	(0.40)	(1.30)	(0.78)	(1.36)
Pop Density	-0.24	3.76	0.00	0.01	0.06	-1.78***
	(0.44)	(4.37)	(0.00)	(0.01)	(0.07)	(0.65)
Year	Yes	Yes	Yes	Yes	Yes	Yes
adj.R^2	0.66	0.42	0.69	0.05	0.38	0.28
R^2: Within		0.56		0.55		0.67
R^2: Between		0.10		0.01		0.02
R^2: Overall		0.11		0.00		0.02
ICC		0.93		0.93		0.98
Observations	26	26	214	214	146	146
Groups		13		107		75

Socio Economic Data Source: Own Calculations using NHIES 2003/04 and 2009/10. NDVI Source: MODIS Terra Vegetation Indices (MOD13A1), Version 5. Published by NASA LP DAAC. Night lights Source: Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the United States Air Force Weather Agency.

Regression also controls for year fixed effects.

POLS refers to Pooled OLS and FE refers to Fixed Effects regression.

ICC refers to the Intra-class correlation.

For comparison with based results in Table 3.3 the model specification used is model 2.

3.5 Conclusion

In most countries in SSA, household survey data is released at low frequency and is only representative at high levels of geographic aggregation. This is specifically the case with surveys designed to measure poverty. This limits what researchers can say about the location and short-term changes in poverty.

A new body of research attempts to overcome these limitations by using satellite data. Satellites capture data at high frequencies and small areas. The night lights data series, has particularly been shown to correlate well with GDP. However, a large share of the poor in sub-Saharan Africa are located in under-electrified, agriculturally dominant regions. Night lights cannot proxy economic activity where it is not present. Daytime RS data has also been used to model welfare cross-sectionally but the research and conclusions have not dealt with variation over time. Henderson et al. (2012) noted that night lights data is more suitable to estimate within region changes, and the same case can be made for daytime

RS indicators. Regional heterogeneity can imply that cross-sectional correlations could be misleading if welfare changes over time are studied.

This study has suggested the use of indicators of vegetation quality, derived from daytime RS as a proxy for welfare in rural regions. The reasoning is as follows: indicators of vegetation quality have been used substantially to predict crop harvests, and given that smallholder agriculture is an important source of income for households in rural areas, it should be that case that vegetation quality can proxy welfare. Indicators of vegetation quality would also overcome the ineffectiveness of night lights in underdeveloped areas.

Namibia was selected to explore theses ideas as it has large regional heterogeneity which could possibly highlight why cross-sectional results could be misleading. To illustrate that daytime RS data can identify economic activities where night lights cannot, the area of economic activity observed, proxied either night lights or cropland, was identified in Kavango - an impoverished, agricultural region. It was indeed observed that cropland was identified for a much larger area than night lights was observed for, indicating that daytime RS data can track economic change where night lights cannot.

Secondly, regression analysis was used to calculate correlations between indicators of vegetation quality and welfare, between regions, and within regions over time. Results showed that vegetation quality and welfare are negatively correlated between regions but are positively correlated over time within rural areas. This highlighted cross-sectional selection effects and also that RS based indicators of vegetation quality hold promise in tracking poverty over time and can do so in areas that night lights cannot. The study was based on a small number of observations, and a larger scale study is needed to verify the findings. The results in this study do justify that a large scale study should be undertaken.

The results also highlight the dependence of the poor in SSA on climate. It raises the question whether decreases in poverty rates were actually due to economic changes or coincidentally bad weather in one survey period and good weather in another. It also suggests that poverty alleviation and prevention policies should focus on poor households' vulnerability to climate change. Of course, effective policies of this kind would imply weaker correlations between vegetation quality and welfare, but RS based vegetation indices could then be used to measure the effectiveness of such policies.

Chapter 4

Poverty Dynamics in Namibia: What can be Learned from Repeated Cross-sections?

4.1 Introduction

Namibia experienced rapid declines in poverty between the early 1990's and the late 2000's. While exact numbers are hard to come by due to survey design changes (Levine and Roberts, 2013), it was at least a 20 percentage point decline. Yet the poverty rate around 2009/10 was still 28.7% and inequality remained stubbornly high over the period with a Gini coefficient of approximately 0.6 (NSA, 2012).

Understanding the nature of poverty and identifying where and why gains were made in certain areas would be of value to policy makers. Barrett (2005) notes that the main argument for studying poverty dynamics is that not all of the poor need policy interventions to escape poverty. At any point in time some of the poor are not chronically poor but are rather in poverty due to stochastic shocks. For these transitively poor households, policy rather needs to focus on ensuring that they do not fall into chronic poverty. Haughton and Khandker (2009) note that for the chronically poor, policies need to focus on raising the average consumption of individuals over time while for the transient poor policies would need to focus on smoothing consumption over time. Barrett (2005) calls the former “cargo net” and the latter “safety net” policies.

Static measures of poverty such as headcount rates have limited value to researchers who want to understand dynamics. Definitions such as chronic and transient poverty are inherently dynamic in nature and require (from an empirical point of view) that the poverty status for the same household is known for more than one time period. Thus, traditionally these measures have always been explored with panel data - data that follows the welfare of households over time.

This requirement of panel data has unfortunately limited the number of countries and periods over which poverty dynamics can be studied. Ironically, countries for which information on poverty dynamics would be most valuable - developing countries - are hit worst by the data requirement.

Namibia is no exception. No household panel data is available in the country. Poverty has been measured using the Namibian Household Income and Expenditure Surveys (NHIES) which were conducted in 1993/94, 2003/04 and 2009/10. A few labour force surveys for the 2010's are also available, yet all data is cross-sectional in nature. Levine and Roberts (2013) argue that good data, and data that is more comparable over time, will be a key guide for policy makers to achieve development goals in Namibia.

The need for more information on the dynamics of poverty, and economic mobility in general, has led researchers to develop methods based on the use of repeated cross-sections. Methods by Bourguignon et al. (2004), Antman and McKenzie (2007) and Dang et al. (2014) have been well documented. While limited in the knowledge the methods can contribute, they do allow researchers insights into mobility that was not possible otherwise.

The method by Dang et al. (2014) is of particular interest for this chapter. This paper will refer to both the method proposed in Dang et al. (2014) and the authors as DLLM. DLLM show how to estimate bounds on the proportion of the population moving in and out of poverty using two cross-sections of data. This can give insights into the distributions of chronic and transitory poverty. DLLM shows that the correlation of residuals of earnings regressions for two points in time determines the extent of earnings mobility over the period. They also show how bounds can be tightened by assuming that the residuals are distributed with a bivariate parametric distribution with correlation coefficient ρ over the period.

In chapter 2, the utility of having good data was highlighted and in chapter 3 daytime satellite data was suggested as a way to overcome data limitations. This chapter uses recently developed econometric techniques to extend the utility of available data, even when the data is not in the ideal format. This study uses the DLLM method to gain a deeper understanding of poverty dynamics in Namibia between 2003/04 and 2009/10 using the NHIES surveys. Following Fields and Viollaz (2013) bounds on the conditional distributions are also calculated. Parametric bounds are also estimated using estimates of ρ obtained from South African panel data.

Results show that poverty is both chronic and transitory in nature. Poverty is also predominantly a rural phenomeno,n with urban areas having low levels of poverty and also low levels of vulnerability to poverty. In mainly poor rural agrarian regions, the dominant nature of poverty varies. In the three regions which account for 50% of the poor in the country, and also have the highest poverty rates, poverty is mainly chronic. These regions - Caprivi, Kavango and Oshikoto - are based in previously disadvantaged areas dominated by smallholder farming.[1] This thesis also highlights that while the method enhances knowledge on mobility in Namibia, it does not overcome the need for panel data. Panel data and data

[1] Note that there have been some changes with regards to regions in Namibia since the release of the data used in this study. Caprivi is now known as Zambezi, and Kavango is split into two regions: Kavango East and Kavango West.

that combines household and agricultural information, such as the data used in chapter 2 would be of great assistance to measure economic mobility. It will also help researchers to understand the intricacies of poverty better in the rural areas of Namibia.

Section 4.2 discusses recent trends in socio-economic outcomes. Section 4.3 discusses concepts and techniques used to study economic mobility with specific focus on instances where panel-data is not available. Section 4.4 discusses the data and methodology used. Section 4.5 presents the results and Section 4.6 concludes.

4.2 Socio-Economic Trends in Namibia

A background of Namibia, and the policies that led to large inequalities were discussed in chapter 3 in section 3.3.2. Here, recent trends in socio-economic outcomes are discussed.

Poverty has fallen dramatically since colonial rule ended in Namibia. Based on the national poverty line the headcount poverty rate was 69.3% in 1993/1994, 37.7% for 2003/04 and 28.7% in 2009/10 (NSA, 2012).[2] Poverty is calculated from the NHIES. Surveys have been conducted in 1993/1994, 2003/2004 and 2009/2010. It has been argued that consumption was under-captured in 1993/1994 and this has shed doubts on the extent of poverty around that period. Levine and Roberts (2013) suggests that the real poverty rate was close to 48.6% in 1993/1994.[3]

Poverty is higher in rural areas and female-headed households are also more likely to be poor. Poverty is particularly high among households that rely on pensions or subsistence farming as their main source of income (NSA, 2012). Poverty is particularly prevalent in the north of the country. Levine and Roberts (2013) note that around two-thirds of the decrease in poverty between 1993/94 and 2003/04 could be attributable to growth in income over the period. Inequality remains a big concern, however, with Namibia regarded as one of the most unequal countries in the world. The Gini-coefficient has remained around 0.6 since 1993/94 (NSA, 2012).

There are large regional disparities in development. For 2009/2010 headcount poverty was above 50% in both Caprivi and Kavango. Khomas and Erongo, the regions with the most important urban centres, have poverty rates of only 7% and 11% respectively (NSA, 2012). Indeed, Levine (2007) estimated the Human Development Index (HDI) for Namibia and found great inequalities in the HDI along lines of language and region. The three highest ranked regions - Karas, Erongo, and Khomas - were all within

[2] The poverty line was estimated using the Cost of Basic Needs approach and is based on per adult equivalent household expenditure (CBS, 2008). This study uses the upper poverty line which was also used in chapter 3.

[3] Levine and Roberts (2013) note that due to the survey changes between 1993/94 and 2003/04 direct comparisons were not possible. To account for this the authors applied techniques originally designed for small-area estimation by Elbers et al. (2003). The method allows researchers to impute income/consumption for datasets that do not have such information. Using this they modelled income using the more reliable NHIES 2003/2004 and then imputed it for the 1993/1994 survey.

areas where settlers could own land before independence, while the three lowest ranked - Ohangwena, Kavango, and Caprivi - were all in regions that were granted self-rule during colonial rule.

Little is known about economic mobility in Namibia. Case studies have looked at the vulnerability of individuals in different settings. Frayne (2004) studied survival strategies of households in Katutura, an informal settlement in Windhoek, Namibia's capital city. He found that households with social links to rural households were less vulnerable to poverty than those with limited ties. Households with uncertain urban incomes and no access to rural farm incomes were the most vulnerable. Furthermore Fara (2001) and Devereux and Naeraa (1996) have discussed household responses and vulnerability to droughts which can possibly push or keep households in poverty.

Levine et al. (2011) have discussed the importance of the social grants system in Namibia. It is a comprehensive and well-established system and the authors argue that it has been effective in alleviating poverty in the country. Using NHIES 2003/04, they calculated what the poverty rate would have been in the absence of social grants, assuming no behavioural changes, externalities and general equilibrium effects. They estimated that poverty would have been 42.0%, instead of 37.8%, implying that the grants decreased that poverty rate by 10.0%. At a lower poverty line they estimated that the effect was a 22.0% decrease.

Even though great strides have been made in poverty alleviation, more than a quarter of the population remain impoverished. Studying economic mobility in the country can greatly improve policy making to support these households and lift them out of poverty.

4.3 Poverty Dynamics

4.3.1 Definition of key concepts

The study of poverty dynamics falls into the broader literature of income mobility. This is the general name given to studies that deal with changes in individuals' income/welfare over time in a society (Jantti and Jenkins, 2013). The literature distinguishes between a number of different measures that each capture different aspects of income mobility and that also have different implications for social welfare.[4]

Fields (2005) notes that when researchers undertake studies in income mobility, they need to be clear about the exact mobility concept being used. Firstly mobility studies can be done on a macro or micro level. Micro mobility studies are concerned with studying the correlates and determinants of mobility at the individual level. Macro mobility studies are concerned with aggregate levels of mobility in a country.

[4] Fields and Ok (1996), Fields (2005) and Jantti and Jenkins (2013) review the different concepts and their implications.

H

Table 4.1: Concepts of Macro Mobility

Concept	Description
Time dependence	How past well-being determines present well-being.
Positional movement	How have people's economic position changed with regards to any number of measures such as ranks or deciles.
Share movement	How people's share of total income has changed.
Income flux	How large peoples' changes in income have been, regardless of the direction.
Directional income movement	How many people have moved (or not moved) in any given direction.
Mobility as an equalizer of longer-term incomes	How income inequality has changed over time.

Source: Fields (2005)

Within macro mobility (which is the focus of Fields (2005)) there are six distinct concepts of mobility which are summarized in Table 4.1. Fields (2005) shows that even on the same dataset, the question of whether mobility increased or decreased can depend on the specific mobility concept used.

Poverty dynamics is concerned with directional income movement and there are a number of conceptual challenges in defining different dynamic poverty statuses. A person could be considered chronically poor if he/she is observed being poor in both period 1 and 2. It is, however, possible that they were observed poor because of stochastic shocks and that they are in fact not chronically poor. These concepts have led researchers to more nuanced definitions of dynamic poverty status.

Haughton and Khandker (2009) distinguish between the following four categories of chronic poverty: The chronically poor - Individuals who are on average below the poverty line over time; The persistently poor - Individuals who are always below the poverty line; The transient poor - Those who move into and out of poverty over time; The never poor - People who never enter poverty.

Carter and May (2001) draw a distinction between stochastic poverty and structural poverty based on individuals' consumption and asset levels. In their framework they assume that household assets are a better indicator of long term well-being than current consumption levels. From this point of departure they estimate an asset poverty line based on the levels of assets that would yield *expected* consumption equal to the consumption poverty line. This framework then allows researchers to distinguish between households that are stochastically poor (below the consumption poverty line, but expected to consume above the poverty line) and those that are structurally poor (consumed and expected to consume below the poverty line). They show that this concept can be expanded to a dynamic framework where a poverty line for the initial period can be defined as the present value of future poverty lines. A "dynamic" asset poverty line can then be inferred which defines individuals with initial asset holdings below the line to be "dynamically" or chronically poor. They have initial asset holdings that are too low for them to escape from poverty. The authors also show how the distinction between stochastic and structural poverty can help distinguish between more detailed definitions of mobility. Households can be structurally or stochastically upward mobile, for example.

In a overview of evidence from developing economies, Baulch and Hoddinott (2000) suggest that the nature of poverty is mostly transient. McKay and Lawson (2003), though, argue that the findings might be misleading. Firstly, the headcount rates in different categories of poverty hide the depth and severity of the poverty which might be worse for chronically poor individuals. Secondly, the influential summary study by Baulch and Hoddinott (2000) uses strict definitions of chronic poverty and lastly, measurement error will tend to overstate the degree of mobility.

4.3.2 Estimation Methods

Jantti and Jenkins (2013) note that while most economic studies on income distribution are cross-sectional in nature, mobility studies take an *"explicitly longitudinal perspective"*. The one defining feature of mobility studies is thus that the same households are followed over time which implies the use of panel data.

Panel data, together with some measure of welfare, can then be used to study some facet of mobility. Popular methods include regressing lagged income on current income, which is referred to as beta-convergence (see Antman and McKenzie (2007)), or calculating transition matrices. Transition matrices can be used to study relative or absolute concepts of mobility. Matrices are estimated with income for 1 period along the rows and income for the other period along the columns. A simple depiction of a poverty transition matrix is shown in table 4.2 which shows the poverty status of individuals for periods 1 and 2 along the rows and columns respectively. Each cell thus shows the percentage of individuals by their joint poverty status over the 2 periods.

Table 4.2: Transition Matrix Example

		Period 2	
		Poor	Not
Period 1	Poor	a	b
	Nonpoor	c	d

Note: $a + b + c + d = 100$

In the hypothetical example in table 4.2, $a\%$ of the population was poor in both periods while $d\%$ was not poor in both. $c\%$ was downwardly mobile while $b\%$ was upwardly mobile. Fields (2005) discusses the use of transition matrices from a theoretical perspective. For empirical examples, readers can refer to Fields et al. (2003) for relative mobility and Carter and May (2001) for an expanded poverty transition matrix. The above definitions can also easily be transferred to a context of positional movement by comparing individuals' ranking in the income distribution over time.

The requirement of longitudinal data has limited the number of countries and time periods where income mobility can be studied. Baulch and Hoddinott (2000) noted (at the time of publication) that the research on poverty dynamics in developing countries was *"remarkably thin"*, especially when compared to the work on poverty from cross-sectional surveys. While the number of countries with panel data surveys has increased substantially over the past decade, it remains scarce. Cuesta et al. (2011) note that while more data has become available in Latin America, the new data only covers short periods and limits analysis.

Apart from the availability of panel data, issues of measurement error and attrition are also of concern to analysts measuring economic mobility. Attrition is of concern because, if it occurs non-randomly, results will be biased. Concerns with measurement error centre around the fact that when income is measured with error, the level of mobility will be overestimated. For example, if households under-report income in one period and correctly report it in the next, it will seem as if households were mobile even with no changes in their true income (Burger et al., 2016).[5] A number of papers have suggested methods for addressing the issue (see for example Burger et al. (2016); Lechtenfeld and Zoch (2014)). Jantti and Jenkins (2013) note that not much is known about the effects of measurement error on mobility outcomes. The small body of research that is available (see Dragoset and Fields (2006); Fields et al. (2003); Gottschalk and Huynh (2010)) downplays the concerns of the problem, but Jantti and Jenkins (2013) caution readers against generalising their findings.

Apart from the above mentioned issues, a less talked about problem is that panel datasets usually have smaller sample sizes than cross-sectional data. Combined with attrition, small samples can call into question the representativeness and also the confidence of estimates, particularly at sub-national levels.

These shortcomings have led researchers to develop methods of estimating dynamic economic relationships in the absence of panel data. McKay and Lawson (2003) suggest, for example, that analysing retrospective information in cross-sections can help estimate the extent of chronic poverty.

Another method which has been used substantially is pseudo-panel analysis - an idea first proposed by Deaton (1985). He suggested that while in repeated cross-sections, specific individuals cannot be followed over time, cohorts, defined as groups with fixed membership, can be followed. Deaton (1985) argued that repeated surveys should generate repeated random samples of cohorts if the samples or cohorts are large

[5] This example assumes that measurement (reporting) error is not correlated over time.

enough. From these repeated random samples it is then possible to create a time series from the summary statistics for the cohorts which can be used to estimate certain economic relationships at cohort level.[6]

Three methods have been developed to measure economic mobility or certain aspects of it when using repeated cross-sections: Bourguignon et al. (2004), Antman and McKenzie (2007), and Dang et al. (2014). These methods will be referred to as BGK, AM and DLLM respectively. This section discusses the broad ideas of each followed by assessments of their effectiveness. For more detail on BGK and AM please refer to the appendix.

Fields and Viollaz (2013) distinguish between the three approaches by the assumptions made about the structural parameters and the mobility question that the method attempts to answer. Along these characteristics they label AM as a mean-based approach, and the other two as dispersion-based approaches. Mean-based approaches follow cohorts of individuals over time from repeated cross-sectional data, while dispersion-based approaches construct estimates of mobility from the variance of the error term.

AM proposed a method to estimate beta-convergence that is useful in dealing with measurement error and also able to estimate beta-convergence in the absence of panel data. They take cohort averages of income from repeated cross-sections. This deals with the measurement error in income and also allows for the creation of a cohort-level pseudo-panel. Based on a few assumptions, they then show how beta-convergence at the cohort level can be estimated, with the most basic specification being:

$$\bar{Y}_{c(t),t} = \alpha + \beta \bar{Y}_{c(t-1),t-1} + \bar{\mu}_{c(t),t}, \tag{4.1}$$

where $Y_{c(t)t}$ refers to the average income of cohort c, observed in time period t, for the period t.[7]

BGK show how certain dynamic parameters of the income generating process can be extracted and then used to estimate individual vulnerability to poverty. They follow work by Deaton and Paxson (1994) who studied the second-order moments of cohorts in a pseudo panel context. BGK simplify their idea by stating:

> "if it may be assumed that all individuals within a cohort face a stochastic earning process that has common characteristics, these characteristics may be recovered at the aggregate level, without observing actual earning paths."

BGK assume cohort specific income generating processes for individual i, who is a member of cohort c in period t of the form $y_{i,t,c} = \beta_{t,c} x_{i,t,c} + e_{i,t,c}$, and assume that the error follows the following autoregressive ($AR(1)$) process:

[6] According to Antman and McKenzie (2007) and Cuesta et al. (2011), the conditions for the consistent estimation of dynamic pseudo-panel models has since been discussed by numerous authors including Moffitt (1993), Collado (1997), Girma (2000), McKenzie (2004) and Verbeek and Vella (2005)

[7] For a clearer understanding of the notation please refer to the appendix.

$$e_{it}^{j} = \rho^{j} e_{it-1}^{j} + \epsilon_{it}^{j} \tag{4.2}$$

where ρ is a parameter measuring the persistence of earnings shocks, and, ϵ_{it}^{j} is what the authors call the innovation in earnings which has a variance $\sigma_{\epsilon jt}^{2}$.

Under a set of simplifying assumptions they show that it is possible to extract ρ^{c} and also $\sigma_{\epsilon,t,c}^{2}$. These parameters can then be used to estimate individual vulnerability to poverty.

DLLM show how to estimate bounds of the joint distribution of poverty over two periods (similar to table 4.2). They start with a data generating proses for income, $y_{i,t} = \beta_t x_{i,t} + e_{i,t}$, and show that the joint distribution of poverty over time is dependant on the correlation between the two residual terms.

If the residuals over two periods are perfectly correlated, mobility will be at its lowest bound. Similarly, if the residuals are independent mobility will be at its topmost bound. This captures the intuition that the less (more) dependant individuals' income is on past income, the more (less) economically mobile they are.

These methods have been tested using true panel data. Apart from the authors who published the methods (and found them) to be accurate, Fields and Viollaz (2013) test all three methods and Cruces et al. (2015) also validate DLLM.

Fields and Viollaz (2013) used 3 waves of Chilean panel data stretching from 1996-2006 to compare true mobility estimates to psuedo-panel estimates. They found that none of the above mentioned methods performed particularly well. They found that unconditional beta convergence using AM was close to the true estimate and that DLLM estimated the bounds correctly except for the proportion that were non-poor in both periods.

On a final point Fields and Viollaz (2013) argue that the DLLM method gives valuable information on poverty movements, but they do not say anything about poverty dynamics (which is a conditional concept). Poverty dynamics is concerned with answering what the likelihood is of being in a certain poverty status in t_2, given a certain poverty status in t_1. The DLLM bounds on the joint distribution of poverty over two periods can easily be tranlated into the bounds of the conditional distribution. Fields and Viollaz (2013) did this using true panel data and found that the true conditional estimates are sandwiched by the bounds, but that the bounds are so large that they do not provide valuable information.

Cruces et al. (2015) tested DLLM using panel data from three countries: Chile, using the same data as Fields and Viollaz (2013), Nicaragua and Peru.[8] They found that the method sandwiched the bounds in all cases. They tested the robustness to different measures of wealth (income and consumption), time

[8] Fields and Viollaz (2013) notes that there are two possibilities why their results differ from Cruces et al. (2015). The first is that used a different age cut-off and the second is that they used a different method of dealing with outliers.

frames and the introduction to multiple thresholds (expanded poverty transition matrix) and found the method to produce the correct results each time. One advantage that the authors note of the DLLM approach compared to the other two is that it imposes fewer structural restrictions on the data generating process.

The development of these methods has allowed researchers to explore aspects of economic mobility in countries where it was not possible before, due to a lack of panel data. The usefulness is not only limited to countries without panel data. In many countries panel data surveys have only been conducted recently and cover short time frames, while repeated cross-sections stretch back further in time and thus allow for a long term perspective on economic mobility. Cuesta et al. (2011) used the AM method to study economic mobility for 14 countries in Latin America over the period 1992 - 2003. Bierbaum and Gassmann (2012) use the methods for the Kyrgyz Republic using the Kyrgyz Integrated Household Survey. While the survey does have a panel component, they prefer to use the cross-sections, as the panel has shortcomings. The development of these methods is also what allows us to study mobility in Namibia.

4.4 Data and Methodology

4.4.1 Semi-Parametric Approach

For the study of poverty dynamics in Namibia, the DLLM method is the only suitable psuedo-panel method. Only two survey periods are available and both BGK and AM are less precise if only a small number of cross-sections are used.

DLLM is practically implemented by predicting an upper and lower level of consumption that individuals would have had in the period in which they are not observed. This follows small-area estimation methods developed by Elbers et al. (2003). The method was originally developed to estimate poverty at small geographic area. Census data contains representative samples at small geographic levels, but usually lacks income data. Household surveys, on the other hand, usually contain accurate income data but samples are only representative at larger geographic regions. Elbers et al. (2003) used household surveys to create a prediction model for income based on variables in both surveys, and then predicted consumption for households in the census data.

DLLM show how the predicted levels of consumption are dependent on the joint distribution of the residuals in consumption regressions. Firstly, individuals' (i) consumption (y), at time t, is expressed as:

$$y_{it} = \beta x_{it} + e_{it}, \tag{4.3}$$

where x_{it} is a set of time invariant observables, and e_{it} is the error.

Assume then that there is some poverty line z_t for period t. We are interested in the joint probability of being either poor or non-poor in the first period and also the in the second. In the case of measuring the extent to which individuals transition out of poverty between two periods one needs to estimate the probability of individuals being in poverty in the first period but not in the second:

$$P(y_{i1} < z_1 \, and \, y_{i2} > z_2). \tag{4.4}$$

In repeated cross-sections we do not observe the same individuals in both periods, which disallows direct estimation of this probability. The probability can be rewritten as:

$$P(e_{i1} < z_1 - \beta_{i1}x_{i1} \, and \, e_{i2} > z_2 - \beta_{i2}x_{i2}). \tag{4.5}$$

This shows that the probability therefore depends on the joint distribution of e_{i1} and e_{i2}, which captures the autocorrelation of the parts of consumption that are not explained by the observable characteristics in those periods. Intuitively, the less correlated e_{i1} and e_{i2}, the more mobility arises. By making assumptions on the joint distribution of the error, bounds can be obtained on the estimates of mobility. At one extreme, it can be assumed that the errors are perfectly correlated. This implies that consumption is completely dependant on past consumption which implies no mobility. Thus, the assumption of perfectly correlated errors will produce a non-parametric lower-bound to mobility. At the other extreme it can be assumed that the errors are not correlated at all. Using a similar logic as above, this will yield a non-parametric upper-bound to mobility.

The formulas (discussed below) which are used to obtain the bounds are dependant on two assumptions. The first is that the underlying population sampled is the same in both surveys. This can be verified by comparing time invariant characteristics of a cohort over time. The second assumption is that the errors are positive quadrant dependent (PQD) which means that the correlation between the error terms cannot be negative. They provide three reasons why the assumption is reasonable. Firstly, income or consumption shocks are expected to be fairly persistent over time. Secondly, poverty is fairly persistent over time, which implies that the joint probability of being in poverty in both periods is expected to be higher than the product of the individual probabilities of being poor in both periods. Finally, the authors note that while for some individuals the correlation between errors could be negative, for the majority it is expected to be positive. To satisfy these assumptions, they note that it is best to limit the sample to an age cohort that is not too young or too old, such as ages 25 - 55, which is a standard practise in pseudo-panel analysis. From these assumptions they propose the following theories to obtain the bounds on mobility. They also show that the estimates calculated in equations 4-11 are robust to classical measurement error and also to non-classical measurement error if the PQD assumption is not violated. Refer to appendix 1 of DLLM for the proof.

For clarity, please note that the upper-bound for mobility, which is obtained when $corr(e_{i1}, e_{i2}) = 0$, is also the lower-bound to immobility. Thus, for the joint distributions of moving into and out of poverty, the assumption $corr(e_{i1}, e_{i2}) = 0$ provides the upper-bound, and for joint distributions of staying out of poverty, the same assumption provides the lower-bound. The opposite holds when $corr(e_{i1}, e_{i2}) = 1$.

Independent Errors

Based on the 2 assumptions required above and $corr(e_{i1}, e_{i2}) = 0$, the upper-bound estimate of the probability of being poor in period 1 and not poor in period 2 is

$$P(y_{i1}^{2u} < z_1 \; and \; y_{i2} > z_2) = P(e_{i1} < z_1 - \beta_1' x_{i2}) P(e_{i2} > z_2 - \beta_2' x_{i2}), \tag{4.6}$$

where y_{i1}^{2u} refers to the estimated upper-bound consumption of period 2 households in period 1 which is $y_{i1}^{2u} = \beta_1' x_{i2} + e_{i1}$. Note that if real-consumptions levels are used $z_1 = z_2$. Given that the errors are independent, e_{i1} is that is randomly sampled (with replacement) from the sample of estimated residuals $(\hat{e}_1)$.

The upper-bound estimate of the probability of not being poor in period 1 but being poor in period 2 is given by

$$P(y_{i1}^{2u} > z_1 \; and \; y_{i2} < z_2) = P(e_{i1} > z_1 - \beta_1' x_{i2}) P(e_{i2} < z_2 - \beta_2' x_{i2}). \tag{4.7}$$

The lower-bound estimate of the probability of not being poor in both periods is

$$P(y_{i1}^{2u} > z_1 \; and \; y_{i2} > z_2) = P(y_{i2} > z_2) - P(y_{i1}^{2u} < z_1 \; and \; y_{i2} > z_2). \tag{4.8}$$

and the lower-bound estimate of the probability of being poor in both periods is

$$P(y_{i1}^{2u} < z_1 \; and \; y_{i2} < z_2) = P(y_{i2} < z_2) - P(y_{i1}^{2u} > z_1 \; and \; y_{i2} < z_2). \tag{4.9}$$

Completely dependent Errors

When $corr(e_{i1}, e_{i2}) = 1$ the lower-bound estimates of mobility can be obtained and similarly the upper-bound estimates of immobility. The lower-bound estimate of the probability of being poor in period 1 and not poor in period 2 is

$$P(y_{i1}^{2l} < z_1 \; and \; y_{i2} > z_2) = P(\eta e_{i2} < z_1 - \beta_1' x_{i2}) - P(e_{i2} \leq z_2 - \beta_2' x_{i2}), \tag{4.10}$$

where y_{i1}^{2l} is the estimated consumption of period 2 households for period 1 for lower-bound mobility. $y_{i1}^{2l} = \beta_1' x_{i2} + \eta e_{i2}$ where $\eta = \sqrt{\frac{Var(e_{i1})}{Var(e_{i2})}}$.

The lower-bound estimate of the probability of not being poor in period 1 but being poor in period 2 is given by

$$P(y_{i1}^{2l} > z_1 \, and \, y_{i2} < z_2) = P(e_{i2} < z_2 - \beta_2' x_{i2}) - P(\eta e_{i2} \leq z_1 - \beta_1' x_{i2}). \tag{4.11}$$

The upper-bound estimate of the probability of not being poor in both periods is

$$P(y_{i1}^{2l} > z_1 \, and \, y_{i2} > z_2) = P(y_{i2} > z_2) - P(y_{i1}^{2l} < z_1 \, and \, y_{i2} > z_2). \tag{4.12}$$

and the upper-bound estimate of the probability of being poor in both periods is

$$P(y_{i1}^{2l} < z_1 \, and \, y_{i2} < z_2) = P(y_{i2} < z_2) - P(y_{i1}^{2l} > z_1 \, and \, y_{i2} < z_2). \tag{4.13}$$

Estimation of Non-parametric Bounds

4.4.2 Parametric Approach

DLLM further show that it is possible to obtain narrower bounds by assuming that the errors follow a parametrized statistical distribution, such as a bivariate normal distribution with a correlation coefficient ρ. Using true estimates of ρ from countries comparable to the country being studied, one can obtain sharper estimates. Values for ρ can be obtained from actual panel data. This can be done by estimating consumption regressions using time-invariant characteristics for each wave of data and then estimating the correlation between the estimated residuals from each regression. Assuming that ρ is equal to one and zero will give the parametric equivalents of the non-parametric estimates above.

Assuming that e_1 and e_2 have a bivariate normal distribution with standard deviations σ_{e_1} and σ_{e_2} and a correlation coefficient $\rho \geq 0$, the probability that households are poor in period 1 and not poor in period 2 is:

$$P(y_{i1} < z_1 \text{ and } y_{i2} > z_2) = P(e_{i1} < z_1 - \beta_1' x_{i2}) P(e_{i2} > z_2 - \beta_2' x_{i2}) = \Phi(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, \frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho), \tag{4.14}$$

where $\Phi(.)$ refers to the bivariate normal cumulative distribution function.

Using the bounds approach it is assumed that the true ρ lies somewhere between an upper and lower bound value $\rho_u \, and \, \rho_l$ where $0 < \rho_l < \rho_u < 1$. A lower value of ρ implies a higher level of mobility. Thus the lower and upper bound estimates of the joint probabilities for the different poverty statuses are:

$$P(y_{i1} < z_1 \, and \, y_{i2} < z_2) = \Phi(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, \frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho) \tag{4.15}$$

$$P(y_{i1} < z_1 \, and \, y_{i2} > z_2) = \Phi(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, -\frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho) \tag{4.16}$$

$$P(y_{i1} > z_1 \, and \, y_{i2} < z_2) = \Phi(-\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, \frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho) \tag{4.17}$$

$$P(y_{i1} > z_1 \, and \, y_{i2} > z_2) = \Phi(-\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, -\frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho) \tag{4.18}$$

where $\rho = \rho_u$ for the upper bound estimates and $\rho = \rho_l$ for the lower bound estimates.

4.4.3 Estimation Approach

To obtain the non-parametric bounds of mobility, consumption is modelled for both periods 1 and 2 using the same set of variables and the model specified in equation 1. From the period 1 model the vectors $\hat{\beta}_1$ and $\hat{e}_1$ are obtained. y_{i1}^{2u} is then calculated using the observable characteristics of households in period 2 with the variable coefficients of period 1 ($\hat{\beta}_1$) and the randomly sampled (with replacement) residuals ($\hat{e}_1$). To obtain consistent estimates of equations 4.6-4.9, this process is repeated 500 times and the average of the estimates is taken. y_{i1}^{2l} is calculated in the same manner but uses the estimated residuals for individuals in period 2, $\hat{e}_{i2}$, which is scaled by $\eta = \sqrt{\frac{Var(\hat{e}_{i1})}{Var(\hat{e}_{i2})}}$. As done in Fields and Viollaz (2013), the bounds obtained above are also reworked to reflect the poverty status in period 2, conditional on poverty status in period 1 which gives a dynamic interpretation to the bounds obtained by DLLM. In the discussion of the results, the nature of poverty - chronic or transitive - will be defined by the conditional distributions as this provides a more dynamic interpretation.

To get an idea of locational transitions, analysis was also done individually for all 13 regions in the country. The validity of these results rely on the assumption that the underlying population within each region did not change systematically over the period. This is discussed in the data section.

4.4.4 Data

The data used for the analysis come from the 2003/04 and 2009/10 NHIES surveys. The purpose of the surveys was to gain accurate information on income, expenditure, and consumption in the country and the surveys serve as the only valid source from which to calculate poverty estimates[9]. The surveys used a stratified two-stage sampling design. The questionnaires were kept the same to make them as comparable as possible. This is unlike the 1993/94 survey, that differed in a number of ways from the 2003/04 data which makes using it in this exercise problematic. It was half the sample size of the 2003/04 survey, later surveys used modern technology for data capturing, infrequent non-food expenditures were captured worse when compared to the later surveys, and finally, to be considered as a household member

[9] A few other cross-sections are available including Demographic and Health Surveys for 1992, 2000 and 2006/07, and Labour Force Surveys for 2012-2014.

you had to be residing in the house for at least 1 week in the past month, compared to two weeks in the later surveys (Levine and Roberts, 2013). For 2003/04 and 2009/10, 10 920 and 9 801 households were sampled respectively. As is suggested by DLLM and other pseudo-panel procedures, the samples are also limited to households where the head was between the ages of 25 and 55 at the time of the first survey. This, together with the exclusion of households with missing data for the estimation, leaves 6 618 for 2003/04 and 5 911 for 2009/10 that is used in the analysis.

The DLLM method requires that consumption is modelled using time invariant characteristics. In the NHIES surveys, few variables satisfy this condition. For this chapter, consumption is modelled using gender, age in the first time period of analysis, level of education, all for the household head, and household language.

To check if the variables remain comparable over time, Table 4.4 shows the means of the variables for both surveys. The differences in means are also shown and whether the differences are significantly different from zero, which was estimated using t-tests. The results suggest the means remained consistent over time. Where the differences were significantly different from zero, they were small. The significant difference in age is less than a year, for example. The largest difference was an increase of 3 percentage points in household heads with secondary education.

As discussed earlier, for the region specific estimates to be valid, the underlying populations within each region should also have remained constant of the period. There are no migration statistics available from the specific period. Namibia's 2011 Census though, does give some insights into migration between 2010 and 2011 and also lifetime migration. Data shows that 1.9% of the population migrated to a different region between 2010 and 2011. Not surprisingly, the proportion of Namibians who lived in different regions from where they were born was 22.5%. The most likely destinations for lifetime migrants in Namibia were Khomas and Erongo - regions with main urban centres (NSA, 2015a).

Yet, while migration is present, DLLM requires that the underlying sample did not change over the two survey periods. To test this, t-tests on the mean differences are conducted for each region. Table 4.5 shows the differences in means, and their significance for the variables of interest for the 13 regions. The results suggests that there were some shifts in the sample characteristics within some of the regions. In Erongo, the proportion of household heads with secondary and tertiary education increased by around 10 percentage points and the proportion with primary education decreases by the same amount. Oshikoto had 10 percentage points more female households in 2009/10 compared to 2003/04. In Khomas, Oshana and Otjozondjupa the average age (in 2004) of household heads decreased by more than 1 and a half years. While there are some significant differences in the samples, there does not seem to be extreme differences over the two survey periods. his study will therefore report the results at the regional level, but they results should be interpreted with caution.

To narrow the bounds using the parametric approach, ρ is estimated using panel data from South Africa. South Africa is similar to Namibia in a number of ways. Both countries were subject to apartheid

Table 4.3: Data Description

	Survey Year	Mean	St. Dev.	Median	Min.	Max
Real Expenditure	*03/04*	19830.16	35288.77	9274.61	429.24	744601.25
	09/10	22285.54	40527.67	10269.37	461.66	773265.63
Age in 2003/2004	*03/04*	38.85	8.30	38.00	25.00	55.00
	09/10	38.08	8.57	37.00	25.00	55.00
Female	*03/04*	0.39	0.49	0.00	0.00	1.00
	09/10	0.40	0.49	0.00	0.00	1.00
Education						
No Education	*03/04*	0.15	0.36	0.00	0.00	1.00
	09/10	0.15	0.35	0.00	0.00	1.00
Primary	*03/04*	0.30	0.46	0.00	0.00	1.00
	09/10	0.29	0.45	0.00	0.00	1.00
Secondary	*03/04*	0.42	0.49	0.00	0.00	1.00
	09/10	0.45	0.50	0.00	0.00	1.00
Tertiary	*03/04*	0.12	0.33	0.00	0.00	1.00
	09/10	0.12	0.32	0.00	0.00	1.00
HHold Language						
Khoisan	*03/04*	0.02	0.14	0.00	0.00	1.00
	09/10	0.01	0.11	0.00	0.00	1.00
Caprivi Languages	*03/04*	0.08	0.27	0.00	0.00	1.00
	09/10	0.07	0.26	0.00	0.00	1.00
Otjiherero	*03/04*	0.08	0.28	0.00	0.00	1.00
	09/10	0.09	0.29	0.00	0.00	1.00
Rukavango	*03/04*	0.08	0.28	0.00	0.00	1.00
	09/10	0.11	0.31	0.00	0.00	1.00
Nama/Damara	*03/04*	0.16	0.37	0.00	0.00	1.00
	09/10	0.16	0.37	0.00	0.00	1.00
Oshiwambo	*03/04*	0.41	0.49	0.00	0.00	1.00
	09/10	0.40	0.49	0.00	0.00	1.00
Setswana	*03/04*	0.00	0.07	0.00	0.00	1.00
	09/10	0.00	0.06	0.00	0.00	1.00
Afrikaans	*03/04*	0.12	0.32	0.00	0.00	1.00
	09/10	0.12	0.33	0.00	0.00	1.00
German	*03/04*	0.01	0.08	0.00	0.00	1.00
	09/10	0.01	0.08	0.00	0.00	1.00
English	*03/04*	0.02	0.13	0.00	0.00	1.00
	09/10	0.02	0.13	0.00	0.00	1.00
Other	*03/04*	0.01	0.10	0.00	0.00	1.00
	09/10	0.01	0.10	0.00	0.00	1.00

Source: Own Calculations using NHIES 2003/04 and 2009/10.
The sample is limited to households where the head is between 25 or older and 55 or younger.
Individual Characteristics are all for the household head.

rule, that created highly unequal societies. In the early 1990's both also abolished apartheid and held democratic elections. Both have large regional disparities in living standards and have similar levels of inequality.

The National Income Dynamics Study (NIDS) data is used to calibrate ρ for South Africa. The first round of NIDS was conducted in 2008 with subsequent waves in 2010, 2012 and 2014. It has to be noted that the period that NIDS covers corresponds with the financial crisis and covers a period after our analysis period for Namibia. This study uses the first three periods to obtain estimates for ρ. Per capita household expenditure was regressed on a set of time-invariant explanatory variables for household heads.

Table 4.4: Sample Mean Differences between 2003/04 and 2009/10

	Mean(03/04)	Mean(09/10)	Difference
Age in 2003/2004	38.85	38.08	0.77***
Female	0.39	0.40	-0.01
Education			
No Education	0.15	0.15	0.01
Primary	0.30	0.29	0.02**
Secondary	0.42	0.45	-0.03***
Tertiary	0.12	0.12	0.00
HHold Language			
Khoisan	0.02	0.01	0.01***
Caprivi	0.08	0.07	0.00
Otjiherero	0.08	0.09	-0.01
Rukavango	0.08	0.11	-0.02***
Nama/Damara	0.16	0.16	-0.00
Oshiwambo	0.41	0.40	0.02*
Setswana	0.00	0.00	0.00
Afrikaans	0.12	0.12	-0.00
German	0.01	0.01	-0.00
English	0.02	0.02	0.00
Other	0.01	0.01	0.00

Source: Own Calculations using NHIES 2003/04 and 2009/10.
The sample is limited to households where the head is between 25 or older and 55 or younger.
Individual Characteristics are all for the household head.

As in Namibia, the sample is limited to households where heads are between (and including) the ages of 25 and 55. Only the balanced sample is used which leaves 2596 households for estimation purposes.

The benefit of having true panel data is that retrospective information can also be incorporated in consumption models. Explanatory variables for the models include age, province, a rural dummy and car ownership, all for 2008. Furthermore race, years of education and gender are also included. Models were estimated on consumption in each of the three years and the residuals were calculated. The correlation of the residuals between the different year pairs was then calculated and all estimates were weighted. Table C.1 shows the regression estimates and table C.2 shows the estimates of ρ.

The estimates of ρ in South Africa for the period are 0.51 (2008-2010), 0.56 (2010-2012) and 0.48 (2008-2012). Correlations are higher for the 2 year comparisons than the four year comparisons. This would be expected, given that the movements of consumption should become less dependant on lagged consumption as the time period increases. DLLM presents estimates of ρ for Bosnia-Herzegovina, Indonesia, Lao PDR, Nepal, Peru an Vietnam for periods more similar to the analysis period in this study, with estimates ranging between 0.39 and 0.66.

In this study, results will be shown for two different combinations of ρ. The first is the parametric equivalent of the non-parametric bounds where $\rho = 0$ for an upper bound estimate to mobility and $\rho = 1$ for a lower bound estimate. In the second set of parametric estimate ρ is taken to be 0.35 for the upper bound to mobility and 0.7 for the lower bound. This covers the range of estimates found in NIDS and the countries used in DLLM. This helps to address concerns about ρ based on South Africa which covers a different time period.

The parametric estimates assume that the error terms are bivariate normally distributed with correlation ρ. It is questionable whether the normality assumption is satisfied, however. In DLLM formal tests dismissed the normality of the error terms. Formal tests of the error terms in NHIES also dismissed the normality assumption. However, normality is an assumption that is generally made with regards to logged income distributions and as such analysis will still be shown but are to be interpreted with caution.

As a validity test for DLLM, the bounds on the joint distribution of poverty over time are also calculated for South Africa for 2008-2012 using NIDS, and compared to the true estimates from NIDS. The results, presented in table C.3, show that the non-parametric bounds do cover the true estimates. Apart from the probablity of being poor in 2008 and not poor in 2012, the parametric bounds do not cover the true estimates, however. Even using the most conservative specification ($\rho = \{0, 1\}$), the estimates fail for the probability of not being poor in both periods. The fact that the true estimates are not covered by the parametric bounds suggests that the model residuals for South Africa from NIDS are not bivariate normally distributed over time.

Poverty is measured using per adult equivalent consumption in Namibia and an annual poverty line of N$4535.52 in 2009/10 prices is used.[10] All models are estimated using real consumption at 2009/10 prices. Models are estimated at the household level using individual characteristics of the household head. Final estimates of the joint distribution of poverty are then estimated taking account of household size to get estimates for the full population. All models and tables are also estimated using survey weights.[11]

[10] See CBS (2008) for calculation of the poverty line.

[11] Estimation is done in STATA using code adapted from DLLM wich is available at `http://siteresources.worldbank.org/DEC/Resources/McKenzie_ReplicationFilesforMobility.zip`

Table 4.5: Sample Mean Differences between 2003/04 and 2009/10 for Regions in Namibia

	Caprivi	Erongo	Hardap	Karas	Kavango	Khomas	Kunene	Ohangwena	Omaheke	Omusati	Oshana	Oshikoto	Otjozondjupa
Age in 2003/2004	-0.07	0.74	0.79	1.50***	0.95*	1.53***	-0.70	0.25	-0.06	-0.50	1.78***	-0.11	1.54***
Female	-0.00	0.01	-0.06*	-0.04	-0.04	0.03	0.06	-0.05	-0.07**	0.01	-0.00	-0.10***	-0.02
Education													
No Education	-0.00	-0.00	0.04*	0.04***	-0.06**	0.00	-0.03	-0.04*	-0.02	-0.01	0.02	0.06**	-0.01
Primary	0.02	0.10***	-0.03	0.01	0.05*	-0.01	0.09**	-0.01	0.01	0.01	0.03	-0.04	0.04
Secondary	-0.00	-0.05*	0.01	-0.08**	0.00	0.02	-0.02	-0.01	-0.02	-0.07**	-0.04	-0.03	-0.01
Tertiary	-0.02	-0.04**	-0.02	0.03	0.00	-0.01	-0.03	0.06**	0.03	0.07***	-0.01	0.01	-0.02
HHold Language													
Khoisan	-0.00	-0.00	0.02***	-0.00	-0.00	-0.00	0.01*	-0.00	-0.00	0.00	0.00	0.04***	0.02
Caprivi	0.04*	-0.00	0.00	-0.01	0.01	0.00	0.01	0.01	0.01	0.01**	-0.01**	0.00	0.01
Otjiherero	0.00	-0.07***	0.00	-0.01*	-0.01	0.03**	-0.03	-0.00	-0.09**	-0.00	-0.00	-0.01	-0.01
Rukavango	-0.02	0.00	-0.03***	-0.03**	-0.02	-0.03***	0.01	0.00	-0.00	0.00	-0.00	0.03***	-0.01
Nama/Damara	-0.00	0.04	0.01	-0.04	0.00	-0.01	-0.02	-0.00	-0.02	0.00	-0.00	0.04**	0.07***
Oshiwambo	-0.01	0.03	-0.02	0.04	-0.01	0.05**	0.02	0.00	0.02	-0.04***	-0.00	-0.12***	-0.04*
Setswana	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00
Afrikaans	-0.00	0.02	0.02	0.03	0.01**	-0.03	-0.01	-0.00	0.07***	0.00	0.01	-0.00	-0.02
German	0.00	-0.00	0.00	0.00	0.00	-0.01	0.00	0.00	-0.00	0.00	0.00	0.00*	0.00
English	-0.01	-0.01	-0.00	0.01*	0.00	-0.01	0.00	-0.00	0.01	0.01	0.01	0.01	-0.01*
Other	0.00	-0.00	-0.00	0.00	0.02*	-0.01	0.00	-0.01	0.00	0.01**	0.00	-0.00	-0.00

Source: Own Calculations using NHIES 2003/04 and 2009/10.
The sample is limited to households where the head is between 25 or older and 55 or younger.
Individual Characteristics are all for the household head.

4.5 Results

To understand the results in context, the headcount poverty rates were calculated for both surveys for households with heads aged between 25 and 55 in 2003/04. Table 4.6 shows the headcount rate for Namibia and all the 13 regions in the country. Table 4.7 shows the regional poverty shares across the country for both survey periods.

Table 4.6: Headcount Poverty in Namibia

	2003/04	2009/10
Caprivi	34.36	46.82
Erongo	10.79	6.91
Hardap	39.45	25.99
Karas	29.46	22.15
Kavango	60.17	54.84
Khomas	7.55	10.40
Kunene	31.67	27.92
Ohangwena	33.65	22.65
Omaheke	41.71	31.17
Omusati	27.83	16.48
Oshana	17.80	14.82
Oshikoto	37.08	42.24
Otjozondjupa	34.73	31.03
Namibia	28.77	26.20

Source: Own calculations using NHIES 2003/04 and 2009/10.
The sample is limited to households where the head is between 25 or older and 55 or younger.

Two important points can be noted from table 4.6. The first is that the headcount rate for the estimation sample is lower than for the full population which was 37.7% (2003/04) and 28.7% (2009/2010) in the two periods. Secondly, the decrease in poverty among the age cohort of interest was smaller than for the full sample. This suggests that poverty was more prevalent in households with older and/or younger heads and that much of the decreases in poverty between the periods was due to changes in expenditure for those households. In this context, the proportion of the population who moved out of poverty over the period will be lower for the estimation sample than for the whole population.

As for the locational aspects of poverty for the estimation sample, poverty is highest in regions that are dominated by smallholder agriculture, such as Caprivi, where it increased from 34% to 47%, and Kavango, where the headcount poverty rate for the region was above 50% for both periods. Kavango is also the region that contains the largest share of the poor in Namibia with 30% of the country's poor in 2009/10. Erongo and Khomas, regions containing the main urban centres of the country, have the lowest poverty rates.

With the poverty rates as background, the bounds on the joint distribution of poverty is discussed next. Firstly, table 4.8 shows the results of OLS regressions on consumption for the two survey periods which are used to obtain β_1 and β_2. For both years, female headed households consume significantly less and

Table 4.7: Regional Distribution of Poverty in Namibia

	2003/04	2009/10
Caprivi	6.47	9.03
Erongo	2.67	2.08
Hardap	5.71	3.61
Karas	4.14	3.40
Kavango	24.93	30.49
Khomas	5.37	8.00
Kunene	3.84	4.06
Ohangwena	8.88	7.45
Omaheke	5.11	3.72
Omusati	7.99	5.11
Oshana	4.95	4.37
Oshikoto	9.60	11.20
Otjozondjupa	10.34	7.48
Namibia	100.00	100.00

Source: Own calculations using NHIES 2003/04 and 2009/10.
The sample is limited to households where the head is between 25 or older and 55 or younger in 2003/04.

consumption increases with education levels. The model produces a relatively high adjusted R^2 of 0.49 for 2003/04 and 0.48 for 2009/10. In DLLM, the authors obtain an adjusted R^2 of 0.395 for Indonesia in their most exhaustive model that is least likely to satisfy the assumptions. Given that these models have good predictive power, it would be expected that the method would produce relatively narrow bounds, which are discussed next.

Tables 4.9 and 4.10 show the bounds on the joint and conditional distributions respectively. To reiterate, the conditional distributions present a interpretation of poverty dynamics whereas the joint distribution provides insights in total poverty movements. This paper uses the conditional distribution to distinguish whether poverty is chronic - $P(Poor_{10}|Poor_{04})$ - or transitive - $P(NotPoor_{10}|Poor_{04})$ - in nature. Between 10.5% and 25.7% of the population remained poor over the period and between 0.5% and 15.7% moved into poverty. The parametric assumptions (which tighten the bounds) suggests that a larger proportion of poor in 2009/10 were poor in 2003/04. The non-parametric bounds cannot distinguish whether more people moved into poverty than out of poverty, but the parametric estimates suggest that the proportion that moved out was larger. This is consistent with the fact that the poverty rate decreased over the period. Finally, between 60% and 70% of the population did not experience poverty over the period.

The conditional distributions produce wide bounds but do provide some insights. The probability of not falling into poverty over the period was at least 79.2%. The parametric estimates place the probability closer to 90%. This suggests that vulnerability to poverty is low in Namibia. The estimates cannot distinguish whether the probability of moving out, or staying in poverty differed from each other, suggesting that the nature of poverty in Namibia was both chronic and transitory.

Table 4.8: OLS Regression models of ln(per Adult Equivalent Expenditure)

	2003/04	2009/10
Age	-0.03**	0.01
	(0.01)	(0.01)
Age^2	0.00**	-0.00
	(0.00)	(0.00)
Female	-0.22***	-0.28***
	(0.02)	(0.02)
Education (ref: None)		
Primary	0.18***	0.27***
	(0.03)	(0.03)
Secondary	0.72***	0.75***
	(0.03)	(0.04)
Tertiary	1.74***	1.65***
	(0.05)	(0.05)
Language	Yes	Yes
Observations	6618	5911
Adjusted R^2	0.49	0.48

Source: Own calculations using NHIES 2003/04 and 2009/10.
The sample is limited to households where the head is between 25 or older and 55 or younger.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.9: Bounds on the Joint Distribution of Poverty Status over Time

Method	$P_{04}; P_{10}$		$N_{04}; P_{10}$		$P_{04}; N_{10}$		$N_{04}; N_{10}$	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Non Parametric	25.7	10.5	0.5	15.7	5.3	14.1	68.5	59.6
Parametric								
$\rho = \{0.7, 0.35\}$	14.5	11.0	5.9	9.4	9.5	13.0	70.1	66.6

Source: Own calculations using NHIES 2003/04 and 2009/10.

The sample is limited to households where the head is between 25 or older and 55 or younger.

P_{04} and P_{10} refer to "Poor" in 2003/04 and 2009/10 respectively. Similarly N refers to "Not Poor". "Lower" and "Upper" refer to the two bounds on **mobility**. For **immobility** "Lower" is the upper bound and "Upper" is the lower bound.

Table 4.10: Bounds on the Distribution of Poverty in 2009/10, Conditional on Poverty in 2003/04

Method	$P_{10}\|P_{04}$		$N_{10}\|P_{04}$		$P_{10}\|N_{04}$		$N_{10}\|N_{04}$	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Non Parametric	82.9	42.7	17.1	57.3	0.7	20.8	99.3	79.2
Parametric								
$\rho = \{0.7, 0.35\}$	60.5	45.8	39.5	54.2	7.7	12.4	92.3	87.6

Source: Own calculations using NHIES 2003/04 and 2009/10.

The sample is limited to households where the head is between 25 or older and 55 or younger.

P_{04} and P_{10} refer to "Poor" in 2003/04 and 2009/10 respectively. Similarly N refers to "Not Poor". "Lower" and "Upper" refer to the two bounds on **mobility**. For **immobility** "Lower" is the upper bound and "Upper" is the lower bound.

The estimates at the regional level discussed next could shed more light on whether the nature of poverty is locational specific. Table 4.11 shows the bounds on the joint probability for each region and Table 4.12 show the conditional distributions. To reiterate, the regional results assume that the underlying populations were the same in both periods, which was largely the case. The results for the non-parametric estimates are discussed as to limit the assumptions made.

In one region, Ohangwena, the method produced a higher lower than upper bound for the proportion of the population which moved out of poverty. This implies that over the period, either the underlying population changed or the unexplained component of consumption was generally negatively correlated. Thus, the results for the region will be discarded.

In both the urban regions Erongo and Khomas more than 80% of the population did not experience poverty in either period. Based on the conditional distributions, the probability of remaining out of poverty was over 90% in both those regions. In Kavango, on the other hand, more than 65% of the population experienced poverty in at least one of the periods.

The conditional distributions highlight a few regions where it is clear that poverty is more chronic in nature - Caprivi, Kavango and Oshikoto. This is based on the fact that the bounds between the conditional probability of moving out of poverty and staying in poverty do not cross. In other regions poverty could also be mainly chronic, but the bounds overlap. Importantly, in both Kavango and Caprivi, the underlying populations did not change significantly. More changes were observed in Oshikoto, however. In Omusati, the likelihood of moving out of poverty was the highest in the country and was between 46% and 77%.

Where previous results have noted changes in poverty in Namibia, the results presented here highlight the dynamic nature of poverty. It has been identified that poverty is low in urban areas, but the re-

Table 4.11: Non-Parametric Bounds on the Joint Distribution of Poverty Status by Region

Region	$P_{04}P_{10}$		$N_{04}P_{10}$		$P_{04}N_{10}$		$N_{04}N_{10}$	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Caprivi	37.30	16.30	9.40	30.40	0.70	10.00	52.50	43.20
Erongo	5.40	1.10	1.70	6.00	1.80	7.80	91.20	85.10
Hardap	24.70	11.40	0.90	14.20	13.00	18.50	61.40	56.00
Karas	22.20	7.10	0.40	15.40	10.40	16.00	67.00	61.50
Kavango	53.50	30.20	1.30	24.60	10.50	17.50	34.70	27.70
Khomas	7.30	2.10	3.20	8.40	1.00	6.80	88.50	82.70
Kunene	24.80	7.00	1.70	19.50	4.80	14.90	68.80	58.70
Ohangwena	22.20	6.20	0.60	16.60	15.80	15.30	61.40	61.90
Omaheke	31.10	14.50	0.00	16.60	13.40	17.50	55.60	51.40
Omusati	16.20	4.30	0.70	12.60	14.10	15.10	68.90	67.90
Oshana	14.40	3.60	0.60	11.30	2.00	11.40	83.00	73.60
Oshikoto	36.00	18.00	6.20	24.20	5.70	16.50	52.10	41.20
Otjozondjupa	28.70	12.80	1.60	17.50	6.00	12.90	63.70	56.80

Source: Own calculations using NHIES 2003/04 and 2009/10.
The sample is limited to households where the head is between 25 or older and 55 or younger.
P_{04} and P_{10} refer to "Poor" in 2003/04 and 2009/10 respectively. Similarly N refers to "Not Poor". "Lower" and "Upper" refer to the two bounds on **mobility**. For **immobility** "Lower" is the upper bound and "Upper" is the lower bound.

sults highlight that vulnerability to poverty is also low in urban areas. In contrasts, in rural regions, vulnerability is higher.

In general, Namibia is dealing with both chronic and transitive poverty. With poverty mainly a rural construct, it is interesting to differentiate between rural regions with regards to the dominant nature of poverty. Three regions in particular are highlighted for dealing with mainly chronic poverty - Caprivi, Kavango and Oshikoto. These are also the three regions with the highest poverty rates and contain the largest proportion of the poor in Namibia. Together they contain 50% of the poor in Namibia. The agricultural nature of poor regions suggests that dealing with both chronic and transitive poverty within the regions, requires that attention be paid to smallholder agriculture. Clearly, a large portion of farmers remain unable to escape poverty while others remain vulnerable to it.

Table 4.12: Non-Parametric Bounds on the Distribution of Poverty in 2009/10, Conditional on Poverty in 2003/04 by Region

Region	$P_{10} \mid P_{04}$		$N_{10} \mid P_{04}$		$P_{10} \mid N_{04}$		$N_{10} \mid N_{04}$	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Caprivi	98.16	61.98	1.84	38.02	15.19	41.30	84.81	58.70
Erongo	75.00	12.36	25.00	87.64	1.83	6.59	98.17	93.41
Hardap	65.52	38.13	34.48	61.87	1.44	20.23	98.56	79.77
Karas	68.10	30.74	31.90	69.26	0.59	20.03	99.41	79.97
Kavango	83.59	63.31	16.41	36.69	3.61	47.04	96.39	52.96
Khomas	87.95	23.60	12.05	76.40	3.49	9.22	96.51	90.78
Kunene	83.78	31.96	16.22	68.04	2.41	24.94	97.59	75.06
Ohangwena	58.42	28.84	41.58	71.16	0.97	21.15	99.03	78.85
Omaheke	69.89	45.31	30.11	54.69	0.00	24.41	100.00	75.59
Omusati	53.47	22.16	46.53	77.84	1.01	15.65	98.99	84.35
Oshana	87.80	24.00	12.20	76.00	0.72	13.31	99.28	86.69
Oshikoto	86.33	52.17	13.67	47.83	10.63	37.00	89.37	63.00
Otjozondjupa	82.71	49.81	17.29	50.19	2.45	23.55	97.55	76.45

Source: Own calculations using NHIES 2003/04 and 2009/10.
The sample is limited to households where the head is between 25 or older and 55 or younger.
P_{04} and P_{10} refer to "Poor" in 2003/04 and 2009/10 respectively. Similarly N refers to "Not Poor". "Lower" and "Upper" refer to the two bounds on **mobility**. For **immobility** "Lower" is the upper bound and "Upper" is the lower bound.

4.6 Conclusion

Since independence Namibia has experienced large decreases in poverty, but by the end of the 2000's the headcount rate was still close to 30%. From a policy perspective it would be of value to understand whether poverty is dominantly chronic or transitive and where different types of poverty are concentrated.

For Namibia, however, there is no panel data - traditionally required for such analyses - available. In its absence, drawing conclusions regarding poverty dynamics relies on pseudo-panel analysis. The method proposed by Dang et al. (2014) was suitable for use in Namibia, given the availability of 2 comparable surveys. It generates bounds on the joint distribution of poverty over time. This was used to study poverty dynamics between 2003/04 and 2009/10 - two years with comparable survey data that measured poverty. Generally, the bounds were too wide to draw strong conclusions on whether poverty was largely transitory or chronic in nature. It did provide some valuable insights though.

Poverty is mainly a rural construct and vulnerability to poverty is low in urban areas. The nature of poverty varies across agricultural regions. In the poorest three regions of Namibia - Caprivi, Kavango and Oshikoto - which account for 50% of the poor in the country, poverty was predominantly chronic in nature.

This thesis has highlighted that while the pseudo-panel method has contributed some insights into the dynamics of poverty, it does not provide enough information to inform policy. Panel data is needed; specifically, panel data that surveys households and their agricultural activities. This will provide much needed information to address poverty in Namibia, with the value of such data being highlighted in chapter 2.

Chapter 5

Conclusions

This thesis started by presenting an overview of poverty in sub-Saharan Africa (SSA), pointing out that the poor mainly reside in rural areas and most poor households practise smallholder agriculture. The shortage of good data was also highlighted - there is no doubt that data limitations have hindered poverty alleviation efforts.[1]

Chapter 1 highlighted that there are too few surveys undertaken in the region. There is also a shortage of data that combines household and agricultural level data to inform, especially, on agricultural productivity. Growth in the agricultural sector has been shown to be highly effective at alleviating poverty.

Against this background this thesis set out to highlight ways in which current data can be used to enhance the understanding of welfare in SSA. Three welfare related topics were studied by extending techniques and data to new settings. In chapter 2 econometric techniques were extended from the formal labour market to a setting of smallholder agriculture. These techniques were used to estimate the causal effect of education on agricultural productivity. In chapter 3 it was shown how daytime satellite data can be used to track welfare over time in under-developed agrarian regions - regions where night lights satellite data do not capture economic activity. In chapter 4 repeated cross-sectional surveys were extended to estimate economic mobility in a region where panel data does not exist.

The rest of this chapter, briefly expands on the studies undertaken, their main findings and gives some suggestions for further research. The findings are then collated to highlight messages for broader developmental concerns in sub-Saharan Africa.

[1] As noted, sub-Saharan Africa was the only region that did not halve poverty between 2000 and 2015 which was one of the aims of the Millennium Development Goals.

5.1 Research Findings, and Suggestions for Future Studies

Chapter 2 estimated whether education increases farmers' productivity in Malawi. The chapter used one of the few available datasets that contains household level and agricultural data. It extended the use of econometric techniques generally used in labour market analysis of the formal sector to smallholder agriculture. Instrumental variable techniques were used to estimate the causal effect of education on agricultural productivity. It is, to my knowledge, the first causal estimates in sub-Saharan Africa of the impact of education on agricultural productivity. The causal returns to education in the formal sector of Malawi were also estimated for comparison and are also a first for the country.

The introduction of free primary education (FPE) and the age of paternal orphanhood were used as instruments for education and these IV's were shown to estimate local average treatment effects of education for a group that would not gone to primary school had in not become free and a group that only left school due to paternal orphanhood.

For individuals who only entered schooling due to FPE, returns were low in both agricultural productivity and the formal sector. In comparison, there is evidence that education can have a sizeable effect on agricultural productivity for individuals who only left school due to paternal orphanhood. The returns in the formal sector are larger. It is evident thus, that increased access to education alone would not suffice to improve agricultural productivity for a large proportion of Malawians.

Chapter 3 addressed current shortcomings in household surveys that measure welfare. Surveys measuring welfare are conducted infrequently and are limited to being representative at large geographic areas. Satellites overcome both these limitations and night lights satellite data has been shown to be useful in tracking GDP over time (Henderson et al., 2012). Chapter 3 noted, however, that night lights cannot detect economic activity in under-electrified areas of sub-Saharan Africa, where poor households practise smallholder farming.

In under-electrified agrarian regions, where night lights are not present, chapter 3 suggested that indicators of vegetation quality, derived from daytime satellite data, can be used to track welfare changes over time. This was based on the fact that such indicators can be used to predict crop yields. Previous studies have only focussed on predicting welfare cross-sectionally with daytime satellite data. This study is the first to study relationships between satellite indicators of vegetation quality and welfare longitudinally. Namibia was used as country of study. Regional inequalities in Namibia highlighted why cross-sectional estimates could be misleading. The poor are largely concentrated in rural agrarian regions. Thus, cross-sectionally, higher vegetation quality correlated negatively with welfare. However, within rural agrarian regions, vegetation quality and welfare were shown to be positively associated with one another over time.

Further research should focus on confirming the findings of chapter 3 in other rural agricultural settings and on developing models with high predictive power over time. These models could then be used to track changes in welfare over the short-term, which could be used to identify areas that suffered income shocks. The predictive power of vegetation indicators for crop yield harvests also suggests that this type of model could be used to pre-empt welfare shocks within regions which could allow governments and aid organisations to respond earlier to those shocks.

Chapter 4 showed how existing cross-sectional data can be utilised to gain insights into poverty dynamics, in a country where panel data does not exist - Namibia. This chapter presented, for the first time, insights into poverty dynamics in that country. Namibia has experienced large decreases in headcount poverty since gaining independence, but headcount poverty remains high, especially in rural areas. In Namibia, understanding the dynamic nature of poverty can potentially be useful for poverty alleviation efforts.

As in much of sub-Saharan Africa, there are only two comparable surveys in Namibia that measure welfare. Most pseudo-panel methods, usually used in the absence of panel data, require at least three surveys for estimation of dynamic models. However, the method by Dang et al. (2014) was suitable for use in Namibia, as it requires only two cross-sections. The method estimates bounds on the proportion of the population observed moving into or out of, and staying in or out poverty over two time periods. Bounds were also adjusted to represent distributions conditional on first period poverty status as suggested by Fields and Viollaz (2013).

Results showed that both chronic and transitive poverty are prevalent in Namibia and are mainly present in regions where smallholder agriculture on communal lands is prevalent. Across these regions, however, the dominant nature of poverty varied. Chronic poverty was shown to be particularly dominant in the three regions with the highest poverty rates and which contain the largest share of poor households in Namibia.

5.2 General Findings for Sub-Saharan Africa

By considering the results of the three chapters, three general points on welfare and its study in sub-Saharan Africa can be made. Firstly, the context of smallholder agriculture matters for policy and analysis. Chapter 2 showed that education does not have the desired returns in smallholder agriculture that it has in the non-agricultural formal sector. Chapter 3 showed that by acknowledging that most rural households practice agriculture, indicators of vegetation quality could be used to track welfare.

The agricultural context also matters for the data that is collected. The value of having data that linked household level data with agricultural level data made the study in chapter 2 possible. On the other hand, the point was made in chapter 4 that such data would be needed in Namibia to create effective poverty alleviation policies, as both chronically and transitive poor households were mainly located in regions practising smallholder agriculture.

Surveys capturing both household and agricultural data, such as those by LSMS-ISA are of great value in Sub-Saharan Africa. Recent follow-up surveys also imply that panel data is now available to study the relationship between welfare and agriculture. Yet, only eight countries are currently conducting such surveys. As noted in the chapter 1, Carletto et al. (2015) cited weak demand as a reason for a shortage of data. Greater use of the surveys can highlight their value, which could convince other countries to participate. Indeed, at the time of publishing, Carletto et al. (2010) only noted six countries participating in LSMA-ISA. Thus, two countries since then have become convinced that such surveys can add value.

Secondly, for poverty alleviation, agricultural productivity needs to be increased. Chapter 2 highlighted that education is on its not own not effective for a large proportion of farmers in Malawi. An alternative program that has been suggested is input subsidy programs (ISPs). ISPs work by supplying farmers with better technologies and are used in a number of countries in SSA. Malawi's Food Input Subsidy program is particularly well known and has been shown to have increased agricultural productivity (Dorward and Chirwa, 2011). Jayne and Rashid (2013) note that in 2011, ten countries in SSA, spent roughly 28.6% of their public expenditure for agriculture on ISPs. However, the overall evidence from these countries suggested that the costs outweigh the benefits. Arndt et al. (2016) dispute this, however, by showing that in Malawi, once spillover effects from the ISP are considered benefits are higher than costs.

Better survey data could potentially highlight effective ways to improve ISPs efficiency and suggest alternative methods to increase productivity. Satellite data could also be of assistance here. Vegetation indices could for example be used to identify agricultural areas with particularly low productivity and thus improve efficiency of ISPs by targeting areas with the largest marginal returns to, for example, fertilizer.

Improving the productivity of female farmers is also important as they are less productive than their male counterparts. As discussed in Chapter 2, both less productive use of resources and worse access to resources (compared to male farmers) are reasons for this (Kilic et al., 2015; Oseni et al., 2015). The Food and Agricultural Organisation (FAO) report that the gender-gap in agricultural productivity in sub-Saharan Africa is between 20% and 30%. The FAO estimates that by improving female farmers' productivity to the levels of males could potentially increase agricultural output by 2.5%-4% and could decrease the number of people living in hunger by 100 million (FAO, 2011).

The third point to be highlighted is that agricultural households are vulnerable to climate change. Chapter 4 found that rural households were more vulnerable to poverty and chapter 3 showed how changes in vegetation quality were positively associated with changes in welfare. Forecasts suggest that climate change will have significant impacts on crop yields in SSA (see for example Challinor et al. (2007) and Schlenker and Lobell (2010)). The ability of satellite data to identify agricultural areas, and to track vegetation quality can be valuable to identify and support vulnerable farmers.

Challinor et al. (2007) note that access to information and knowledge are central to farmers' ability to adapt to climate change. Even though chapter 2 found that education did not increase productivity for farmers with low expected returns to such education, education could still be key in addressing their vulnerability to climate change. As was discussed in chapter 2, Ram (1980) argued that the marginal benefits of new information could increase with education. Thus, in rapidly changing environments, the benefits of education in smallholder agriculture could become more significant.

The vulnerability of farming households, particularly to climate fluctuations and climate change, remains a key future challenge in addressing poverty in sub-Saharan Africa. The generation of new insights and knowledge on these households is needed to address this vulnerability. This thesis has highlighted pathways to generate new insights even without the ideal data source. Methods can be extended from other economic fields and existing data sources can be extended to provide information on topics that they were not designed for.

Methods highlighted in this thesis, combined with continually improving survey data and econometric techniques, suggest that Africa's "statistical tragedy" (Devarajan, 2013) could give way for a (much needed) statistical revolution.

Bibliography

Ainsworth, M., Beegle, K., and Koda, G. (2005). The Impact of Adult Mortality and Parental Deaths on Primary Schooling in North-Western Tanzania. *Journal of Development Studies*, 41(3):412–439.

Alam, S. A. (2015). Parental health shocks, child labor and educational outcomes: Evidence from Tanzania. *Journal of Health Economics*, 44:161 – 175.

Alene, A. and Manyong, V. (2007). The Effects of Education on Agricultural Productivity under Traditional and Improved Technology in Northern Nigeria: an Endogenous Switching Regression Analysis. *Empirical Economics*, 32(1):141–159.

Angrist, J. D. and Imbens, G. W. (1995). Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association*, 90(430):431–442.

Anthony, M. (2015). Night lights and regional income inequality in Africa. Working Paper Series UNU-WIDER Research Paper, World Institute for Development Economic Research (UNU-WIDER). Available at `https://ideas.repec.org/p/unu/wpaper/wp2015-085.html`.

Antman, F. and McKenzie, D. J. (2007). Earnings Mobility and Measurement Error: A Pseudo-Panel Approach. *Economic Development and Cultural Change*, 56:125–161.

Appleton, S. and Balihuta, A. (1996). Education and Agricultural Productivity: Evidence from Uganda. *Journal of International Development*, 8(3):415–444.

Arndt, C., Pauw, K., and Thurlow, J. (2016). The economy-wide impacts and risks of malawi's farm input subsidy program. *American Journal of Agricultural Economics*, 98(3):962 – 980.

Asadullah, M. N. and Rahman, S. (2009). Farm Productivity and Efficiency in Rural Bangladesh: The Role of Education Revisited. *Applied Economics*, 41(1):17–33.

Barrett, C. B. (2005). Rural poverty dynamics: development policy implications. *Agricultural Economics*, 32(s1):45–60.

Baulch, B. and Hoddinott, J. (2000). Economic mobility and poverty dynamics in developing countries. *The Journal of Development Studies*, 36(6):1–24.

Beegle, K., Christiaensen, L., Dabalen, A., and Gaddis, I. (2016). *Poverty in a Rising Africa.* World Bank, Washington, DC. Available at https://openknowledge.worldbank.org/handle/10986/22575.

Bhattacharya, H. and Innes, R. (2012). Income and the Environment in Rural India: Is There a Poverty Trap? *American Journal of Agricultural Economics*, 95(1):42–69.

Bierbaum, M. and Gassmann, F. (2012). Chronic and Transitory Poverty in the Kyrgyz Republic: What can synthetic panels tell us? UNU-MERIT Working Paper Series 2012-064, UNU-MERIT, Maastricht.

Blumenstock, J., Cadamuro, G., and On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264):1073–1076.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American Statistical Association*, 90(430):pp. 443–450.

Bourguignon, F., Goh, C.-c., and Kim, D. I. (2004). Estimating individual vulnerability to poverty with pseudo-panel data. Policy Research Working Paper Series 3375, The World Bank.

Bryceson, D. F. (2006). Ganyu Casual Labour, Famine and HIV/AIDS in Rural Malawi: Causality and Casualty. *The Journal of Modern African Studies*, 44:173–202.

Burger, R. P., Klasen, S., and Zoch, A. (2016). Estimating income mobility when income is measured with error: the case of South Africa. REDI3x3 Working paper 14, REDI3x3.

Cao, Z., Wu, Z., Kuang, Y., Huang, N., and Wang, M. (2016). Coupling an Intercalibration of Radiance-Calibrated Nighttime Light Images and Land Use/Cover Data for Modeling and Analyzing the Distribution of GDP in Guangdong, China. *Sustainability*, 8(2):108.

Carletto, C., Jolliffe, D., and Banerjee, R. (2015). From Tragedy to Renaissance: Improving Agricultural Data for Better Policies. *The Journal of Development Studies*, 51(2):133–148.

Carletto, G., Beegle, K., Himelein, K., Kilic, T., Murray, S., Oseni, M., Scott, K., and Steele, D. (2010). Improving the Availability, Quality and Policy-Relevance of Agricultural Data: The Living Standards Measurement Study – Integrated Surveys on Agriculture. Paper presented at the Third Wye City Group Global Conderence on Agricutlural and Rural Households Statistics, Washington D.C. Avaialble at http://www.fao.org/fileadmin/templates/ess/pages/rural/wye_city_group/2010/May/WYE_2010.2.1_Carletto.pdf.

Carr-Hill, R. (2014). Measuring development progress in Africa: the denominator problem. *Canadian Journal of Development Studies*, 35(1):136–154.

Carter, M. R. and May, J. (2001). One Kind of Freedom: Poverty Dynamics in Post-apartheid South Africa. *World Development*, 29(12):1987 – 2006.

Case, A., Paxson, C., and Ableidinger, J. (2004). Orphans in Africa: Parental Death, Poverty, and School Enrollment. *Demography*, 41(3):483–508.

CBS (2006). 2003 / 2004 Namibia Household Income & Expenditure Survey: Main Report. Central Bureau of Statistics, Windhoek, Namibia.

CBS (2008). A review of Poverty and Inequality in Namibia. Central Bureau of Statiscts and National Planning Commission, Windhoek, Namibia.

Challinor, A., Wheeler, T., Garforth, C., Craufurd, P., and Kassam, A. (2007). Assessing the vulnerability of food crop systems in africa to climate change. *Climatic Change*, 83(3):381–399.

Chen, X. and Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594.

Chirwa, E. (2005). Adoption of Fertiliser and Hybrid Seeds by Smallholder Maize Farmers in Southern Malawi. *Development Southern Africa*, 22(1):1–12.

Chirwa, E. W., Kumwenda, I., Jumbe, C., Chilonda, P., and Minde, I. (2008). Agricultural Growth and Poverty Reduction in Malawi: Past performance and Recent Trends. ReSAKSS Working Paper 8, Regional Strategic Analysis and Knowledge Support System for Southern Africa, Pretoria, South Africa.

Chirwa, E. W. and Matita, M. M. (2009). The Rate of Return on Education in Malawi. Working Paper 2009/01, University of Malawi, Chancellor College, Department of Economics, Zomba.

Chirwa, E. W. and Zgovu, E. K. (2001). Does the Return to Schooling Depend on the Type of Employment? Evidence From the Rural Labour Market in Malawi. Paper presented at a conference on "understanding poverty and growth in sub-saharan africa" held at the centre for the study of african economies, held on 18-19 march, 2002, University of Oxford. Available at `http://www.csae.ox.ac.uk/conferences/2002-upagissa/papers/Chirwa-Zgovu-case2002.PDF`.

Christiaensen, L., Demery, L., and Kuhl, J. (2011). The (evolving) role of agriculture in poverty reduction—An empirical perspective. *Journal of Development Economics*, 96(2):239–254.

Collado, M. D. (1997). Estimating dynamic models from time series of independent cross-sections. *Journal of Econometrics*, 82(1):37 – 62.

Cruces, G., Lanjouw, P., Lucchetti, L., Perova, E., Vakis, R., and Viollaz, M. (2015). Estimating poverty transitions using repeated cross-sections: a three-country validation exercise. *The Journal of Economic Inequality*, 13(2):161–179.

Cuesta, J., Nopo, H., and Pizzolitto, G. (2011). Using Pseudo-Panels to Measure Income Mobility in Latin America. *Review of Income and Wealth*, 57(2):224–246.

Dang, H.-A., Lanjouw, P., Luoto, J., and McKenzie, D. (2014). Using Repeated Cross-Sections to Explore Movements into and Out of Poverty. *Journal of Development Economics*, 107:112 – 128.

Davis, B., Giuseppe, S. D., and Zezza, A. (2017). Are African households (not) leaving agriculture? Patterns of households' income sources in rural Sub-Saharan Africa. *Food Policy*, 67:153 – 174. Agriculture in Africa – Telling Myths from Facts.

Davis, K., Nkonya, E., Kato, E., Mekonnen, D. A., Odendo, M., Miiro, R., and Nkuba, J. (2010). Impact of Farmer Field Schools on Agricultural Productivity and Poverty in East Africa. IFPRI Discussion Paper 00992, International Food Policy Research Institute.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30(1-2):109–126.

Deaton, A. and Paxson, C. (1994). Intertemporal Choice and Inequality. *Journal of Political Economy*, 102(3):437–467.

Deininger, K., Carletto, C., Savastano, S., and Muwonge, J. (2012). Can diaries help in improving agricultural production statistics? Evidence from Uganda. *Journal of Development Economics*, 98(1):42 – 50. Symposium on Measurement and Survey Design.

Desiere, S., Staelens, L., and D'Haese, M. (2016). When the Data Source Writes the Conclusion: Evaluating Agricultural Policies. *The Journal of Development Studies*, 52(9):1372–1387.

Devarajan, S. (2013). Africa's Statistical Tragedy. *Review of Income & Wealth*, 59:S9 – S15.

Devereux, S. and Naeraa, T. (1996). Drought and Survival in Rural Namibia. *Journal of Southern African Studies*, 22(3):421–440.

Dorward, A. and Chirwa, E. (2011). The Malawi Agricultural Input Subsidy Programme: 2005/06 to 2008/09. *International Journal of Agricultural Sustainability*, 9(1):232–247.

Dragoset, L. M. and Fields, G. S. (2006). U.S. Earnings Mobility: Comparing Survey-Based and Administrative-Based Estimates. ECINEQ Working Paper 2006-55, Society for the Study of Economic Inequality.

Easterly, W. (2009). How the Millennium Development Goals are Unfair to Africa. *World Development*, 37(1):26–35.

Ebener, S., Murray, C., Tandon, A., and Elvidge, C. C. (2005). From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. *International Journal of Health Geographics*, 4:5 – 17.

Elbers, C., Lanjouw, J. O., and Lanjouw, P. (2003). Micro-level Estimation of Poverty and Inequality. *Econometrica*, 71(1):355 – 364.

Elvidge, C., Hsu, F.-C., Baugh, K., and Ghosh, T. (2014). National Trends in Satellite-Observed Lighting: 1992–2012. In *Remote Sensing Applications Series*, pages 97–120. CRC Press.

Elvidge, C. D., Baugh, K. E., Anderson, S. J., Sutton, P. C., and Ghosh, T. (2012). The Night Light Development Index (NLDI): a spatially explicit measure of human development from satellite data. *Social Geography*, 7(1):23–35.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., Davis, E. R., and Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, 18(6):1373–1379.

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., and Bright, E. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, 35(8):1652 – 1660.

Evans, D. K. and Miguel, E. (2007). Orphans and Schooling in Africa: A Longitudinal Analysis. *Demography*, 44(1):35–57.

Fafchamps, M. and Quisumbing, A. R. (1999). Human Capital, Productivity, and Labor Allocation in Rural Pakistan. *Journal of Human Resources*, 34(2):369–406.

FAO (2010). Crop calendar - An information tool for seed security. Online. Available at `http://www.fao.org/agriculture/seed/cropcalendar/cropcalendar.do[Accessedon20/07/2017]`.

FAO (2011). The state of food and agriculture. women in agriculture: Closing the gap for development. Food and Agricultural Organisation. Rome, Italy. Avaiable at `http://www.fao.org/docrep/013/i2050e/i2050e.pdf`.

Fara, K. (2001). How Natural Are 'Natural Disasters'? Vulnerability to Drought of Communal Farmers in Southern Namibia. *Risk Management*, 3(3):47–63.

Fields, G. and Viollaz, M. (2013). Can the Limitations of Panel Datasets be Overcome by using Pseudo-Panel to Estimate Income Mobility. Fifth meeting of the Society for the Study of Economic Inequality (ECINEQ), held at the University of Bari, Italy, July 22-24, 2013. Available at `http://www.ecineq.org/ecineq_bari13/filesxbari13/cr2/p90.pdf`.

Fields, G. S. (2005). The Many Facets of Economic Mobility. Retrieved 21-07-2015, from Cornell University, ILR School site: `http://digitalcommons.ilr.cornell.edu/articles/230/`.

Fields, G. S., Cichello, P. L., Freije, S., Menéndez, M., and Newhouse, D. (2003). For Richer or for Poorer? Evidence from Indonesia, South Africa, Spain, and Venezuela. *The Journal of Economic Inequality*, 1(1):67–99.

Fields, G. S. and Ok, E. A. (1996). The Meaning and Measurement of Income Mobility. *Journal of Economic Theory*, 71(2):349 – 377.

Filmer, D. and Pritchett, L. H. (2001). Estimating Wealth Effects without Expenditure Data-Or Tears: An Application to Educational Enrollments in States of India. *Demography*, 38(1):115 – 132.

Foster, A. D. and Rosenzweig, M. R. (1995). Learning by Doing and Learning from Others: Human capital and Technical Change in Agriculture. *Journal of Political Economy*, 103(6):1176–1209.

Foster, A. D. and Rosenzweig, M. R. (2003). Economic Growth and the Rise of Forests. *The Quarterly Journal of Economics*, 118(2):601–637.

Foster, A. D. and Rosenzweig, M. R. (2004). Technological change and the distribution of schooling: evidence from green-revolution India. *Journal of Development Economics*, 74(1):87 – 111. New Research on Education in Developing Economies.

Frayne, B. (2004). Migration and urban survival strategies in Windhoek, Namibia. *Geoforum*, 35(4):489 – 505.

Fuglie, K. O. (2008). Is a Slowdown in Agricultural Productivity Growth Contributing to the Rise in Commodity Prices? *Agricultural Economics*, 39(supplement):431–441.

Ganguly, S., Friedl, M. A., Tan, B., Zhang, X., and Verma, M. (2010). Land surface phenology from MODIS: Characterization of the Collection 5 global land cover dynamics product. *Remote Sensing of Environment*, 114(8):1805 – 1816.

Girma, S. (2000). A quasi-differencing approach to dynamic modelling from a time series of independent cross-sections. *Journal of Econometrics*, 98(2):365 – 383.

Gottschalk, P. and Huynh, M. (2010). Are earnings inequality and mobility overstated? The impact of non-classical measurement error. *Review of Economics and Statistics*, 92:302–315.

Haughton, J. and Khandker, S. R. (2009). *Handbook on Poverty and Inequality*. Number 11985 in World Bank Publications. The World Bank.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153–161.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, 102(2):994–1028.

Huang, J. and Han, D. (2014). Meta-analysis of influential factors on crop yield estimation by remote sensing. *International Journal of Remote Sensing*, 35(6):2267–2295.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., and Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1):195 – 213. The Moderate Resolution Imaging Spectroradiometer (MODIS): a new generation of Land Surface Monitoring.

Huffman, W. E. (1980). Farm and Off-Farm Work Decisions: The Role of Human Capital. *The Review of Economics and Statistics*, 62(1):14–23.

Imran, M., Stein, A., and Zurita-Milla, R. (2014). Investigating rural poverty and marginality in Burkina Faso using remote sensing-based products. *International Journal of Applied Earth Observation and Geoinformation*, 26:322 – 334.

Irz, X., Lin, L., Thirtle, C., and Wiggins, S. (2001). Agricultural Productivity Growth and Poverty Alleviation. *Development Policy Review*, 19(4):449–466.

Jantti, M. and Jenkins, S. P. (2013). Income Mobility. IZA Discussion Paper Series 7730, IZA, Bonn.

Jayne, T. and Rashid, S. (2013). Input subsidy programs in sub-Saharan Africa: a synthesis of recent evidence. *Agricultural Economics*, 44(6):547–562.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794.

Jerven, M. (2016). Research Note: Africa by numbers: Reviewing the database approach to studying African economies. *African Affairs*, 115(459):342–358.

Kadzamira, E. and Rose, P. (2001). Educational Policy Choice and Policy Practice in Malawi: Dilemmas and Disjunctures. IDS Working Paper 124, Institute of Development Studies, Brighton. Available at `https://www.ids.ac.uk/files/Wp124.pdf`.

Kadzamira, E. and Rose, P. (2003). Can Free Primary Education Meet the Needs of the Poor? Evidence from Malawi. *International Journal of Educational Development*, 23(5):501–516.

Kassie, G. T., Abate, T., Langyintuo, A., and Maleni, D. (2014). Poverty in Maize Growing Rural Communities of Southern Africa. *Development Studies Research*, 1(1):311–323.

Kilic, T., Palacios-López, A., and Goldstein, M. (2015). Caught in a productivity trap: A distributional perspective on gender differences in malawian agriculture. *World Development*, 70:416–463.

Klemens, B., Coppola, A., and Shron, M. (2015). Estimating local poverty measures using satellite images : a pilot application to Central America. Policy Research Working Paper Series 7329, The World Bank.

Larson, D. F., Otsuka, K., Matsumoto, T., and Kilic, T. (2014). Should African Rural Development Strategies Depend on Smallholder Farms? An Exploration of the Inverse-Productivity Hypothesis. *Agricultural Economics*, 45(3):355–367.

Lechtenfeld, T. and Zoch, A. (2014). Income Convergence in South Africa: Fact or Measurement Error? Stellenbosch Economic Working Papers 10/14, Department of Economics and the Bureau for Economic Research at the University of Stellenbosch, Stellenbosch.

Levine, S. (2007). Trends in Human Development and Human Poverty in Namibia. Background paper to the namibia human development report., United Nations Development Program.

Levine, S. and Roberts, B. (2013). Robust Estimates of Changes in Poverty and Inequality in Post-Independence Namibia. *South African Journal of Economics*, 81(2):167 – 191.

Levine, S., van der Berg, S., and Yu, D. (2011). The Impact of Cash Transfers on Household Welfare in Namibia. *Development Southern Africa*, 28(1):39 – 59.

Lobell, D. B. (2013). The use of satellite data for crop yield gap analysis. *Field Crops Research*, 143:56 – 64.

Lockheed, M. E., Jamison, D. T., and Lau, L. J. (1980). Farmer Education and Farm Efficiency: A Survey. *Economic Development and Cultural Change*, 29(1):37–76.

Lockheed, M. E., Jamison, D. T., and Lau, L. J. (1987). Farmer Education and Farm Efficiency: Reply. *Economic Development and Cultural Change*, 35(3):643–644.

Lucas, A. M. and Mbiti, I. M. (2012). Access, Sorting, and Achievement: The Short-Run Effects of Free Primary Education in Kenya. *American Economic Journal: Applied Economics*, 4(4):226–53.

Lunduka, R., Ricker-Gilbert, J., and Fisher, M. (2013). What are the Farm-Level Impacts of Malawi's Farm Input Subsidy Program? A Critical Review. *Agricultural Economics*, 44(6):563–579.

Matita, M. M. and Chirwa, E. W. (2009). The Impact of Education on Self-Employment, Farm Activities and Household Incomes in Malawi. Working Paper 2009/02, University of Malawi, Chancellor College, Department of Economics, Zomba.

McKay, A. and Lawson, D. (2003). Assessing the Extent and Nature of Chronic Poverty in Low Income Countries: Issues and Evidence. *World Development*, 31(3):425 – 439. Chronic Poverty and Development Policy.

McKenzie, D. J. (2004). Asymptotic theory for heterogeneous dynamic pseudo-panels. *Journal of Econometrics*, 120(2):235 – 262.

Mingat, A., Tan, J., and Hoque, M. (1985). Recovering the Cost of Public Higher Education in LDCs: To What Extent are Loan Schemes an Efficient Instrument? Education and Training Department Discussion Series EDT14, The World Bank.

Mkhabela, M., Bullock, P., Raj, S., Wang, S., and Yang, Y. (2011). Crop yield forecasting on the Canadian Prairies using {MODIS} {NDVI} data. *Agricultural and Forest Meteorology*, 151(3):385 – 393.

Mkhabela, M. S., Mkhabela, M. S., and Mashinini, N. N. (2005). Early maize yield forecasting in the four agro-ecological regions of Swaziland using {NDVI} data derived from NOAA's-AVHRR. *Agricultural and Forest Meteorology*, 129(1–2):1 – 9.

Moffitt, R. (1993). Identification and Estimation of Dynamic Models with a Time Series of Repeated Cross-Sections. *Journal of Econometrics*, 59(1-2):99 – 123.

Morikawa, R. (2014). Remote Sensing Tools for Evaluating Poverty Alleviation Projects: A Case Study in Tanzania. *Procedia Engineering*, 78:178 – 187. Humanitarian Technology: Science, Systems and Global Impact 2014, HumTech2014.

Myneni, R. and Hall, F. (1995). The interpretation of spectral vegetation indexes. *Geoscience and Remote Sensing, IEEE Transactions on*, 33(2):481–486.

NASA LP DAAC (2012a). MOD12Q2 Land Cover Dynamics – Known Issues. Available at `http://www.bu.edu/lcsc/files/2012/08/MCD12Q2_Known_Issues.pdf`.

NASA LP DAAC (2012b). User Guide for the MODIS Land Cover Dynamics Product (MCD12Q2). Available at `http://www.bu.edu/lcsc/files/2012/08/MCD12Q2_UserGuide.pdf`.

NASA LP DAAC (2017a). MODIS Terra Vegetation Indices 16-Day L3 Global 500m SIN Grid. Version 5. (MOD13A1). NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov). Accessed on 23 May 2017 at `https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod13a1`.

NASA LP DAAC (2017b). V005 MODIS Land Cover Dynamics (MCD12Q2).

National Geophysical Data Center (2016). Version 4 DMSP-OLS Nighttime Lights Time Series. Available at `https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html`.

NPC (2015). Namibia Poverty Mapping. Republic of Namibia, National Planning Commission, Windhoek, Namibia.

NSA (2012). Poverty Dynamics in Namibia: A comparative study using the 1993/94, 2003/04 and the 2009/10 NHIES surveys. Namibia Statistics Agency, Windhoek, Namibia.

NSA (2013). Namibia - National Household Income and Expenditure Survey 2009-2010: Metadata. Namibia Statistics Agency, Windhoek, Namibia.

NSA (2015a). Namibia 2011 Census Migration Report. Namibia Statistics Agency, Windhoek, Namibia.

NSA (2015b). Namibia Census of Agriculture 2013/2014: Communal Sector Report. Namibia Statistics Agency, Windhoek, Namibia.

NSO (2012a). Integrated Household Survey 2010-2011 - Household Socio-Economic Characteristics Report. National Statistical Office, Zomba, Malawi.

NSO (2012b). Malawi Third Integrated Household Survey (IHS3) 2010-2011: Basic Information Document. National Statistical Office, Zomba, Malawi.

Odendaal, W. (2011). Land Grabbing in Namibia: A case study from the Omusati region, northern Namibia. Paper presented at the international conference on global land grabbing. 6-8 april 2011, Land Deals Politics Initiative, University of Sussex.

Orr, A. (2000). 'Green Gold'?: Burley Tobacco, Smallholder Agriculture, and Poverty Alleviation in Malawi. *World Development*, 28(2):347–363.

Oseni, G., Corral, P., Goldstein, M., and Winters, P. (2015). Explaining gender differentials in agricultural production in nigeria. *Agricultural Economics*, 46(3):285–310.

Pfaff, A. S. (1999). What Drives Deforestation in the Brazilian Amazon? *Journal of Environmental Economics and Management*, 37(1):26 – 43.

Phillips, J. M. (1987). A Comment on Farmer Education and Farm Efficiency: A Survey. *Economic Development and Cultural Change*, 35(3):637–641.

Phillips, J. M. (1994). Farmer education and farmer efficiency: A meta-analysis. *Economic Development and Cultural Change*, 43(1):149–165.

Pinkovskiy, M. and Sala-i-Martin, X. (2016). Lights, Camera ... Income! Illuminating the National Accounts-Household Surveys Debate. *The Quarterly Journal of Economics*, 131(2):579–631.

Pritchett, L. (2001). Where Has All the Education Gone? *World Bank Economic Review*, 15(3):367–391.

Ram, R. (1980). Role of Education in Production: A Slightly New Approach. *The Quarterly Journal of Economics*, 95(2):365–373.

Ram, R. and Singh, R. D. (1988). Farm households in rural Burkina Faso: Some evidence on allocative and direct return to schooling, and male-female labor productivity differentials. *World Development*, 16(3):419 – 424.

Ravallion, M. and Chen, S. (2007). China's (Uneven) Progress Against Poverty. *Journal of Development Economics*, 82:1–42.

Reserve Bank of Malawi (2014). Financial and Economic Review. Volume 46 2, Lilongwe.

Rigobon, R. and Stoker, T. M. (2005). Instrumental Variable Bias with Censored Regressors. Mit working paper. Available at `http://web.mit.edu/tstoker/www/Rigobon\_Stoker\_IV\_March\_05.pdf`.

Sandefur, J. and Glassman, A. (2015). The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics. *The Journal of Development Studies*, 51(2):116–132.

Schlenker, W. and Lobell, D. B. (2010). Robust negative impacts of climate change on african agriculture. *Environmental Research Letters*, 5(1):014010.

Schultz, T. W. (1975). The Value of the Ability to Deal wit Disequilibria. *Journal of Economic Literature*, 13(3):827–846.

Skole, D. and Tucker, C. (1993). Tropical Deforestation and Habitat Fragmentation in the Amazon: Satellite Data from 1978 to 1988. *Science*, 260(5116):1905–1910.

Small, C., Elvidge, C. D., Balk, D., and Montgomery, M. (2011). Spatial scaling of stable night lights. *Remote Sensing of Environment*, 115(2):269 – 280.

Staiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586.

Sutton, P. C. and Costanza, R. (2002). Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation. *Ecological Economics*, 41(3):509–527.

Sutton, P. C., Elvidge, C. D., and Ghosh, T. (2007). Estimation of Gross Domestic Product at Sub-national Scales Using Nighttime Satellite Imagery. *International Journal of Ecological Economics and Statistics*, 8:5 – 21.

Taylor, S. and Spaull, N. (2013). The effects of rapidly expanding primary school access on effective learning: The case of Southern and Eastern Africa since 2000. Stellenbosch Economics Working Papers 1/13, Stellenbosch.

USGS (2015). Landsat - Earth Observation Satellites. Fact Sheet 2015-3081, U.S Geological Survey.

Uwaifo Oyelere, R. (2010). Africa's Education Enigma? The Nigerian Story. *Journal of Development Economics*, 91(1):128 – 139.

Valente, C. (2015). Primary Education Expansion and Quality of Schooling: Evidence from Tanzania. Discussion Paper 9208, IZA, Bonn, Germany. Available at `http://ftp.iza.org/dp9208.pdf`.

Verbeek, M. and Vella, F. (2005). Estimating Dynamic Models from Repeated Cross-Sections. *Journal of Econometrics*, 127(1):83 – 102.

Villa, J. M. (2016). Social Transfers and Growth: Evidence from Luminosity Data. *Economic Development and Cultural Change*, 65(1):39–61.

Watmough, G. R., Atkinson, P. M., and Hutton, C. W. (2013). Predicting socioeconomic conditions from satellite sensor data in rural developing countries: A case study using female literacy in Assam, India. *Applied Geography*, 44:192 – 200.

Watmough, G. R., Atkinson, P. M., Saikia, A., and Hutton, C. W. (2016). Understanding the Evidence Base for Poverty Environment Relationships using Remotely Sensed Satellite Data: An Example from Assam, India. *World Development*, 78:188 – 203.

Weeks, J. R., Getis, A., Stow, D. A., Hill, A. G., Rain, D., Engstrom, R., Stoler, J., Lippitt, C., Jankowska, M., Lopez-Carr, A. C., Coulter, L., and Ofiesh, C. (2012). Connecting the Dots Between Health, Poverty and Place in Accra, Ghana. *Annals of the Association of American Geographers*, 102(5):932–941.

Weir, S. (1999). The Effects of Education on Farmer Productivity in Rural Ethiopia. Working Paper Series 1999-7, Centre for the Study of African Economies, University of Oxford.

Weir, S. and Knight, J. (2000). Adoption and Diffusion of Agricultural Innovations in Ethiopia: The Role of Education. Working Paper Series 2000-4, Centre for the Study of African Economies, University of Oxford.

Welch, F. (1970). Education in Production. *Journal of Political Economy*, 78(1):35–59.

Werner, W. (1993). A brief history of land disposession in Namibia. *Journal of Southern African Studies*, 19(1):135.

Whiteside, M. (2000). Ganyu labour in Malawi and its Implications for Livelihood Security Interventions - An Analysis of Recent Literature and Implications for Poverty Alleviation. Agricultural Research and Extension Network Paper No. 99, Overseas Development Institute, London.

World Bank (2008). Republic of Namibia - Addressing Binding Constraints to Stimulate Broad Based Growth : A Country Economic Report. World Bank Other Operational Studies 12601, The World Bank.

World Bank (2015a). Malawi — Data.

World Bank (2015b). World Development indicators. Available at `http://databank.worldbank.org/data/views/variableSelection/selectvariables.aspx?source=world-development-indicators`.

World Bank (2017a). Data on statistical capacity. Available at `http://datatopics.worldbank.org/statisticalcapacity/`.

World Bank (2017b). Statistical capacity indicators. Available at `http://databank.worldbank.org/data/reports.aspx?source=statistical-capacity-indicators`.

Zant, W. (2012). The Economics of Food Aid under Subsistence Farming with an Application to Malawi. *Food Policy*, 37(1):124–141.

Zeller, M., Diagne, A., and Mataya, C. (1998). Market Access by Smallholder Farmers in Malawi: Implications for Technology Adoption, Agricultural Productivity and Crop Income. *Agricultural Economics*, 19(1-2):219–229.

Zhang, X., Friedl, M. A., and Schaaf, C. B. (2006). Global vegetation phenology from Moderate Resolution Imaging Spectroradiometer (MODIS): Evaluation of global patterns and comparison with in situ measurements. *Journal of Geophysical Research: Biogeosciences*, 111(G4).

Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., Reed, B. C., and Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84(3):471 – 475.

Appendix A

CGA Report on Cropland Classification in Erongo and Kavango, Namibia

This appendix presents a report by the Centre for Geographic Analysis (CGA) at Stellenbosch University on the process they used, to classify cropland in two regions in Namibia - Erongo and Kavango. The report follows on the next page.

CGA
CENTRE FOR GEOGRAPHICAL ANALYSIS

11 May 2016

Thomas Ferreira
Department of Economics
Stellenbosch University

Classification of agriculture/non-agriculture in Erongo and Kavango regions in Namibia

At the behest Thomas Ferreira from the Department of Economics in Stellenbosch University, the Centre for Geographical Analysis (CGA) undertook to classify agriculture and non-agriculture areas in the Kavango and Erongo regions of Namibia for the years of 1993, 2003 and 2009. Landsat 5 and 7 imagery was used for the classification in the Kavango region. The classification in the Erongo region was initially derived from Google Earth, with Landsat 5 and 7 imagery being used to detect changes over time. This document elaborates on the methodology used to perform the classification and discusses the quality and limitations of the resulting product.

1. Landsat imagery

Landsat images are available freely from the website of the United States Geological Survey (USGS). The Landsat 5 Thematic Mapper (TM) sensor has six bands in the visible and near-infrared (VNIR) spectrum with a spatial resolution of 30 metres, and a seventh thermal infrared band with a spatial resolution of 120 metres. Landsat 7's Enhanced Thematic Mapper Plus (ETM+) sensor has the same six VNIR bands with 30 metre spatial resolution, but the thermal band has a spatial resolution of 60 metres. Furthermore it adds a panchromatic band with a spatial resolution of 15 metres. This allows the VNIR bands to be resampled to 15 metre resolution through pan-sharpening.

Landsat 7 was only launched in 1999, and its scan-line corrector failed in mid-2003, rendering large areas in each scene unusable. Hence suitable Landsat 7 imagery was only available for the 2003 classification, with the 1993 and 2009 classifications being based on Landsat 5.

2. Data Pre-processing

2.1 Geometric Correction

Geometric correction of satellite imagery is necessary to mitigate locational errors introduced by sensor geometry and terrain relief. Landsat imagery is usually supplied in terrain corrected form (level 1 terrain corrected or L1T), but two of the 2009 Landsat 5 images used for the project were only available at the lower systematic correction processing level (level 1G / L1G). Upon inspection the locational accuracy of these two scenes was deemed insufficient for analysis, and further geometric correction was deemed necessary. This correction was performed in PCI OrthoEngine using Rational Functions correction based on Ground Control Points (GCP's). GCP's are points with known locations that can be clearly identified on the image to be corrected. In this case the coordinates of these features were obtained from the already terrain corrected Landsat 7 images of the same scene, while the elevation values for those locations were extracted from the 90 metre spatial resolution Shuttle Radar Topography Mission (SRTM) global digital elevation model (DEM).

2.2 Radiometric Correction

During radiometric preprocessing the brightness values of an image are adjusted to compensate for sensor malfunctions and atmospheric influences. VIR imagery obtained from data suppliers usually represents radiation as relative radiation, denoted as digital numbers (DNs). However, if values from different images are to be compared, this relative radiation has to be converted to physical parameters such as radiance or reflectance. If images taken at different times are to be compared, reflectance should be calculated. Reflectance is the ratio between the amount of energy incident upon a surface and the amount of energy reflected by that surface. It compensates for differences in solar irradiation. Calibration parameters for the specific sensor at the time of the image acquisition are required by these calculations and these are usually stored in the header accompanying the image file. However, to obtain true reflectance or radiance values for the surface of the earth, atmospheric correction should be applied in addition to these calculations. Atmospheric correction aims to eliminate the contribution of the atmosphere to brightness values in remotely sensed imagery. For this study, the implementation of

ATCOR2 in PCI Focus was used for atmospheric correction. The ATCOR algorithm uses physically-based radiative transfer models to model atmospheric conditions and their effect on surface reflectance.

2.2 Pan-sharpening

Image fusion, or pan-sharpening, refers to the fusion of high spatial resolution panchromatic data with lower spatial resolution, higher spectral resolution multispectral data to obtain a combined dataset with both high spectral and spatial resolution. This was only applicable to the 2003 Landsat 7 imagery, as Landsat 5 lacks a panchromatic band. It was performed using the Pansharp algorithm in PCI Focus.

3 Classification

The objective of the classification exercise was to separate agriculture (planted crops) and non-agriculture (built-up, water, natural and semi-natural vegetation, rocky and sandy bare ground).

3.1 Kavango

In Kavango, the approach was to classify the higher spatial resolution pan-sharpened 2003 scenes using a supervised classifier and a maximum of one hour of manual corrections per scene. The classification of scenes from the other two dates was based on manual editing of the classification from the first date, with two hours allocated per scene.

The images were first clipped to the extent of the study area, and then according to each other’s boundaries so that only small overlapping strips remained to avoid gaps in the classification result. The same clipping extents were used for all three years.

The supervised classification was performed through Geographic Object-Based Image Analysis (GEOBIA). This methodology groups individual pixels into readily usable objects based on spectral and contextual information through the process of image segmentation. The classification is then performed on these objects rather than on individual pixels. The software used was eCognition Developer.

Multiresolution segmentation was used to segment the pixels into objects. The strategy was to derive relatively small objects, thereby avoiding objects containing more than one class.

For each scene, several objects for both classes were manually selected as classification samples. These were used to train the RandomForest classifier. The classification was then performed and the results evaluated. The process was iterative; if the results were unsatisfactory more samples were added in problematic areas and the classification reran. When the classification reached the point where adding more samples did not meaningfully improve the result, further sample collection was abandoned. Manual correction was then used to correct misclassified objects. This was also performed within eCognition Developer software. Finally the classification was exported as a shapefile.

The results of the (higher spatial resolution) 2003 classifications were then imported in new projects in eCognition, and compared to the 1993 and 2009 images respectively. Objects that differed from the 2003 classification were manually reclassified. Finally the results were exported as shapefiles.

The classification shapefiles obtained from eCognition were then merged through the union tool in ArcMap. Agriculture took precedence where the classifications were mismatched in areas where images overlapped. Finally the resulting shapefile was converted to a thematic raster.

3.2 Erongo

The low density of agriculture in the Erongo study area rendered the approach used in Kavango impractical. It was likely that a supervised classification would produce more false positives than correct results. Fields were smaller and more dispersed – typically only one or two fields occurred on an extensive farm close to its homestead. Furthermore the lower density of the natural vegetation in this area made it more difficult to identify areas cleared for agriculture. Hence the relatively low spatial resolution of the Landsat imagery would pose a problem.

The solution was to perform a thorough search for current or recent agricultural fields on Google Earth's high spatial resolution imagery. The fields identified on Google Earth were then compared to the Landsat imagery from the three dates to determine changes in classification. In this case the manual reclassification was performed in ArcMap: the kml (Keyhole Markup Language) file created in Google Earth was converted to a shapefile which was copied for each of the three years. These shapefiles were then manually edited.

4. Results and comments

A formal accuracy assessment was not included in the scope of the project; hence a numerical measure of accuracy cannot be provided. However the CGA is confident that the accuracies of the classifications will meet the expectations of the end-users.

The Kavango region is characterised by relatively dense natural vegetation (mostly woodland), however large tracts of it have been cleared for agriculture. The contrast between the bare soil and low crops in agricultural fields and the denser woodlands surrounding them was palpable. However some areas cleared for other reasons (e.g. open spaces around houses) could have been misclassified as agriculture. Natural clearings could also have been misclassified, particularly in the valleys between the parallel linear dunes in the southern section of the area. Fields are complex and fragmented (large commercial and formally irrigated fields are restricted to the Northern section of the area, close to the Kavango river). Hence misclassification of individual fields is unavoidable, but this has been kept to a minimum. The lower spatial resolution of the Landsat 5 imagery might lead to some misclassifications but this is minimised by basing the initial classification on the higher spatial resolution Landsat 7.

In Erongo, the rarity of fields meant that a thorough search for relatively small fields was possible. However there is some difficulty in comparing the lower spatial resolution Landsat imagery to the high resolution recent images from Google Earth. Furthermore fields that were present in 1993 that have since been abandoned could potentially have been missed if they were completely overgrown by the time the Google Earth image was acquired.

5. Conclusion

This has been an interesting and insightful project. Even though the accuracy of the classification results has not been assessed, the CGA is confident that a high quality product has been provided. The CGA keenly awaits the results of the research being conducted by Thomas, and will be happy to be involved in similar projects in future.

Appendix B

Alternative Pseudo-panel Methods to Study Mobility

Antman and McKenzie (2007)

Antman and McKenzie (2007) show how pseudo-panels can be used to measure income mobility in the absence of panel data and also deal with situations where income is measured with error. They argue that pseudo-panels address measurement error in two ways: firstly. aggregating individual level data to cohort level, averages out individual measurement error; secondly, the fact that individuals are only observed in one time period implies the measurement error cannot be correlated over time.

Their mobility measure of interest is the coefficient of lagged income on income in a regression model:

$$Y^*_{i,t} = \alpha + \beta Y^*_{i,t-1} + \mu_{i,t}, \tag{B.1}$$

where $Y^*_{i,t}$ is income for individual i in period t , and β is coefficient of interest. A value of β greater than 1 indicates income divergence while a value less than 1 indicates income convergence in the economy. Income is measured with error at the individual level. Instead of observing true income, Y^* , we observe $Y_{i,t}$ which is equal to:

$$Y_{i,t} = Y^*_{i,t} + \epsilon_{i,t} \tag{B.2}$$

Substituting B.2 into B.1 leads to the following equation:

$$Y_{i,t} = \alpha + \beta Y_{i,t-1} + \mu_{i,t} + \epsilon_{i,t} - \beta\epsilon_{i,t-1}. \tag{B.3}$$

As is the case in the absence of panel data, we do not observe individual i in both periods t and $t-1$. To deal with this problem the authors start by rewriting equation B.3 in terms of cohort averages:

$$\bar{Y}_{c(t),t} = \alpha + \beta\bar{Y}_{c(t),t-1} + \bar{\mu}_{c(t),t} + \bar{\epsilon}_{c(t),t} - \beta\bar{\epsilon}_{c(t),t-1}, \tag{B.4}$$

where, for example, $\bar{Y}_{c(t),t}$ is the average income for all individuals belonging to cohort c observed in time t for time t [1]. $\bar{Y}_{c(t),t-1}$ is the average income for cohort c, observed in time t, for period $t-1$. In the absence of panel data this is also not observed. To deal with this the authors replace $\bar{Y}_{c(t),t-1}$ with $\bar{Y}_{c(t-1),t-1}$ which leads to:

$$\bar{Y}_{c(t),t} = \alpha + \beta\bar{Y}_{c(t-1),t-1} + \bar{\mu}_{c(t),t} + \bar{\epsilon}_{c(t),t} - \beta\bar{\epsilon}_{c(t),t-1} + \lambda_{c(t),t}, \tag{B.5}$$

where, $\lambda_{c(t),t} = \beta(\bar{Y}_{c(t),t-1} - \bar{Y}_{c(t-1),t-1})$.

To get to an estimable equation, the authors assume that the law of large numbers holds. This implies that as $n_c \to \infty : \frac{1}{n_c}\sum_{i=1}^{n_c} \epsilon_{i,t} \to 0$ and $\lambda_{c(t),t} = \beta(\bar{Y}_{c(t),t-1} - \bar{Y}_{c(t-1),t-1}) \to 0$. Thus B.5 becomes the estimable equation:

$$\bar{Y}_{c(t),t} = \alpha + \beta\bar{Y}_{c(t-1),t-1} + \bar{\mu}_{c(t),t} \tag{B.6}$$

Equation B.6 can be modified to include cohort fixed effects. The trade-off between cohort-size and number of individuals influences the requirements for consistent estimation of the model.

Bourguignon et al. (2004)

The authors follow on work by Deaton and Paxson (1994) who studied the second-order moments of cohorts in a pseudo panel context. Their model assumes earnings of individual i from cohort c at time t follows the following equation:

$$y^c_{i,t} = \beta^c_t x^c_{i,t} + e^c_{i,t}, \tag{B.7}$$

where y is the log of earnings, x is a group of individual characteristics, and e is the residual term that contains two components of earnings: the transitory component, and, the unobserved component of permanent earnings.

[1] $\bar{Y}_{c(t),t} = \frac{1}{n_c}\sum_{i=1}^{n_c} Y_{i(t),t}$

Assume that $e^c_{i,t}$ follows an autoregressive process $AR(1)$:

$$e^c_{i,t} = \rho^c e^c_{i,t-1} + \epsilon^c_{i,t} \tag{B.8}$$

where, ρ is a parameter measuring the persistence of earnings shocks, and, $\epsilon^c_{i,t}$ is, what the authors call, the innovation in earnings which has a variance $\sigma^2_{\epsilon c,t}$.

In the absence of panel data it is not possible to observe both $e^c_{i,t}$ and $e^c_{i,t-1}$ which implies that the above model cannot be estimated. They argue, however, that under the assumption that individuals enter and exit the labour market at random between 2 periods, it is still possible to estimate ρ^c and $\sigma^2_{\epsilon c,t}$ by estimating the following equation:

$$\sigma^2_{e,c,t} = \rho^{c2}\sigma^2_{e,c,t-1} + \sigma^2_{\epsilon c,t}, \tag{B.9}$$

where $\sigma^2_{e,c,t}$ is the within cohort variation of the residual $e^c_{i,t}$ and $\sigma^2_{\epsilon c,t}$ is the within cohort variation of $\epsilon^c_{i,t}$. The above equation can easily be estimated if a minimum of three time periods are available. Only three periods though will not, however, lead to very precise estimates.

Bourguignon et al. (2004) note that it is possible to estimate an individual's (observed in period t) vulnerability to poverty in period $t+1$ if two further assumptions are made. Firstly, it must be assumed that the innovation term, $\epsilon^c_{i,t}$, has a mean of zero and variance $\sigma^2_{\epsilon c,t}$. Secondly, it must be assumed that future predictions of variables, coefficients (apart from the constant) and the variance in innovation are available. Variables such as age and gender are easy to predict in the future. If time variant characteristics are used as predictors, stationarity must be assumed (Bourguignon et al., 2004, 5). The probability of an individual in period t falling below a poverty line, Z_{t+1}, in period $t+1$ is:

$$P(y^c_{i,t+1} < Z_{t+1} | x^c_{i,t}, x^c_{i,t+1}, \hat{\beta}^c_{t+1}, \hat{\sigma}^2_{\epsilon c,t+1}) = \Phi(\frac{Z_{t+1} - \hat{X}^c_{i,t+1}\hat{\beta}^c_{t+1} - \hat{\rho}^c e^c_{i,t}}{\hat{\sigma}^2_{\epsilon c,t}}). \tag{B.10}$$

The authors note that the assumption of random entry and exit into the labour market, and also, that the method only estimates parameters for individuals who are continually employed, is an obvious weakness. This is because job losses are possibly the most important source in movements into and out of poverty.

Appendix C

Additional Tables relating to Chapter 4

Table C.1: Regression Models on ln(per Capita Consumption) for South Africa

	2008	2010	2012
Race (ref: Black)			
Coloured	0.07	0.14*	-0.09
	(0.09)	(0.09)	(0.08)
Indian	0.63***	0.47***	0.68***
	(0.16)	(0.16)	(0.15)
White	0.70***	0.71***	0.72***
	(0.09)	(0.09)	(0.08)
Female	-0.54***	-0.52***	-0.44***
	(0.04)	(0.04)	(0.04)
Educ (yrs)	-0.07***	-0.08***	-0.09***
	(0.02)	(0.02)	(0.02)
Educ^2	0.01***	0.01***	0.01***
	(0.00)	(0.00)	(0.00)
Owned Car (2008)	0.65***	0.49***	0.44***
	(0.06)	(0.06)	(0.06)
Age (2008)	0.04	0.02	0.02
	(0.02)	(0.02)	(0.02)
Age^2	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)
Rural (2008)	-0.30***	-0.29***	-0.26***
	(0.05)	(0.05)	(0.05)
Province (2008)	Yes	Yes	Yes
Observations	2581	2585	2585
Adjusted R^2	0.47	0.43	0.45

Source: Own Calculations using NIDS for 2008, 2010 and 2012.
Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
The sample is limited to households where the head is between 25 or older and 55 or younger.
Individual Characteristics are all for the household head.

Table C.2: Estimates of ρ for South Africa

	2008-10	2010-12	2008-12
ρ	0.51	0.56	0.48

Source: Own Calculations using NIDS for 2008, 2010 and 2012.

Table C.3: True Estimates of Economic Mobility in South Africa vs. Bounds Generated using Dang et al. (2014)

Method	$P_{08}; P_{12}$		$N_{08}; P_{12}$		$P_{08}; N_{12}$		$N_{08}; N_{12}$	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Non Parametric	39.4	21.3	0.1	18.3	8.6	17.0	51.9	43.4
Parametric								
$\rho = \{1, 0\}$	30.2	16.7	0.0	13.6	7.6	21.2	62.1	48.5
$\rho = \{0.7, 0.35\}$	24.5	20.2	5.8	10.1	13.4	17.7	56.4	52.1
True Estimate	29.1	.	10.5	.	15.9	.	44.4	.

Source: Own calculations using NIDS for 2008 and 2012.
The sample is limited to households where the head is between 25 or older and 55 or younger.
P_{08} and P_{12} refer to "Poor" in 2008 and 2012 respectively. Similarly N refers to "Not Poor". "Lower" and "Upper" refer to the two bounds on **mobility**. For **immobility** "Lower" is the upper bound and "Upper" is the lower bound.