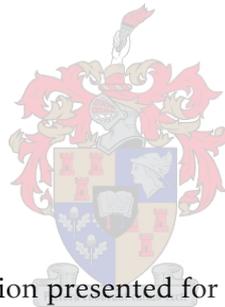# Manifestations of inequality in three developing countries: An investigation of differential education and labour market outcomes

by

Joel Roosevelt Gondwe

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

Supervisor :	Prof. Servaas Van Der Berg
Co-Supervisor :	Prof. Rulof Burger
Co-Supervisor :	Dr. Debra Shepherd

Degree of confidentiality: A   March 2021

PLAGIARISM DECLARATION

1. Plagiarism is the use of ideas, material and other intellectual property of another's work and to present it as my own.

2. I agree that plagiarism is a punishable offence because it constitutes theft.

3. Accordingly, all quotations and contributions from any source whatsoever (including the internet) have been cited fully. I understand that the reproduction of text without quotation marks (even when the source is cited) is plagiarism.

4. I also understand that direct translations are plagiarism.

5. I declare that the work contained in this dissertation, except otherwise stated, is my original work and that I have not previously (in its entirety or in part) submitted it for grading in this dissertation or another dissertation.

J.R. Gondwe	March 2021
Initials and surname	Date

ACKNOWLEDGEMENTS

Growing up, I never thought I would ever go as far as pursuing a doctorate in anything, much less one in economics. While I could have given many reasons why that was the case, they can all be summed up in just one overwhelming sentiment—I didn't think I was capable. I didn't think I would be capable of the hours required, I didn't think I would be capable of the intellectual rigour, of the mental toughness, the stamina, the focus, and a whole host of other attributes I had told myself were required to attain such a high achievement. In short, I didn't really believe in myself or my abilities.

This thesis is therefore a direct product of every single person throughout my life's journey that took the time to communicate very firmly to me—in word and in deed—that I was capable. This thesis, I dedicate wholly to them. While there are far too many names to mention, I list a few here.

First, to God, words will always fail me, as they rightly should, but I am grateful for EVERY-THING.

Second, to my supervisors, Prof. Servaas van der Berg, Prof. Rulof Burger, and Dr. Debra Shepherd, thank you so much for your tireless guidance, support, assistance and encouragement over the past four years. I feel very privileged to have been under your tutelage and mentorship for as long as I have been. Thank you for your investment in me, thank you for your patience and thank you for taking interest in me that went beyond what was required of you. To Prof. van der Berg in particular, thank you for the brief chat we had in my master's year when you suggested I stay on for a PhD. Thank you for seeing potential in me and sparing no expense in my development. Your thoughtfulness and leadership have impacted me as a person tremendously.

I further thank the teaching staff of the Economics Department led by Prof. Andrie Schoombee. Right from my Honours year, the department has been nothing but good to me, extending kindness, funding, and opportunities not just to me, but also to my fellow Malawians who came after me in the department. On behalf of all of them too, I wish to say THANK YOU. I further thank my fellow researchers in the Research on Socio-Economic Policy group for their support and encouragement. In addition, I am grateful to the GEMS programme led by Dr. Jaco Franken who provided funding for my PhD.

To my parents, Roosevelt and Ebba Gondwe, thank you so much for the unwavering love and

support you've shown me my whole life. Dad, thank you for showing me the virtues that lie in hard work, in being inquisitive, and asking questions. Mum, thank you so much for helping me to love books from a young age. Thank you in particular for ensuring the competitive spirit my siblings and I all share and were raised in never burnt us out, but rather caused us to always strive for more. Dad, Mum, thank you for your sacrifices that made my education possible. To my siblings, Temwa, Kondwani, and Ruth thank you for being the best family I could have ever asked for.

To Duwa Gondwe, my wife and mother of our first little girl, THANK YOU. In many ways, my being here wouldn't have even been possible without your constant love, encouragement and support all these years. Thank you for believing in me. Thank you in particular for all the sacrifices you have gladly made, to ensure that I keep on and pursue this goal. I would not be where I am today without you.

To My Fam, THANK YOU. I appreciate all the support during this entire time. You have all been a Godsend. I couldn't have asked for a better support system. As we would loosely say back home, "You know yourselves!".

Lastly, to all the teachers that I, and many others like me, have had over the years, who have tirelessly and repeatedly poured themselves out into their students, I thank you immeasurably for your commitment.

ABSTRACT

Economic inequality is a pervasive phenomenon that has long been a feature of both developed and developing countries. Such inequality is particularly problematic in developing countries because of its potential to exacerbate already established and yet detrimental features that characterise most developing countries, such as high unemployment, a large informal sector, and high population growth rates. This thesis investigates how inequality manifests itself through differential student outcomes and also attempts to examine whether or not technological growth destroys jobs and worsens wage inequality.

Chapter 2 investigates the relationship between student performance and socio-economic status (SES) among grade six students in Malawi and Namibia. Malawi was specially included in this cross-country comparison because it is likely that the underlying mechanisms that govern the typical/expected SES-performance relationship do not hold in countries as poor as Malawi as they would in more advanced developing countries like Namibia. Using OLS regressions and hierarchical (multilevel) models, the results show an approximately flat socio-economic gradient for student education performance for Malawi, for both the full sample and the reduced samples (urban and rural). In Namibia, in contrast, SES appears to be correlated with student performance. However, this is primarily driven by students who live in urban areas, whereas, like Malawi, rural Namibia also has an approximately flat socio-economic gradient.

Chapter 3 builds on this by taking special interest in research in low-income countries like Malawi and the challenge that arises when, in the absence of income/expenditure information, one has to rely on an asset index to distinguish among individuals of comparable SES levels. This follows on discussions in the literature that have well articulated the difficulty asset-based measures have of doing so especially among various shades of poor individuals, that is, differentiating the poor from the very poor: a feature which a measure used in very poor countries like Malawi should have. Chapter 3 explores the use of finite mixture modelling as an alternative approach to achieving this goal. The findings suggest that using this approach makes it possible to distinguish between individuals' relative SES level in a meaningful way.

Lastly, Chapter 4 is primarily interested in examining if technological growth in South Africa contributed to exacerbation of wage inequality and job loss during the period from 1997 to 2015.

This analysis is done through the lens of a routine-biased technological change framework whose main hypothesis is that recent technological advancements are biased towards replacing labour in routine tasks. This chapter presents findings from descriptive analysis, OLS regressions, as well as a non-linear systems estimator applied to a normalised CES production function. The results show both descriptive and empirical evidence of a hollowing out of middle-skilled work (done by workers whose occupations typically involve a high share of routine tasks). Further, these findings are differentiated by gender and race.

OPSOMMING

Ekonomiese ongelykheid is ń algemene verskynsel in beide ontwikkelde en ontwikkelende lande. Sulke ongelykheid is veral problematies in ontwikkelende weens sy potensiaal om reeds bestaande negatiewe kenmerke van die meeste sulke lande te vererger, soos hoë werkloosheid, ń groot informele sektor en snel bevolkingsaanwas. Hierdie proefskrif ondersoek hoe ongelykheid in verskille in leerling-uitkomste manifesteer, en bestudeer ook in watter mate tegnologiese groei werkgeleenthede uitwis en loonongelykheid vererger.

Hoofstuk 2 ondersoek die verhouding tussen leerling-uitkomste en sosio-ekonomiese status (SES) onder graad 6 leerlinge in Malawi en Namibia. Malawi is veral by hierdie vergelyking tussen twee ontwikkelende lande ingesluit, omdat die meganismes wat gewoonlik die verband tussen SES en leerling-prestasie onderlê moontlik nie so goed in ń land so arm soos Malawi sal geld as in ń land wat verder ontwikkel is soos Namibia nie. Met gebruik van gewone kleinste-kwadrate regressies en hiërargiese (multivlak) modelle word bevind dat Malawi se sosio-ekonomiese gradiënt byna heeltemal plat is, beide vir die volle steekproef en vir aparte landelike en stedelike groeperings. In Namibia, daarenteen, styg leerling-uitkomste oënskynlik met SES, maar dit word veral deur stedelike leerlinge verklaar, terwyl die sosio-ekonomiese gradiënt in landelike gebiede plat is, netsoos in Malawi.

Hoofstuk 3 bou hierop voort deur spesiale aandag te gee aan navorsing in lae-inkomste lande soos Malawi en die uitdaging wat ontstaan wanneer navorsers weens die afwesigheid van inkomste of bestedingsdata verplig is om ń bate-indeks te gebruik om tussen leerlinge met soortgelyke SES te onderskei. Dit volg op besprekings in die vakliteratuur wat uiteensit hoe moeilik bate-indekse veral tussen arm en baie arm individue kan onderskei – ń kenmerk wat erg nodig is in baie arm lande soos Malawi. Hoofstuk 3 gebruik sogenaamde "finite mixture"-modelle as alternatiewe metode om hierdie doel te bereik. Die resultate dui daarop dat dit ń sinvolle benadering bied om tussen verskillende individue se relatiewe SES te onderskei. Ten slotte is die vraag in Hoofstuk 4 in watter mate tegnologiese verandering in Suid-Afrika van 1997 tot 2015 tot loonongelykheid en verlies aan werkgeleenthede bygedra het. Hierdie analise gebruik ń roetine-gebaseerde tegnolo-giese veranderingsraamwerk. Die belangrikste hipotese van hierdie raamwerk is dat onlangse tegnologiese verandering geneig is om arbeid te vervang vir roetine take. Hierdie hoofstuk bied resultate van beskrywende analise, OLS-regressies, sowel as ń nie-liniêre stelsel-beramer toegepas

op ń genormaliseerde CES produksiefunksie. Die resultate toon beide beskrywende en empiriese getuienis van ń uithol-proses van beroepe met ń groot mate van roetine-take, meestal beroepe in die middel van die geskoolheidsverdeling. Die resultate word ook volgens geslag en ras gedifferensieer

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1

CHAPTER 1

Introduction

1.1 Background

From the 1950s, just when development economics was emerging as a major sub-discipline of economics, the commonly held view among economists was that a period of growing income inequality was an inevitable feature of a country in pursuit of economic growth (Ravallion 2014, Stiglitz 2015). What this meant for developing countries was that the policies that took centre stage were those aimed at economic growth that was characterised by poverty reduction even at the expense of rising income inequality. To make it worse, it was equally accepted that policies aimed at addressing both inequality and economic growth would in fact impede economic growth, so that poor countries would achieve only limited economic growth due to a lack of policy focus. This idea went on to be the mainstream view amidst high and rising inequality that was accompanied by (at times) persistent poverty in these developing countries. As (Ravallion 2014, p.851) put it, "the policy message for the developing world was clear: you cannot expect to have both lower poverty and less inequality while you remain poor, and, if you choose to give poverty reduction highest priority, then focus on growth". It would not be until the 2000s that this long-standing view of inequality would experience sufficient scrutiny, leading to new ways of thinking about how inequality could in fact threaten a country's prospects of development.

For developing countries specifically, this shift marked a much-needed departure from thinking of income inequality as an inconsequential and temporal feature of economic progress that self-corrects with limited external impetus and action to one that requires as much attention as had previously been given to poverty reduction and economic growth. Indeed, studies have now shown that economic growth in itself does not unequivocally lead to reductions in income inequality in all developing countries (Fields 1989, Ravallion 2006). To the contrary, evidence from cross-national studies strongly point to a more mixed finding: for some developing countries like Brazil and Indonesia, economic growth was accompanied by decreasing income inequality, whereas there are many instances where economic growth was accompanied by high and rising income inequality, like South Africa, Russia, and China (OECD 2011). However, even though eco-

nomic growth in Brazil and Indonesia was poverty reducing, the presence of inequality tended to impede pro-poor growth. Since that time, a lot of effort has been given to investigating the pervasive phenomenon of income inequality in developing countries.

There are several reasons why income inequality in a developing country context would threaten future development, or indeed undermine and nullify efforts aimed at achieving pro-poor growth. Ravallion (2006, 2014) suggests three related ways. Firstly, the extent to which income inequality changes—moving up or down—during the process of economic growth has a bearing on how much the poor benefit from realised growth. Secondly, an initially high incidence of income inequality in a country often implies that the share of the economic growth that actually accrues to the poor is small. Thirdly, even in the absence of increasing income inequality, a high level of income inequality leads to less growth, because it reduces potential contributors to that economic growth. Thus, the degree to which an economy enjoys the potential gains from economic growth is greatly impacted by inequality.

Over the years, several sources of income inequality that are unique to developing countries have been put forward. Some of the more common ones include widespread gaps in education, missing or insufficient political will, a large and growing informal sector, increasing bargaining power of those with high earnings, and a labour market which discriminates against parts of the labour force (typically women) when it comes to hiring and lifetime career mobility (Coady & Dizioli 2018, Fields 2011). Generally, a direct consequence of having such discriminatory practices is that labour force participation is always lower than what it could be if all groups were allowed equal participation, and that the labour force is itself comprised of an abundance of low-skilled workers. Despite its potential to impede growth, and how widespread it is, inequality in developing countries is not without a remedy, nor is it inevitable. Its complexity, however, demands a multipronged approach that is both robust in its methods and nuanced in its appreciation of contextual intricacies that distinguish between inequality in one context, to the next. That is to say, while some policies may work in a variety of contexts, a remedy for income inequality in one context cannot always be applied in another context in the exact same way and achieve similar results.

Despite this, one policy instrument that has (rightfully) grown in popularity and become seen

as an important tool to combat income inequality is education (Coady & Dizioli 2018). Because of this, governments and policymakers in developing countries especially, have concerned themselves with expanding access to education, given that, historically, large proportions of children in these countries were not enrolled in schools (which itself is in part a result of inequality). What followed was an unprecedented expansion of access to schools in several developing countries, which in some instances was accompanied by economic growth (Barro 2001, 2013, Hanushek 2013). Unfortunately, as further research would soon show, despite the expansion of (mostly primary) school access, many other developing countries did not experience economic growth that was accompanied by steep declines in inequality and marked reductions in poverty. What became abundantly clear was that the different experiences with inequality had less to do with equitable access to education alone (i.e. enrolment), but also, and more importantly, equitable access to quality education (i.e. actual learning once a student is enrolled). Indeed, a country's experience of declining income inequality in the medium to long term was significantly enhanced when it gave definite attention to reducing inequality in both access to education, and school quality.

Specifically, equitable access to quality education works to mitigate inequality in several ways, some of which include breaking the generational transmission of poverty (Walker et al. 2019), increasing labour force participation (Fields 2011, Pritchett 1999), and crucially, reducing inequality of opportunity (Alvaredo et al. 2018). This last point, opportunity, is of particular significance and warrants further explanation due to the negative relationship between inequality and equality of opportunity. That is, the more unequal a country is in terms of income, the fewer opportunities those at the bottom of the distribution have to excel. What this means is that children born to low-income parents are already born at a significant disadvantage, even if they possess equal talent to children born to high-income parents. As (Corak 2013, p.98) thoughtfully put it, "inequality … shapes opportunity" and in so doing determines not only academic outcomes, but also and more importantly, labour market prospects, which ultimately determine one's quality of life (at least economically).

This alludes to what is perhaps one of the greatest appeals of education as a policy tool over and above other policy instruments—the idea that education has the potential to level the playing field and give all individuals equal opportunity to make something of themselves (Walker et al. 2019). Unfortunately, despite their promise, education systems more often than not fall short of

this crucial goal: failing to impart in school leavers the necessary technical skills and flexibility required to integrate into a modern economy—what most would imagine to be the yardstick by which, at least from an economic perspective, educational systems must be measured (Van der Berg 2002).[1]

This gives further credence to why international interest has shifted towards ensuring provision of quality learning in schools. In fact, the timing of these interest shifts is important, and some would argue even critical, as more recent research continues to show evidence of changing employment trends in the labour market that are characterised by increases in relative labour demand for skilled and educated workers in place of uneducated and unskilled workers, what some have loosely referred to as the "4th industrial revolution" (Acemoglu & Autor 2010, Goldin & Katz 2007a, Manda & Ben Dhaou 2019). Among the competing reasons, it is argued that these changes stem from globalisation, offshoring and increased adoption of technologies that are biased against unskilled workers—all of which have the net effect of changing the landscape of the modern economy even further, and making it hostile towards certain kinds of skills (Goos et al. 2014, Wood 1998). Where workers continue to be paid their marginal product, this development can only lead to more inequality if little is done to prepare the labour force of the future to function productively in such a labour market. Fortunately, there has not been a lot of penetration of such technologies in developing countries (Das & Hilgenstock 2018). While time is on their side, therefore, it is imperative that developing countries do more with their education systems to reduce the proportion of their population living in poverty, reduce the gap between the rich and the poor and take advantage of the delays in technological adoption to train and equip a new workforce. It is this broad context that serves as the setting for the discussions in this thesis.

1.2 A research response: overview of this thesis

1.2.1 Research questions, data, and methods

This thesis is an empirical contribution that is primarily concerned with investigating different ways in which inequality manifests itself in developing countries. It does so chiefly by looking

---

[1]While this remains the benchmark schools are held to, it is certainly an oversimplification as other critical factors also come into play to ensure student learning. Some of these include home background and parental factors, student health, and distance to learning facility

at how student performance is affected by socio-economic status in the earlier years, and goes on to analyse how the employment trends of workers with different levels of skill (low, middle, and high) respond to the adoption of new technologies in the labour market in the later years. The specific research questions that guide these discussions are: (1) Does socio-economic status have an impact on student performance in Malawi and Namibia? (2) Are these impacts differentiated by urban or rural location? (3) Is there a better way of identifying socio-economic clusters, given a socio-economic status index? (4) Is there evidence of routine-biased technological change in the South Africa labour market? (5) How has routine-biased technological change affected employment trends across demographic groups, economic sectors and occupations? (6) Are the trends in employment and wages consistent with labour demand that would be produced by technological progress? (7) Are the results robust to using different measures of routine?

In answering these questions, use is made of several data sources. For Chapters 2 and 3 I use nationally representative third wave data collected by the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) that assesses grade six students in both literacy and numeracy and also collects asset-ownership information which is used to construct a socio-economic index. This data collection process commenced in 2006 and was completed during 2011. The main econometric techniques applied to this data are ordinary least squares regressions, hierarchical linear modelling and finite mixture modelling. On the other hand, the primary source of information used for Chapter 4 is the Post-Apartheid Labour Market Series (PALMS). This is a stacked cross sectional dataset created by DataFirst at the University of Cape Town (UCT) and currently represents South African labour market data from 1995 to 2019. Other supplementary sources of information used in Chapter 4 included output data from the South African Reserve Bank (SARB) and occupation task content data from the World Bank's Skills Towards Employability and Productivity (STEP) survey (2012-2017). Here, econometric analysis was based on a combination of OLS regressions and the application of a non-linear systems estimator to a normalised CES production function.

The thesis limits itself to three developing countries that are, however, at different points along the development paths: Malawi, a typical example of a low-income country; Namibia, which only recently transitioned from a lower middle-income country to an upper middle-income country (in 2008); and South Africa, which has long been categorised as an upper-middle income country,

but has still struggled to incorporate the bulk of its population into the economic mainstream. Sections 1.2.2 and 1.2.3 below offer a brief discussion of each of the chapters. Section 1.2.2 specifically includes discussions on both Chapters 2 and 3 due to their related nature in that they both analyse different facets of socio-economic status in developing countries, whereas Chapter 4 is discussed in section 1.2.3 by itself, since it rather focuses on the labour market instead.

1.2.2 Socio-economic status and student reading performance

Socio-economic status has long been shown to positively impact a student's classroom performance with the advantages of students from economically advantaged homes persisting from as early as kindergarten through to university, and ultimately the labour market (Bailey & Dynarski 2011, Van der Berg 2002, Reardon 2013). It is therefore important when analysing education systems to have this finding in mind, as opposed to merely looking at average student performance that does not take socio-economic status into consideration.

Chapter 2 does exactly this by investigating reading achievement amongst primary school students in Malawi and Namibia using the aforementioned third wave of SACMEQ data. The main focus is on inequalities in performance within and between these two countries, and the impact of student socio-economic status on student performance. The analysis goes a step further to assess whether student reading performance is differentiated by school even after accounting for student SES. For Malawi especially, the findings are disheartening. Despite universal evidence of the expected impact of socio-economic status on student performance being monotonically increasing, the socio-economic gradient for Malawi is approximately flat for the entire range of SES scores. While this implies a relatively fair education system because the achievement gap between the poorest and richest student in Malawi is small, this fair system performs at levels even below that of the SACMEQ regional average. In other words, student performance in Malawi plateaus far below accepted achievement scores, even for most children from more affluent homes. This means that for stakeholders such as parents, there appears to be no visible impact on a students reading performance itself from them becoming wealthier.

For Namibia, the benefits of coming from a richer background are more pronounced. There is a significant difference in reading performance between low SES students and high SES students. Further analysis, however, reveals that these differences are primarily driven by students in urban

areas, while student performance in rural Namibia, like that of Malawi, seems unresponsive to socio-economic advantage. These findings relay the crucial need for further research that is aimed at investigating factors that actually impact student performance in low-income areas, since the results go against the norm. It further exemplifies the difficulty met when dealing with areas or communities that were previously disadvantaged or poor, since the typical methods often applied may not work with as much vigour.

In an attempt to draft better pro-poor policies for such areas, one critical step is being able to group individuals into socio-economic clusters that best describe their actual level of wealth. This informs both targeting of policy interventions and the ability to retrieve pertinent information about the said cluster. In the presence of income data, this exercise would be as simple as grouping individuals/students from households with selected per capita incomes into wealth brackets, since the ranking of incomes is more straightforward. However, doing so in low-income countries is not always possible because most surveys do not collect information on income, but rather asset ownership. With this information, the norm, following Filmer & Pritchett (2001), is to create some index which acts as a proxy for socio-economic status. At this point, clustering individuals into wealth groups is then achieved by some arbitrarily defined cut-off point such as a 40-40-20 classification. That is, the index is split into the poorest 40 percent, the middle 40 percent, and the richest 20 percent. While this is beneficial given the circumstances, it is not a statistically robust approach that importantly allows the data to "speak for itself" . Using SACMEQ III asset ownership data for Malawi, Chapter 3 explores a data-driven approach to clustering which is applied to the SES wealth index: finite mixture modelling. The findings suggest that finite mixture modelling excels at being able to distinguish between the different clusters within socio-economic status, going a step further to reveal a rural population that is, perhaps surprisingly, more heterogenous than the urban population.

1.2.3 Does technological growth destroy jobs and exacerbate wage inequality in middle-income countries? Routine Biased Technological Change and the South African Labour Market

From around the late 1970s, global wage inequality started increasing (Katz & Autor 1999). Research suggested that this increase was mainly a result of wage differentials between the different types of skills and occupations in the labour market. Among the several hypotheses put forward

to explain these troubling trends, one that grew in prominence and popularity among researchers was skills-biased technological change. Put simply, this was the idea that the technology being adopted and used in the production process was changing in ways that were biased towards workers with skills, and against workers with little or no skills. This had the unfortunate consequence of raising relative labour demand (and wages) of workers with high skills, while demand and wages of low-skilled workers declined. While this model proved successful in explaining the wage and employment trends of that period, it failed to account for more recent trends that were instead characterised by the disappearance of middle-skilled jobs, but increases in low- and high-skilled jobs. In response, Autor et al. (2003) put forward a more nuanced and refined version of skills-biased technological change called Routine-Biased Technological Change (RBTC).

The central idea of this RBCT framework is that work is thought of as being a series of tasks, and using this task framework, Autor et al. (2003) argue that the workers whose jobs are disappearing are the ones doing jobs that involve a high share of routine tasks, as those are the tasks that are easily replaceable by advancements in technology. These tasks happen to be performed by workers in the middle of the skills distribution, and that is why there is a hollowing out in the middle of the skills distribution. While this model successfully explained the pervasive phenomenon of job polarisation in the labour market, it has mostly been applied to the U.S labour market (Autor et al. 2003, Dorn & Autor 2013), and more recently to other developed countries, primarily in Europe (Goos & Manning 2007, Green 2012, Spitz-Oener 2006). Because of this, the norm in the literature has been to measure routine using a composite routine task intensity measure that uses US-based occupation schemes. Evidence for RBTC in developing countries is sparser, and to the best of my knowledge no study has explicitly been done so far for South Africa.

Chapter 4 commences with an overview of the literature as pertains to skills-biased technological change and how its shortcomings directly led to the development of a more refined version of it—routine-biased technological change. The empirical section starts by providing a descriptive analysis of the overall employment share changes in the private sector of the South African labour market from 1997 to 2015, with the intention of affirming whether South Africa, like some developed countries, also experienced a hollowing out of middle skilled jobs that were displaced to the bottom and the top of the skills distribution. These trends are presented for the full private sector sample, as well as by race and gender. The results are further disaggregated by occupation

to see which particular occupations were driving the observed trends, and a standard shift-share analysis is performed to decompose the observed changes into their *within* and *between* industry components. Descriptive analyses for wage trends during this period are also presented and discussed.

The multivariate analysis component in Chapter 4 attempts to do two things. Firstly, to demonstrate that the observed trends in wages and employment are indeed driven by routine task intensity. This is done using a measure of routine that is based on US occupation schemes, and for the first time in South Africa, measures of routine that are based on occupation schemes developed specifically for use in developing country contexts. Secondly, to go a step further than this, non-linear system estimation techniques are used to estimate a fully specified RBTC model. This model is used to investigate whether RBTC can provide a coherent account of observed wage and employment trends in a labour market where technological growth decreased the relative productivity of workers in the middle of the skills distribution.

The findings in the chapter reveal that there was indeed a hollowing out of middle-skilled jobs in the South African private sector during the 1997 to 2015 period. Unfortunately, the absorption of those workers into low- and high-skilled occupations shows that while previously advantaged race groups (white and Indian) move into high-skilled jobs, it is the previously disadvantaged race groups (black and coloured) that are absorbed into low-skilled jobs. These results are further differentiated by gender, with men more often moving to high-skilled occupations while women tend to move to the less-skilled occupations. Further, the results from the econometric analysis affirm that technological growth occurred in such a way that it reduced the relative productivity of middle- relative to low- and high-skilled workers.

1.3   Summary: Contributions and limitations

This thesis contributes to the current literature by adding further analysis and tools for examining socio-economic status versus outcomes relationships and socio-economic gradients in education. The tools can, however, be applied to other areas of research apart from education. In addition, the thesis shows clearly that when it comes to developing countries, a one-size-fits-all approach is inadequate to understanding the mechanisms that influence student performance in these con-

texts. This is seen by how unresponsive student performance in Malawi and rural Namibia is to socio-economic status, which deviates from standard economic theory about how household economic background interacts with individual students to produce improved learning outcomes.

Further, to the best of my knowledge, this thesis represents a first attempt at explicitly testing for RBTC in South Africa. To add to this, it is the first attempt to do so using task content data that is specifically constructed for use in a developing country context. As South Africa is the most developed country in the sub-Saharan Africa (SSA) region, and indeed on the African continent, it is befitting that South Africa be used as the testing ground for the effects of RBTC on the labour market. Indeed, South Africa is atypical, with aspects of both a developed and a developing country. While most developing countries in the region are yet to adequately adopt new technologies (Das & Hilgenstock 2018), South Africa as an upper-middle income country is further along that process and would therefore be more likely to see implications of doing so in the labour market, which this thesis discusses. The findings for South Africa therefore offer a much-needed cautionary tale for other developing countries in the region to learn from. Particularly, seeing that most developing countries are net-importers of technologies and not creators, it is even more critical for them to focus on reducing educational inequality if the kind of students that are produced are to function and contribute meaningfully to the modern economy.

Despite these contributions, this thesis is not without its limitations. Perhaps the most notable of these lies in the data in that one can only use what is available. In turn, this influences both the kind of questions that can be answered, and the extent to which they can be researched. In addition, there are several factors that contribute to inequality in developing countries and it was not possible for all of them to be included in this thesis, nor indeed could the thesis provide analysis for all developing countries. Instead it is limited to the aforementioned three. However, the use of nationally representative surveys means that the bulk of the findings discussed can be generalised to other developing countries albeit to varied degrees of statistical confidence.

CHAPTER 2

Socio-economic status and student reading performance in Malawi and Namibia: Findings from SACMEQ

2.1   Introduction

The study of the association between student achievement and its relationship to parent and household socio-economic status has a long-standing tradition in the social sciences, dating back to the seminal works of Coleman (1968) and White (1982). While one aspect of this research is primarily concerned with what types of schools are better suited to educate students from differing socio-economic backgrounds (Raudenbush & Willms 1995), another aspect focuses primarily on student performance itself, and how it is affected by the socio-economic status (SES) of students' parents and/or household (see for example Caro et al. (2009), Sewell & Hauser (1975)). This study follows the latter aspect.

Motivation for this kind of research is abundant: by the time students enter kindergarten, those from households with a higher socio-economic status already begin to exhibit superior performance to students from relatively poorer households. This income achievement gap is not only maintained, but in some situations becomes wider as students progress through school (Reardon 2013). This implies that the advantage bestowed on children from a higher-income/SES group, even prior to kindergarten, may persist through the duration of compulsory education.[1]

Moreover, once these children enter school, they are not only less likely to perform well, but, as research has shown, are also more likely to complete fewer years of schooling and to leave school earlier (Crane 1991, Janosz et al. 1997, Rumberger 1995, Sirin 2005, van Ewijk & Sleegers 2010, Polidano et al. 2013). This would, in turn, make it more difficult for them to enter the labour market or, once they enter it, be less upwardly mobile than their more educated (often higher

---

[1]Some explanations that have been put forward to explain this include: the family investment model which argues that parents from a higher socio-economic position are more willing to invest in their children in ways that have a positive effect on the child's academic performance (Xiaofeng et al. 2018); the toxic stress theory which offers evidence that suggests that children that experience early life toxic stress and live in high-risk circumstances – as poorer households often do - can have damaging effects on educational outcomes (Williams Shanks & Robinson 2013); and some have shown evidence that children who grow up in poorer families tend to have poorer cognitive development than those who grow up in wealthier homes (Hertzman & Wiens 1996)

SES background) counterparts. It is no stretch of the mind that this can easily become a self-perpetuating cycle, persisting from one generation to the next, locking an entire population of students with low SES backgrounds into a state of subpar performance.

Historically, an abundance of research has shown conclusive evidence of a positive relationship between student achievement and parental SES (Sirin 2005, White 1982, van der Berg 2008, Taylor & Yu 2009). Typically, this relationship is referred to as the socio-economic gradient as it shows socio-economic outcomes across the entire range of SES: that is, it shows how student performance changes as a students' socio-economic status improves or declines (Caro et al. 2009). Compared to the 'traditional' approach of assessing student performance, this serves as a better method of doing the same. The traditional approach, though insightful, relied solely on what are widely described as 'league tables' — where school systems were ranked based on the average test score performance of their students. The obvious shortcoming of ranking school systems in this way is that it ignores universal research findings that show unequivocally that children from higher socio-economic backgrounds outperform their lower socio-economic background counterparts (Ross & Zuze 2004). Doing so, Ross & Zuze (2004) argue, is not a fair approach of assessing student performance specifically, nor education systems more generally. Both of these assessments, however, can be answered using socio-economic gradients.

This chapter contributes to the literature on socio-economic gradients by conducting a comparative cross-national study that looks at the reading score performance and its relationship to student socio-economic status in two African countries: Malawi and Namibia. These countries are at least comparable in three ways: first, they both introduced multiparty-led democracies in the early 1990s (1993 for Malawi, and 1990 for Namibia); secondly, they both have sizeable proportions of their populations living in rural areas (81 percent for Malawi and 50 percent for Namibia); and thirdly, they are both designated as being part of the sub-Saharan Africa (SSA) region.

Other than these similarities, some of their geographic and demographic differences also present interesting points of departure that warrant investigation as pertains to student performance in these two countries. First, there are notable differences in population size: Malawi's population size of about 18 million is notably larger than that of Namibia (about 2 million). Second, Namibia is more richly endowed with natural resources than Malawi. Third, whereas Malawi is classified

13

as a low-income country, Namibia is classified as an upper middle-income country (World Bank 2010, 2017, Fantom & Serajuddin 2016)[2]. Finally, Malawi was one of the earlier adopters of free primary education (FPE) in 1994, whereas Namibia did so only in 2013. It can be expected that all these differences culminate in student performance that interacts differently with one's socio-economic background, as this chapter seeks to investigate.

The remainder of this chapter is organised as follows. Section 2.2 discusses briefly the international push for expanded education access in Africa and the background and profiles of Malawi and Namibia. This serves as a contextualisation of the environment where the students learn. Section 2.3 describes the data and methodologies used in this chapter. Section 2.4 presents the main findings, and finally, Section 2.5 concludes the discussion.

2.2 Background and country profiles

2.2.1 Education for All (EFA) in Africa

The international commitment to Education for All (EFA) was initially made at the World Conference on Education in Jomtien, Thailand, in 1990. Specifically, governments had determined and agreed that primary education was a right that needed to be extended to every child, and that they should do everything in their power to ensure that children of eligible age be guaranteed access to, at the very least, primary school education (Kattan 2006). What followed in Africa was a substantial expansion in school enrolment, even though most countries continued to be far from attaining the enrolment and survival rates observed in developed and developing countries elsewhere (Majgaard & Mingat 2012).

Despite this positive development, much of the literature on schooling in Africa maintains that this (often rapid) increase in student enrolment led to a deterioration in the ability of education systems to produce actual learning (Chimombo 2009, Colclough et al. 2010). In other words, evidence points to the fact that as access to schools improved, the quality of education offered declined over time, leading to increased numbers of children failing to acquire basic literacy and numeracy competencies. This is a critical finding, especially given the findings of Hanushek &

---

[2]This well-known method of classifying countries by their income has been in use by the World Bank since 1989 and divides countries into four groups—low income, lower middle income, upper middle income, and high income—using gross national income (GNI) per capita valued annually in US dollars (Fantom & Serajuddin 2016)

Woessmann (2008) that point towards educational quality as more important in determining economic growth and labour market outcomes than educational attainment alone.

Indeed, more recent times have witnessed a steady shift in attention away from policies geared solely to improving educational access to ones that also address the need for improved educational quality. An example of this new emphasis on access to quality learning is clearly reflected in the United Nations Sustainable Development Goals (SDGs).[3] Whereas Goal 2 of the former Millennium Development Goals (MDG) agenda stipulated the achievement of universal primary education, Goal 4 of the current SDGs is, instead, to improve and ensure quality education for all students.

While some countries have managed to simultaneously improve access to schooling and learning, this has not been the case everywhere (Crouch & Vinjevold 2006). Research has shown that in the region of sub-Saharan Africa specifically, rapid expansion of schooling appears to have reduced school quality due to, among other reasons, overcrowding and resource constraints that could not keep pace with increasing demand (Glewwe et al. 2014, Taylor & Spaull 2015). Malawi, one of the countries under investigation in this chapter, has been no exception: after hasty expansion that lacked adequate and/or sufficient planning, the Malawian education system was not ready to accommodate the sudden influx in enrolment rates. Studies analysing the impact of this sudden surge of students due to the introduction of FPE in Malawi have well documented its adverse effects on education quality (see for example Chimombo (2009)).

As an aside, it is important to note that in most cases, education quality is proxied by student performance in standardised tests. Using this measure, education quality would, therefore, be said to decline over time if students' average test scores on these tests were trending downwards, and improving if, over time, students' average test scores trend upwards. While this traditional way of assessing school quality strengthened the view that increased student enrolment led to declines in education quality, more recent research by Taylor & Spaull (2015) offers an alternative perspective of assessing an education system's performance that takes into consideration increases in student enrolment. In doing so, they introduce what they refer to as a measure of 'access-to-learning'. In contrast to adopting comparisons of mean test scores over time, this approach measures an

---

[3]Sustainable Development Solutions Network (United Nations SDSN 2016).

education system's performance through combining access to schooling (i.e., enrolment) with the quality of learning (i.e., test scores). To operationalise this, Taylor & Spaull (2015) propose taking the product of the proportion of children surviving to a particular grade (typically grade 6) and the proportion of children in that grade who reach acceptable levels of literacy and numeracy competence. This provides an additional tool with which to assess education systems that better captures the progress (or lack thereof) of an education system even in the event of declines in student average test scores in response to increased enrolments.

This is shown to be exactly the case for Mozambique, where rapidly expanded access to primary schooling was associated with lower average test scores as a result of the inclusion of previously excluded children (Taylor & Spaull 2015). However, this higher enrolment simultaneously allowed for more children to reach higher competency levels between 2000 and 2007. While this latter result is not intended to draw attention away from the necessary efforts required to improve education quality, alternate measures such as these help one to appreciate how even in the face of lower average test scores, expanded access may still have allowed more students to reach acceptable and functional levels of literacy and numeracy than would otherwise have been the case.

2.2.2 Malawi and schooling

Malawi is among one of the poorest countries in the world with a per capita GNI of just US\$320 in 2017, and around 50 percent of its population living below the international poverty line of US\$1.90 per day (Government of Malawi 2016). From the 1990s, several strategies have been put in place by the Malawi government to significantly reduce poverty and ensure that the current consumptions of the poor do not fall below some crucial level. Additionally, policies that lift poor people out of the poverty trap through some form of wealth gain have been promoted (Mussa 2017). The more prominent of these strategies include the Poverty Alleviation Program (1994), the Malawi Poverty Reduction Strategy (2002-2005), and the more recent Malawi Growth and Development Strategy (MGDS) (2006-2001, 2011-2016 and 2017-2022) (Government of Malawi 1994, 1998, 2002, 2012, 2016).

The most recent of these, the MDGS III, is aimed at moving Malawi to a productive, competitive

and resilient nation through, among other things, sustainable economic growth and infrastructural development. One of the key priority areas stipulated in this strategy is education and skills development in which government, in coalition with its development partners, hopes to implement educational reform that can lead to both improved quality and relevance of primary, secondary, and tertiary education (Government of Malawi 2016). This is particularly important for Malawi seeing as it is a fairly young country with approximately half of its population below the age of 15, and up to 73 percent below the age of 30 (Government of Malawi 2016).

As with most countries in the region, Malawi was a British colony up until 1964 when it gained its independence led by Dr. Hastings Kamuzu Banda and his party, the Malawi Congress Party (MCP). Dr. Banda was the Prime Minister and later (first) president of Malawi from 1964 to 1994. In 1993, Malawi held a referendum where a majority voted in favour of a multiparty-led government that would end the Malawi Congress Party's 37-year monopoly on power. In 1994, in its first multiparty election, Dr. Bakilii Muluzi won the election and proceeded to be the first democratically elected president of the Republic of Malawi.

As was the norm for newly independent governments at the time, Dr Bakili Muluzi of the United Democratic Front (UDF) introduced Free Primary Education (FPE) in 1994, making Malawi the first country in the sub-Saharan Africa region to start the initiative (Chimombo 2009). The initiative was not only a major selling point during his campaign run (Kadzamira & Rose 2003, Ligomeka 2002), but also coincided with global emphasis and attention on equitable access to education.[4] As a result, total enrolment in Malawi experienced a steep increase: the gross enrolment rate (GER) grew from 93.4 percent in the 1993/94 academic year, to 134 percent in 1994/95, whereas the net enrolment rate (NER) increased from 71.4 percent to 83 percent in the same period (Chimombo 2009, Kadzamira & Rose 2003).[5]

In response to this sudden increase in demand for education the Malawi Government through the Ministry of Education recruited approximately 18 000 untrained teachers (Kunje 2002). Despite the efforts, these teacher recruitments still proved inadequate in reducing class numbers to ac-

---

[4]Part of the reason this was a major selling point for Dr. Muluzi stems from the fact that a majority of citizens were unable to send their children to school citing school fees as a prohibitive and constraining factor (for example, see Burchfield & Kadzamira (1996)).

[5]This sudden jump in enrolment rates after the introduction of FPE was not unique to Malawi but was commonplace in several other countries that had also implemented FPE see (Lewin 2009).

ceptable sizes, and also meant that a large proportion of teachers in Malawi were unqualified and lacked the experience to do so. As a direct result of this, compared to the 16 percent of teachers that were not qualified to teach in 1993/94, over 50 percent of the teachers were not qualified to teach by 1997 (Ministry of Education 1994, 1997). In addition, the abrupt implementation of FPE meant that teaching facilities such as classrooms could not be constructed to keep pace with demand which consequently led to overcrowding of available classes and an increasing number of classes that had to be conducted in the open air (Kadzamira & Rose 2003).

In an attempt to mitigate the likely negative impact on learning quality that these new and untrained teachers were going to have, the government embarked on a number of teacher training initiatives. The first of these was called the 'Malawi Integrated In-service Teacher Program' (MIITEP) which was instituted in 1997. This initiative was set up primarily to improve the quality of teaching and learning in primary schools by increasing the number of qualified teachers with demonstratable professional skills and knowledge (Milner et al. 2001). However, this program lacked the necessary support for effective implementation and as a result most of the training components were not carried out as stipulated. This ultimately led to the inability of the program to focus on the in-depth professional development of the trainees. The government then attempted different modes of teacher training. The main challenge continued to be how to provide high quality teaching training in a relatively short period of time, as the student enrolment had already expanded and it was the supply of teachers that was striving to keep up with that expansion. Eventually, the government reverted back to an older teacher training program known as the 'Initial Primary Teacher Education' (IPTE) that required a trainee to be at a Teacher Training College (TTC) for one year, followed by a year of actual teaching practice.[6] This program proved to be more successful than the MIITEP program by not only producing better prepared teachers, but also increased their annual output. In so doing, the issue of teacher shortages was (at least in part) addressed.

Despite such efforts, primary school students in Malawi continue to perform poorly in reading. This was seen quite vividly in 2010 when the government, in conjunction with the United States Agency for International Development (USAID), supported an Early Grade Reading Assessment

---

[6]Primary teacher training in Malawi has traditionally been a two year program even though the country has also had programs that were just for a year, and some for up to three years.

(EGRA) to evaluate the reading skills of primary students on a national scale. The results were overwhelming: Over 70 percent of Standard 2 students and 40 percent of Standard 4 students could not read a single word given to them to read in Chichewa. Worse still, an even larger share of students was unable to answer even one comprehension question associated with the reading (USAID 2017).

In response, the Government of Malawi's Ministry of Education, Science, and Technology (MoEST) in 2014 approved a five-year National Reading Strategy (NRS) aimed at ensuring children learn to read with understanding by the end of Standard 3 (USAID 2017). In practice, this has become referred to as the 'National Reading Program' (NRP), which is a MoEST-led program aimed at facilitating the up-scaling of early grade reading interventions for both students and teachers in Standards 1 through 4. In time, other related and complimentary programs have sprung up from the NRS, including: the Strengthening Early Grade Reading in Malawi (SEGREM) (2014-2017); the Malawi Early Grade Reading Improvement Activity (MERIT) (2015-2020), and the Girls' Empowerment through Education and Health Activity (ASPIRE) (2014-2018).

As for the structure of the education system, Malawi follows an 8-4-4 education structure; that is, 8 years of primary education, 4 years of high/secondary school, and 4 years of tertiary education. The official language of business and commerce is English, which also serves as the official language of instruction except in grades 1 to 4 where teachers are encouraged to teach in the language which the students understand best; usually a local (indigenous) language. More often than not, this is Chichewa, which is the most common of these languages and is spoken by an estimated 65 percent of the population. In fact, in the public primary education system, Chichewa, not English, is the medium of instruction in Standards 1 through 4 in all subjects except English.

At grade 4, students are expected to switch from mainly 'learning to read' to 'reading to learn', an undertaking that is not always the easiest transition to make. Secondary school education is split into two categories, junior secondary (form 1 and form 2), and senior secondary (form 3 and form 4). Progression from one level to the next is via successful completion of that phase. Primary school students sit for the Primary School Leaving Certificate Examinations (PSLCE); those that are successful are then allocated to different government secondary schools around the country. The official school starting age for primary schooling is six and goes up to thirteen. Even though

thirteen is noticeably quite high, it was determined taking into account the various difficulties and challenges students have in Malawi to start school at the young age of six.

2.2.3 Namibia and schooling

The Republic of Namibia, located on the south-west coast of Africa and bordering South Africa, Botswana and Angola, attained its independence from the former South African apartheid government in 1990. Their independence, as was common for former colonies, came after many years of violent political struggles. The country covers an area of 824 269 square kilometres and is the driest country south of the equator owing to the Namibia and Kalahari deserts.

The country's population pattern is significantly influenced by the weather patterns with areas in the northern parts of the country where rainfall averages around 700 mm per year being more densely populated than the rest of the country. For emphasis, whereas about 60 percent of the population is located in the north, less than 10 percent of the population is located in the south. Due to the rainfall in the north, crop farming is also mainly situated in that region of the country, while participation in animal husbandry is practiced throughout the country.

The education system in Namibia comprises 7 years of primary school and five years of secondary education. Primary school is split into five years of lower primary which includes grades 1 to 4 and a year of pre-primary which is intended to be a bridge between pre-school (kindergarten) and grade 1, and upper primary, which includes grades 5 to 7. Teachers are encouraged to teach grades 1 to 3 in the students' mother tongue, while from grade 4, English is the language of instruction. Secondary school education is also split into two: junior secondary (grades 8 to 10); and senior secondary (grades 11 to 12). A student is only eligible to move to the next grade if their academic performance is adequate.

As soon as Namibia gained its independence, an integral part of the government's policy direction was aimed at reducing poverty, and in so doing, amend the historical imbalances created and perpetuated in the society during the apartheid era. After the successful execution of the three-year Transitional National Development Plan (TNDP) (1991-1994), the Namibian government decided to organise its efforts around clearly defined and agreed upon medium-term development programmes. These have to date served as a roadmap to steer all of the nations development efforts

and have been in use from the 1995/1996 financial year. These are: the National Development Plan (NDP) 1 (1995/96-2000); NDP 2 (2002-2007); NDP 3 (2007-2011); NDP 4 (2012-2016/2017); and the most recent iteration of it, the NDP 5 (2017-2022). A close look at these development programmes soon conveys the fact that right from the onset, the government was squarely committed to poverty reduction through, among other things, improvements in the education sector. To this end, both NDP 1 and NDP 2 clearly acknowledged that the education sector was crucial to addressing some of Namibia's most pertinent and pressing concerns, one of which was poverty and the lack of a sufficiently trained workforce.

With this in mind, one of the main issues that quickly became of pertinence was the need to achieve the goal of equitable access to education. Therefore, in the same year that Namibia was declared independent, the newly adopted constitution guaranteed equitable access to education to every child, with Section 20 in part stating that: "All persons shall have a right to education". This article goes on to stipulate that primary education will be compulsory for all Namibian children and that students will not be allowed to leave school until they have completed this phase of school or reach the age of sixteen

Unfortunately, despite the elaborate constitutional provision for primary education, the Namibian education system continued to levy tuition (and other student-related fees such as school uniforms) on the parents and guardians, thereby nullifying in practice the strength of the constitutional provision. This coupled with the high inequality in Namibia meant that students from poorer segments of the population were still excluded from the education system despite the Namibian government's goodwill as demonstrated in the constitution's provisions for primary education.

Interestingly, as the 2015 deadline for meeting the targets of the Education for All and Millennium Development Goals approached, the Namibian government used this as motivation to speed up their implementation of universal primary education, as was expected of all United Nations member states. Although it took two decades — 1990 to 2012 — Namibia eventually launched universal primary education in 2013 and joined the list of countries which had taken firm and decisive action to ensure that by 2015, children everywhere had access to education.

However, despite this commendable progress in access to schooling, the actual quality of learning did not keep apace. Indeed, the Namibian education system remains rife with problems related to the quality of education, which itself may partly be due to the limited availability of information on Namibia's performance in international tests (so far only participating in only one international assessment system, being the SACMEQ) (Van der Berg et al. 2020). Among other reasons, what this lack of information means in practice is that education stakeholders lack crucial information they need to assess the education system as a whole, especially in comparison to other countries. Education stakeholders, therefore, find it difficult to identify key priority areas that could have significant impact on learning outcomes if addressed.

2.3   Data and methodology

2.3.1   SACMEQ data

This chapter makes use of the third wave of data collected by SACMEQ in the period beginning from 2006 and ending in 2011. SACMEQ is a consortium of education ministries in member states, policymakers and technical officials (often researchers) that are organisationally linked with UNESCO's IIEP (Kotzé & Van der Berg 2015). It aims to improve training and research needs geared towards enabling member states to enhance the quality of education they offer in their respective countries through collaborative training and policy research (Murimba 2005). Initially the consortium was only made up of seven countries[7] but that has since grown to sixteen.[8] These are: Angola, Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania (Mainland), Tanzania (Zanzibar), Uganda, Zambia and Zimbabwe.

There have so far been four education research projects facilitated by SACMEQ which are commonly referred to as "SACMEQ I" , "SACMEQ II" , "SACMEQ III" and SACMEQ IV.[9] SACMEQ is ideal for this study as it not only assesses grade six student performance in numeracy and literacy, but further aims to collect contextual household information, some of which can be used to

---

[7]These were: Kenya, Mauritius, Malawi, Namibia, Zambia, Zanzibar, and Zimbabwe. Note also that since Malawi and Namibia were member countries of the consortium right from the beginning, they have participated in all rounds of the data collection process.

[8]Though there are officially only sixteen countries in the list, the IIEP was added as a seventeenth member in acknowledgement of its long-term association with the grouping prior to the formal launching of SACMEQ.

[9]A fourth SACMEQ was conducted but official results are yet to be released.

construct an indicator of socio-economic status (SES). SACMEQ III, which included 15 member countries, collected data on approximately 61 000 Grade 6 students (and their households), 8 000 language and mathematics teachers, and 2 800 schools. This study is limited to non-missing data on student reading score information in Malawi and Namibia that altogether consists of 2 781 students at 139 schools in Malawi, and 6 938 students at 267 schools in Namibia. Data collection for the SACMEQ III projects employed both explicit and implicit stratification procedures. For the explicit stratification, the variable "Region" was applied by separating each sampling frame into separate regional lists prior to undertaking the sampling. This was done in response to the representative ministries' of Education desire to have sample estimates of the population whose characteristics resembled accurately those of each region (Ross & Saito 2008). The implicit stratification variable was "school size" and this aspect of the exercise employed a two-stage survey design. Firstly, to ensure some control over the total sample size as would not be the case if the primary sampling unit (i.e. schools) were selected with equal probability at the first stage, schools were instead stratified according to size by employing the probability proportional to size (PPS) approach. The second stage then involved the selection of a simple random sample of a fixed number of students within the sampled schools using randomly generated numbers.

2.3.2 Measures

2.3.2.1 *Dependent variable*

The dependent variable used was the SACMEQ reading score. Information on students' reading scores was analysed using Rasch Modelling (Moloi & Strauss 2005) that performs a linear weighted transformation of test scores into a scale with a predetermined international mean of 500 and a standard deviation of 100.[10] Based on these standardised scores for literacy, Malawi performed the worst out of all member countries with a score of 433.5, and Zambia performing second-worst with a score of 434.4, whereas Namibia's performance was more than half a standard deviation higher at a mean score of 496.9.

---

[10]Specifically, sampling weights that allow the sample of each country to be treated as if they were the same size are applied (Moloi & Strauss 2005).

2.3.2.2   Socio-economic status (SES)

While there is no universally agreed-upon conceptual meaning of SES (Bornstein & Bradley 2002), this chapter follows the standard definition that conceptualises a family's SES as their relative position in some hierarchical social structure determined by their ownership of or access to such things as wealth, prestige, assets and influence (Mueller & Parcel 1981). One advantage of such a broad definition is that it allows for easily measurable units, such as a household and/or individual's level of income, to proxy for SES, while at the same time allowing for broader index-type measures to be constructed for the same purpose.

Historically, SES has predominantly been measured using household responses collected from survey questionnaires on information pertaining to household income and/or consumption. Though researchers are well aware of the limitations of such measures, the absence of a better measure has meant that this approach has remained the preferred one, as proven by its continued use in the literature on issues concerned with a household's socio-economic status/rank.[11] While there remains an enduring debate as to which of the two, income or consumption, makes for a better indicator or proxy of household wealth, the latter is often preferred and regarded by many as "the best measure of the economic component of living standards" (Deaton & Grosh 2000, p.95).

This is especially true in LICs, where a significant segment of the population is classified as living in rural areas, and consumption data is thought to be both easier to collect and more reliable than income data (Filmer & Pritchett 2001). Notwithstanding this, LICs pose a unique challenge to data-collecting agencies. This is because in addition to the common problems of recall and refusal of respondents to be forthcoming with questionnaire responses, the extensive amount of data collection needed to achieve reliable responses tends to be time-consuming and, therefore, costly. Given such constraints, there arose a need to find alternative methods for measuring SES.

One such alternative is an asset-based approach that uses household asset information to proxy for wealth. This approach has especially grown in popularity due to the increasing number of household surveys that either lack or poorly measure information on income or consumption, but instead include questions aimed at collecting important asset ownership information (Wittenberg

---

[11]See for example Howe et al. (2008).

2009). Asset-related questions often include areas such as: a household's ownership of a range of durable assets (e.g. car, bicycle, television screen); housing features (e.g. wall and floor building material); and access to basic services and utilities (e.g. water and electricity).

As with the debate on usage of either household income or consumption data, not all researchers agree on both the usefulness and accuracy of using asset data to act as a proxy for household wealth. However, those in favour argue that asset data is more reliable than income or consumption data, since asset questions are not only more direct, but there is also potential for the interviewer to validate the responses through their own observation of the surroundings. This significantly minimises the risk of measurement error that often biases and plagues other forms of data collection (Filmer & Pritchett 2001, Sahn & Stifel 2003).[12] For example, it would be harder for an interviewer to verify a household's monthly income than it would be to verify the pieces of furniture they own. Further, it has been shown that asset-based approaches are an improvement over income or consumption expenditure, since they better reflect the long-run welfare of households (Wittenberg 2009).

On this point, an opposing view holds that asset indices fail to respond, and therefore capture, the impact on a household's income of potential short-run economic shocks (Filmer & Pritchett 2001). To illustrate, consider a study investigating the impact of seasonal income shocks through grants on household expenditure patterns. In such a case, this sudden shock in income may not immediately reflect in the composition of assets owned by the household, and therefore bias the findings of such a study if the household's SES was constructed based on its asset ownership.

An additional critique of wealth indices based on asset ownership is that they are a mere listing (usually of a binary 'Yes/No' nature) of assets present in the house, and say little to nothing of their quantity and/or quality. Collecting information on vehicle ownership alone, then, would not differentiate a household that owns a more recent make and model of a car from a household that merely received a run-down, lower quality car donated by a more affluent (often urban dwelling) relative. A mere listing would only tell you that both these households own a car, which by itself says very little of their relative wealth.[13] Surprisingly, whereas it is expected that this would alter

---

[12]This assertion has however been questioned by Onwujekwe et al. (2006).

[13]Related to this point is an issue of conspicuous consumption which by definition could make wealth indices a problematic measure at the upper end. This is because households who only own a certain expensive asset in at

a household's wealth ranking, a study conducted in a significantly large number of countries by Falkingham & Namazie (2002) shows that this is not the case.

Related to this, an additional concern that arises when using assets is that depending on the subgroup in question, the same asset may imply different rankings (Kotzé & Van der Berg 2015, Vyas & Kumaranayake 2006). This simply means that access to electricity, for example, may be more a sign of wealth in a rural area than it would be in an urban area, simply because it is more a product of geographical location than an indicator of individual wealth for the individual who lives in an urban area. In a typical setting, it is more likely then, that the asset index would rank a household with access to electricity in a rural area among the wealthy, while not doing the same for a household with access to electricity among those who live in urban areas. If one did not take these area-specific idiosyncrasies into account when constructing an asset index, it is possible to arrive at an erroneous conclusion with regard to an asset's ability to determine socio-economic rank and proxy for wealth.

A final point to consider is how the researcher ought to aggregate the asset variables into a unidimensional measure of SES that is able to uniquely differentiate households with a different mix of assets (Vyas & Kumaranayake 2006). As already alluded to, this is necessary because a single variable (such as access to electricity) is not sufficient to differentiate and determine household SES. As an example, consider an approach that merely sums the number of assets a household owns, with the crude conclusion that a household with a higher number of (even inferior or poorer quality) assets is from a wealthier segment of the population than a household with fewer (more superior or expensive) assets. Such an erroneous conclusion is, in this case, possible because this approach assumes and applies equal weights to all items in the asset list.[14]

Filmer & Pritchett (2001), therefore, propose — as has become a norm in the literature — using the technique of Principal Component Analysis (PCA) to assign weights to assets more systematically. Crucially, Filmer & Pritchett (2001) showed that weights for a wealth index determined through PCA make for a better proxy of actual wealth in the population than alternative approaches at-

---

attempt to fit in will be up-weighted and classified with the wealthy, even when they are not.

[14]However not all asset-based approaches assign equal weights. See Bollen et al. (1995), Montgomery et al. (2000) and Setel et al. (2003) for more comprehensive discussions of alternative methods of weight assignment.

tempting to do the same. Unsurprisingly, not all researchers agree.[15] Despite this, PCA continues to gain traction, aided perhaps in part by the fact that even renowned research institutions such as the World Bank use this same approach in some of their analyses of inequality, particularly in their Demographic and Health Surveys (Howe et al. 2008).

Therefore, a unique SES index was derived for both countries by performing a Principal Component Analysis (PCA) on an identical set of 31 'Yes-No' responses to questions on household ownership of items.[16] The technical Note in Appendix A.1 provides further detail on the application of PCA, and how it is able to provide a single, continuous index of SES. Following Willms et al. (2006), the SES index was scaled to have a mean of zero and a standard deviation of 100.

2.3.3 Methodology

2.3.3.1 *Socio-economic Gradient*

A socio-economic gradient graphically depicts the relationship between any social outcome of interest and the socio-economic status of individuals within a particular community. The social outcome can be any measurable attribute such as health, education, marital status or likelihood of spending time in prison. Furthermore, the outcome can be—depending on the question of interest—either continuous or dichotomous. The term 'gradient' is intended to point towards a relationship that is both gradual and increasing (or decreasing) across the range of SES (Adler et al. 1994, Willms 2003). Some scholars have conceptualised this relationship between SES and any social outcome slightly differently, opting rather to refer to it as the socio-economic-gap. In doing so, they emphasise the gap that exists between and among stratified communities (Caro et al. 2009).

Though impressively insightful, the technique used to derive SES gradients is straightforward. One carries out an Ordinary Least Squares (OLS) linear regression of the test scores on the socio-economic status variable for each community of interest. This produces the following simple

---

[15]See for example Setel et al. (2003) and Bollen et al. (1995).

[16]The 31 items were: daily newspaper, weekly or monthly magazine, clock, piped water, bore hole, table to write on, bed, private study area, bicycle, donkey/horse cart, car, motorcycle, tractor, electricity (mains, generator, solar), refrigerator/freezer, air-conditioner, electric fan, washing machine, vacuum cleaner, computer, internet, radio, TV, VCR player, DVD player, CD player, audio-cassette player, camera, digital camera, video camera, telephone/cell-phone (from Question 14 in SACMEQ III Student Questionnaire).

estimated model:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 * SES_i \qquad (2.1)$$

where $\hat{Y}_i$ is the predicted outcome for student i , $\hat{\beta}_0$ is the intercept, and $\hat{\beta}_1$ measures the impact of a unit (in this case, one standard deviation) increase in student SES, and therefore represents the slope of the SES gradient line.

According to Willms (2004), the socio-economic gradient comprises three components that provide information about the relationship between SES and the social outcome of interest: the level/intercept; the slope/SES coefficient; and the strength of the relationship. Since the SES variable has been transformed to have a zero mean, the level of the gradient is the expected performance of a student with an average socio-economic status. In comparison to the typical student mean score, Ross & Zuze (2004) refer to the intercept as the 'adjusted quality' — referring specifically to the fact that an average student test score has been adjusted for socio-economic status. This term is otherwise referred to as the height of the gradient line. In the context of Malawi and Namibia therefore, even prior to performing any analysis, the overwhelming evidence in favour of a positive correlation between higher SES and student performance (Sirin 2005) points to students in Namibia outperforming their Malawian counterparts by virtue of their being relatively wealthier than Malawi.

The second of these components—the slope of the gradient—represents the degree of inequality attributable to SES; that is, the steeper the gradient, the higher the level of inequality in outcomes between low and high SES individuals, and the milder the gradient, the lower the level of inequality. An equally related and important aspect of the SES gradient line and how it relates to student inequality in terms of performance is its length. Following Willms et al. (2006), this is best captured by a gradient line from the 5th percentile value of the SES index to the 95th percentile value of the SES index. Therefore, this measure encompasses the SES range of the middle 90 percent of students in each country.[17] The length of the line indicates how dispersed and spread the student population is in terms of SES, making it possible to assess how differently the most- and least-

---

[17] For an alternative approach to calculating the length of the line, see Ross & Zuze (2004).

advantaged students perform. A longer line represents a more socio-economically heterogeneous and diverse student population, meaning that education systems have to deal with students from a wider range of socio-economic backgrounds. In contrast, a relatively shorter line signifies a narrower distribution of socio-economic backgrounds making it a more homogenous population and thus relatively easier for education systems to cope with.

The final component—the strength of the socio-economic status gradient—refers to how much of the observed variation in individual student test scores can be explained by student socio-economic status. The most common measure of the strength of the relationship is the coefficient of determination, more commonly referred to as the r-squared statistic. A strong relationship, represented by a large r-squared value, implies that variation in socio-economic status accounts for a considerable amount of the variation in academic performance. On the other hand, a weak relationship, represented by a lower r-squared value, would imply that other variables exist that would better account for the observed differences in student outcomes.

2.3.3.2 *Hierarchical Linear Modelling*

Hierarchical linear models (HLM), otherwise known as multilevel models, are characterised by complex patterns of variability that typically take on a nested structure. This is indeed the case with the SACMEQ data where students are nested within schools, implying that one could make use of a two-level HLM (Raudenbush et al. 2004, Lee & Zuze 2011, Raudenbush & Bryk 2001). Socio-economic gradient lines based on models such as that given by equation 2.1 assign a single intercept and slope to all individuals in the sample data, irrespective of the school attended. It is apparent, however, that the relationship between SES and performance could differ significantly by school. This could be owing to several reasons, some of which include: teacher pedagogy, teaching and/or learning resources available to the school, school leadership via the principal, and school entrance exams which work by ensuring the school only allows students with a perceived predetermined level of ability. Also, parent involvement tends to be higher amongst higher SES households. So a higher density of high SES households in a given school may add to accountability.

HLMs are at times referred to as mixed-effects models as they are characterised by two effects:

29

fixed effects, which are the intercept and slope describing the relationships in the full sample as in equation 2.1; and random effects which are intercepts and slopes that are unique to subgroups (schools). Crucially, what this means in practice is that one is able to decompose the student reading performance into its within- and between-school components. To put it differently, for the within-school component, this is somewhat akin to one performing individual regressions within/for each school to determine how a student within the school itself differs from another student in terms of reading performance as you would a standard regression.[18] This is the first step/level. In the second step, HLMs help one to identify what proportion of overall student variance in reading performance lies systematically between schools. This is sometimes referred to as the intraclass correlation (ICC) denoted by rho (ρ).

As a starting point, consider the level-1 equation represented by:

$$Y_{ij} = \beta_{0j} + \beta_{1j} * SES_{ij} + e_{ij} \tag{2.2}$$

where Y and SES are the outcome and SES of student i in school j, respectively. Here, β_{0j} represents the average student score in school j and β_{1j} the SES gradient of school j. The level-1 error term, e_{ij} indicates how much a student's score deviates from the mean score of the school they attend.

Building of this, the intercept term β_{0j} can be considered as comprised of a grand mean—the mean across all students and schools—y_{00} , plus a random error term, u_{0j} , which provides for the level-2 equation:

$$\beta_{0j} = y_{00} + u_{0j} \tag{2.3}$$

The error term u_{0j} , therefore, indicates how the mean score of a school deviates from the grand

---

[18]The word somewhat here is used intentionally because the predicted intercepts and slopes from an HLM would not actually be identical to the individually and separately run ordinary regressions using standard OLS. In summary, the reason for this would be because the regression coefficients in the HLM are shrunk back towards the mean coefficients for the whole data set using empirical bayes (EB) estimation. In other words, the multilevel estimates are weighted and because those from the individual OLS regression are not, the results from the separate analysis would be more variable. For a more thorough discussion, see Hox et al. (2017).

mean across all schools. In the same way, the slope term, β_{1j}, can itself be considered as comprised of a grand mean term and random error term, providing for a second level-2 equation:

$$\beta_{1j} = y_{10} + u_{1j} \tag{2.4}$$

Replacing each of 2.3 and 2.4 into equation 2.2 produces the final mixed-effects model:

$$Y_{ij} = y_{00} + y_{10} * SES_{ij} + u_{0j} + u_{1j} * SES_{ij} + e_{ij} \tag{2.5}$$

All estimation takes into account the complex survey design.

2.4 Empirical Results

2.4.1 Main findings

The results from estimating equation 2.1 are presented in Table 2.1. Before interpreting the results, it is worth noting that SES rankings are not uniform across the two countries, but are unique to Malawi and Namibia. For example, the median SES of Namibia might map to a higher position, say, the 75th percentile, in the SES distribution of Malawi, whilst the median SES of Malawi might map to a position that is well below the median of Namibia. However, relative to their country's population, individuals at the median SES would hold the same position; that is, 50 percent of the population would lie either side of this SES value.

With this in mind, the values in column (3) of Table 2.1 can be interpreted as the associated increase in the expected reading score given a one standard deviation increase in SES. For Malawi, this is estimated to be 8 points, while a similar one standard deviation increase in SES is associated with a 40-point increase in expected reading scores of Namibian students. It is rather unsurprising that, given the historic background of Namibia, the SES slope is steeper than that of Malawi: each standard deviation increase in SES is associated with a higher return on performance for students in Namibia than is the case for a student in Malawi, where moving up the wealth (asset ownership) ladder is associated with a relatively smaller return.

The value of the level of the gradient, shown in column (2), indicates that the expected test score for a student with average SES is lower in Malawi than in Namibia. It is worth noting that the adjusted and unadjusted measures of average school performance are similar. Lastly, the strength of the relationship between the SES and the student outcome (column (4)) is higher for Namibia than for Malawi. This means that socio-economic status accounts for a larger variation in learning outcomes and student performance in Namibia than in Malawi. The very low r-squared statistic in the case of Malawi could be expected given the low SES gradient that is indicative of an education system that offers more equal opportunities to its students, regardless of SES background. While this is in no way meant to excuse the low performance in the Malawian education system, it is hoped that as the system improves–leading to better student and school performance–that this equality will not be lost: allowing for a system where everyone has an opportunity to thrive irrespective of their socio-economic background. Indeed, equality in the absence of excellence is an undesirable outcome (Lee & Zuze 2011).

As pertains to the length of the lines, the 5th and 95th percentiles of the SES scores are -0.83 and 2.4 respectively for Malawi, making the line 3.23 points long. On the other hand, the 5th and 95th percentile SES scores for Namibian students are -1.24 and 1.91 respectively, making the line length 3.15. Therefore, 90 percent of Malawian and Namibian students fall in these ranges. Interestingly, the line lengths are only marginally different. However, by looking at the corresponding range of predicted scores of these middle 90 percent of students, one sees that the education system in Malawi is more equitable than that of Namibia. Indeed, while a less advantaged (5th percentile) student in Malawi on average is expected to achieve a reading score of 427, a student of comparably low (relative to the country) SES from Namibia can expect to get a reading score of 451.

As for the more advantaged Malawian student, the expected reading score of 453 is substantially lower than their Namibian counterpart whose expected reading score is 576. This indicates that despite similar line lengths, there are larger differences (126 versus 26 points) in achievement between wealthier and poorer students in Namibia than in Malawi. The same conclusion can be drawn by taking the product of the length of the line and the slope of the line, as done in column (6) of Table 2.1. The flatter slope in the case of Malawi implies that performance is not distributed

differentially across SES, whilst in Namibia it is.

Table 2.1: Traditional and alternative views of performance for Malawi and Namibia

Country	Traditional view	Alternative View				
	(1)	(2)	(3)	(4)	(5)	(6)
		Quality	Social equity	Strength	Distributional equity	
	Mean (unadjusted) reading score	Line height ($\hat{\beta}_0$)	Line slope ($\hat{\beta}_1$)	R-squared	Line length	Length ∗ Slope
Malawi	434	433	8	0.023	3.23	26
Namibia	497	500	40	0.212	3.15	126

Notes: Own calculations from SACMEQ III

Though equitable education is a much sought after feature of education systems, it is evident that the bulk of the Malawian student population performs below the SACMEQ average. This brings to the fore the need for what (Willms et al. 2006, p.11) refers to as "raising and levelling the bar" , where the term 'learning bar' is synonymous to the socio-economic gradient line. He argues that the learning bars of most countries need either improvements in educational performance (i.e. raising the bar), reductions in inequalities among students from different socio-economic backgrounds (i.e. levelling the bar), or both. Based on this understanding, Table 2.1 suggests that while Malawi is in more need of raising its learning bar, Namibia has a more pressing need to level theirs. This can further be seen graphically in Figure 2.1 where Malawi's gradient line appears more flatter and yet firmly below the SACMEQ average score of 500 for the entire range of SES.
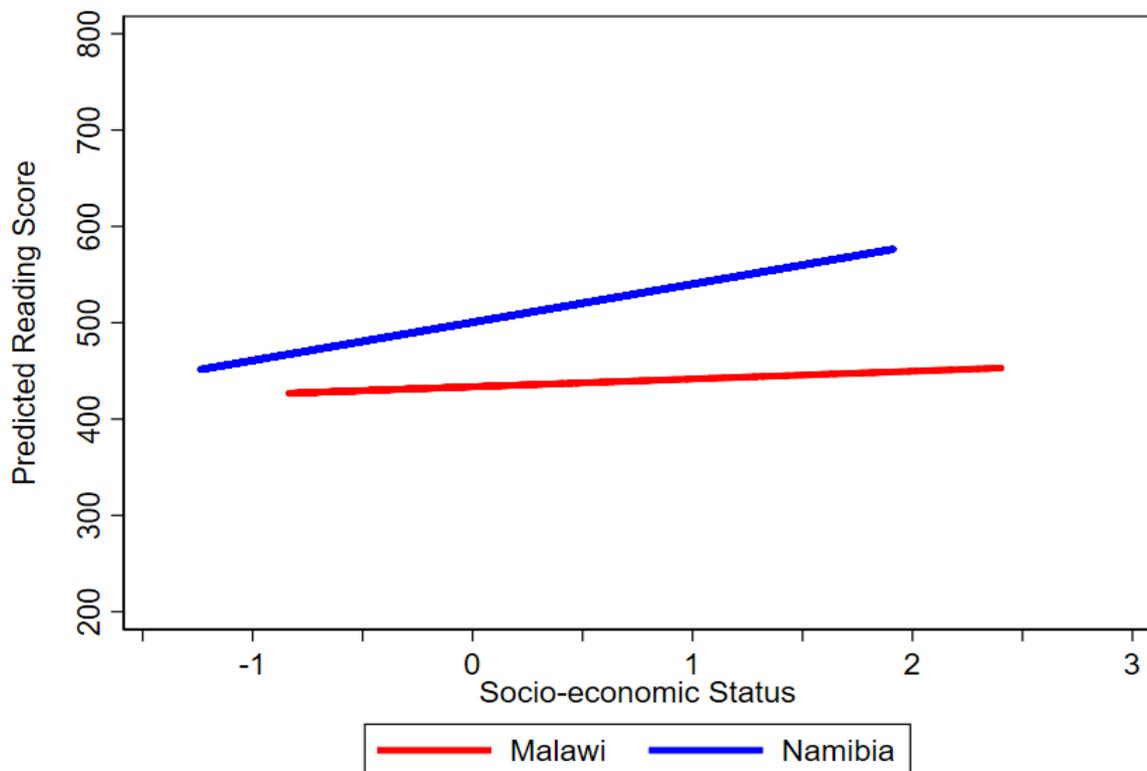
Figure 2.1: Socio-economic gradient lines for Malawi and Namibia drawn from the 5th to the 95th percentile value of SES score for each country

To put these learning bars into broader regional context, Figure 2.2 represents the SES gradients of Malawi and Namibia in comparison to those of the other SACMEQ member countries. One notices that, even among this wider selection of countries, Malawi does indeed have one of the flattest SES gradient lines and Namibia one of the steepest within the SACMEQ member countries.
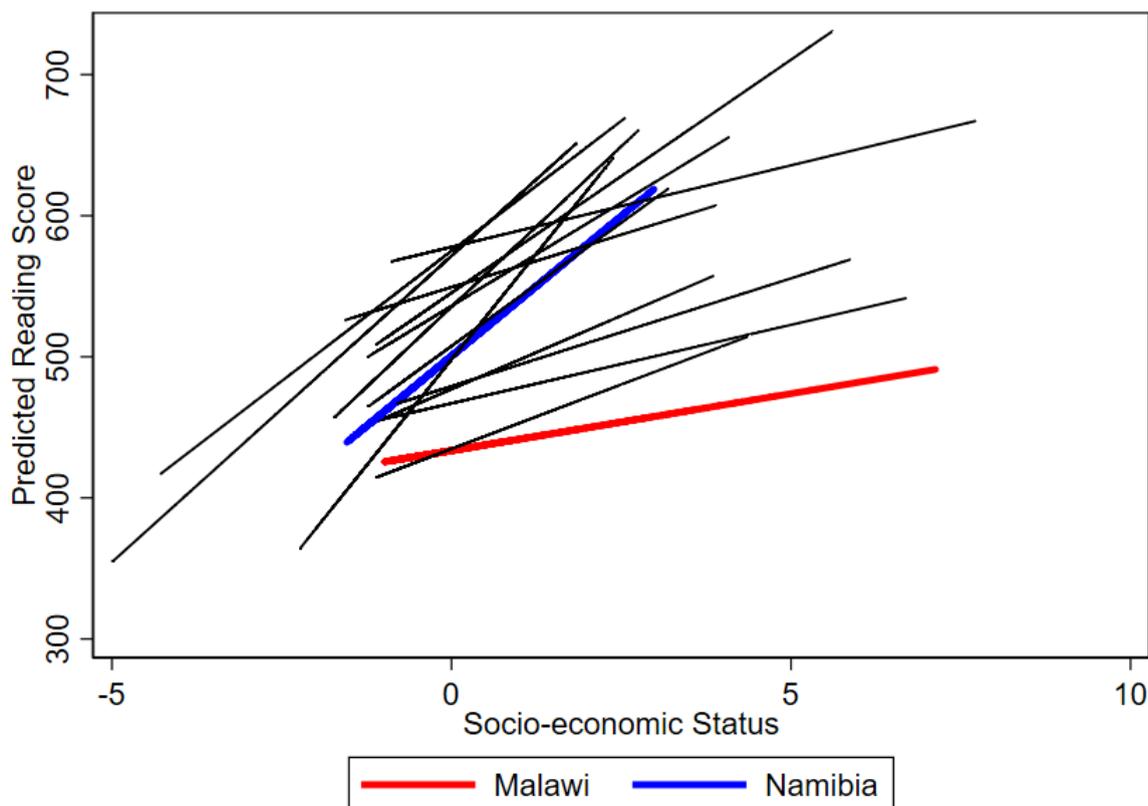
Figure 2.2: Socio-economic gradient lines for Malawi, Namibia, and all other SACMEQ member countries

2.4.2 SES slopes by school

The socio-economic gradient lines obtained from the HL models discussed in equations 2.2 through 2.5 are presented in Figure 2.3. Separate gradient lines are estimated and plotted for each school: 139 in total for Malawi, and 267 for Namibia. For additional context, the overall gradient lines presented in Figure 1 for Malawi and Namibia are superimposed on the new school-specific gradient lines. In doing so, it is possible to demonstrate that overall, students within each school perform quite differently depending on the school they attend than what would derive by merely looking at the overall SES gradient line discussed up until this point.
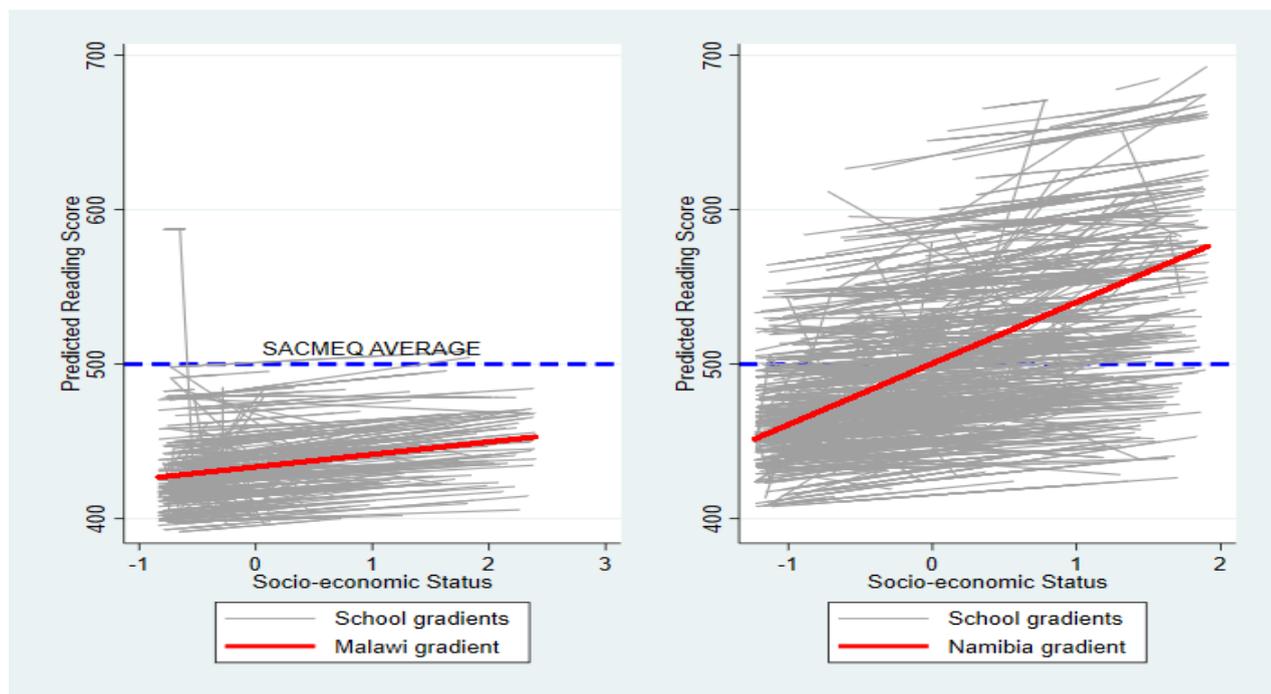
Figure 2.3: Socio-economic gradient lines for Malawian and Namibian schools

From Figure 2.3, it is clear that students are differentiated by school, both in terms of reading score performance and the slope of their gradient. This is true for students in both Malawi and Namibia. Interestingly, unlike Namibia, some Malawian schools appear to reveal negatively sloped gradient lines (one in particular has an almost vertical gradient line).[19] Again, this gives credence to the benefit of analysing data that is of a nested nature using models that are specifically designed to account for those data complexities. Furthermore, it is clear that the majority of schools in Malawi perform below the SACMEQ average reading score of 500. The story is strikingly different in Namibia where a sizeable proportion of schools have students that perform well above the SACMEQ average reading score. Again, given the aforementioned context Malawian schools function in, this is not entirely surprising.

---

[19] Further analysis shows that the vertical gradient line represents a unique school with relatively poor students who surprisingly perform really well on the reading tests.

2.4.3 SES slopes by urban and rural sub-samples

In addition to disaggregating student performance by school for each of the two countries, it is additionally worth doing the same for rural and urban sub-samples. This is mainly for two reasons. Firstly, following the rationale discussed concerning HLM, it is likely that the data generating process (DGP) differs by rural and urban sub-samples. A 'national' socio-economic gradient, as discussed up until this point, likely masks and glosses over the idiosyncrasies that characterise and differentiate students in rural areas from those in urban areas. For students living in rural areas, typically made up of poorer individuals in the population, SES may have a different level of impact on learning outcomes and performance than students who live in relatively wealthier urban areas. In this way, one would therefore expect the urban SES gradient to be steeper than the rural gradient as the wealth profile in rural areas would be more homogenous. Another way of looking at it is merely to consider the population size; specifically, one has to consider that in low-income countries, the rural population accounts for the larger proportion of the population. In the case of Malawi, that would be 85 percent (National Statistical Office 2012).[20] That would mean that both the slope and levels of a national socio-economic gradient would disproportionately be influenced by the sheer size of the rural population, and so mask the intricacies of urban areas.

The second motivation for analysing SES gradients within sub-samples is that, as has been shown elsewhere, constructing an SES variable using principal component analysis in low- and middle-income countries runs the risk of having a variable that is too urban in its construction,[21] and additionally one that is not able to distinguish the poorest individuals from other poor households (Rutstein 2008). This second weakness of the SES variable implies that the SES variable performs poorly in less well-to-do segments of the population in clearly separating the ultra-poor from the mildly poor.[22] One proposed remedy for this is to construct the wealth index at community level

---

[20]The more recent NSO survey reports a lower rural percentage of 81 percent, signifying more rural-urban migration (National Statistical Office 2017). The reason for instead citing the earlier version of results is to remain consistent with the SACMEQ survey itself conducted around 2010.

[21]In addition, accounting for the urban/rural sub-samples may partly be taking into account differences in household composition whose make-up is likely different in urban relative to rural areas. This will form the focus of the next chapter of this thesis.

[22]Weakness of the SES variable constructed via PCA and recommendations of how to overcome them have well been documented and expertly discussed in the literature. See for example Howe et al. (2008), Rutstein (2008).

such as at urban and rural as opposed to at national level (Rutstein 2008, Vyas & Kumaranayake 2006). The results are presented in Table 2.2 below.

Table 2.2: Estimates and regression statistics for socio-economic gradients by community in Malawi and Namibia

	Malawi		Namibia	
	Urban	Rural	Urban	Rural
Line height	449	429	548	465
Line slope	8	4	34	5
p-value	0	0.033	0	0
R2	0.0213	0.0058	0.1364	0.0065
Observations	662	2119	2846	3552

As expected, the levels of the SES gradients in the urban communities are higher relative to the rural communities. This is the case in both Malawi and Namibia. Furthermore, socio-economic status appears to explain variation in student performance in urban Namibia only. That means in the full sample analysis, the high r-squared was most likely driven by students in urban as opposed to rural Namibia. This not only reaffirms the fact that the SES variable may in fact be too urban in its construction as previously mentioned, but also that the unobservable factors correlated with this urban SES have a stronger influence on reading performance in urban relative to rural areas. As stated earlier, one very striking observation is that the levels of both the urban and rural communities in Malawi are lower than even the rural level of the SES gradient for Namibia. What this means is that after accounting for SES, the average student from both urban and rural Malawi does not perform as well as a student from rural Namibia.

Further, since Malawi's level of inequality is far lower than that of Namibia, it makes intuitive sense that the gap in performance between Malawi's urban and rural population is lower than that of Namibia's urban and rural population. In addition, it has been documented that urban areas in Namibia are significantly more unequal than rural areas which may explain the difference in SES gradient between urban and rural Namibia (Namibia Statistics Agency & World Bank 2017). Though in the same country, a one percent increase in SES yields two different kinds of leaps in performance. On the one hand, one only sees their expected performance increase by 5 points, while a similar standard deviation increase in SES yields a 34 point increase in expected reading

score in urban areas.

2.4.4   Local polynomial smoothing

One of the most common drawbacks of analysing the relationship between student SES and performance using the socio-economic gradient line as discussed up until this point, is that it assumes and imposes (parametrically) a linear relationship between the two. There is, however, no reason to expect that the underlying relationship between SES and reading performance is linear. Using non-parametric techniques such as a kernel-weighted local polynomial regression is one approach of allowing the data to 'speak for itself' and show the 'true' mapping of SES and student performance for Malawi and Namibia.

The outcome of this exercise is presented in Figure 2.4. The shape of the relationship between student performance and SES appears to be non-linear for both Malawian and Namibian students. As was seen with the OLS analysis, the slope for Malawi is relatively flat and increases only gradually along the full range of the SES. The slope for Namibian students, on the other hand, becomes steeper towards the higher end of the socio-economic spectrum. Additionally, the widening of the 95 percent confidence interval band towards the top end of the Malawian student SES distribution indicates the presence of small samples beyond an SES value of 3. An overlaid scatterplot drawn for Malawian students with an SES score of 4 or higher indicates that only 20 students lie to the right of that value. As an aside, this gives credence to limiting the line length to the middle 90 percent of the population, since the very top of the SES distribution is prone to outliers.

At every point along the SES axis, it is noticed that Namibian students outperform their Malawian counterparts. This observation for Malawi is disheartening and certainly warrants further investigation into what other channels can be isolated and improved upon that may potentially better affect educational attainment in Malawi other than or in addition to SES. Alternatively, one may also be interested to find out what system-wide blockages or hindrances exist in the education system as a whole that are dampening and almost rendering the socio-economic status of a student to be inconsequential to academic performance. As mentioned earlier, some of these may be the lack of educated teachers, overcrowding, and an above average teacher-to-student ratio.
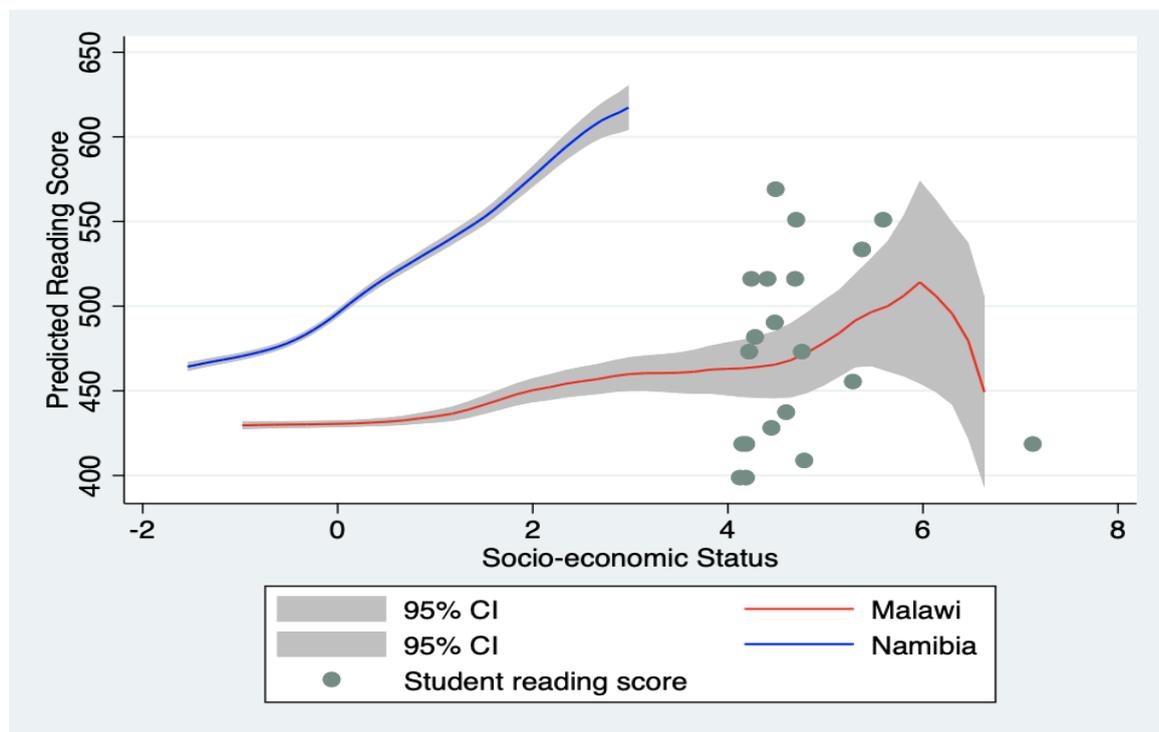
Figure 2.4: Local polynomial regression of student reading score as a function of socio-economic background for Malawi and Namibia.

2.4.5 Disaggregating the socio-economic gradient by quintile

While SES gradients offer insightful information on how student performance differs by SES, an alternative approach towards describing the same relationship is to show the distribution of reading performance for students from different wealth groups. Typically, this is achieved by plotting kernel densities of students grouped by quintile of SES: poorest to richest, as in Figure 2.5 and 2.6 for Namibia and Malawi respectively.
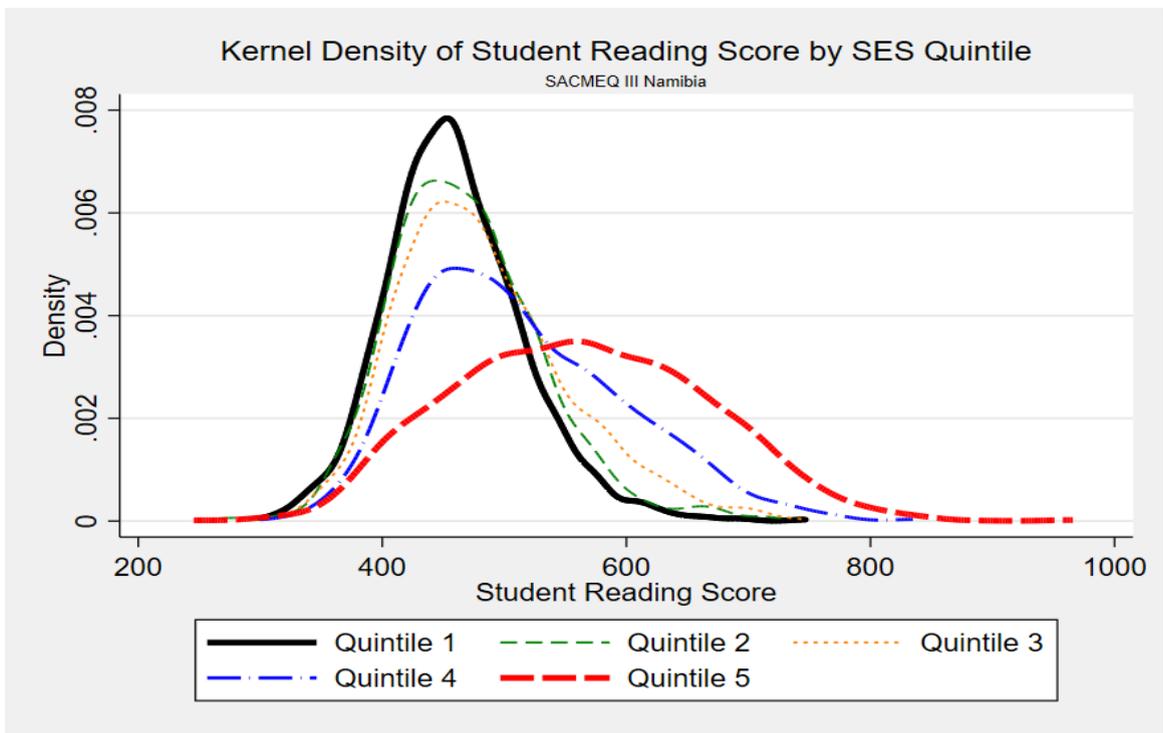
Figure 2.5: Kernel density of student reading score by SES quintile in Namibia
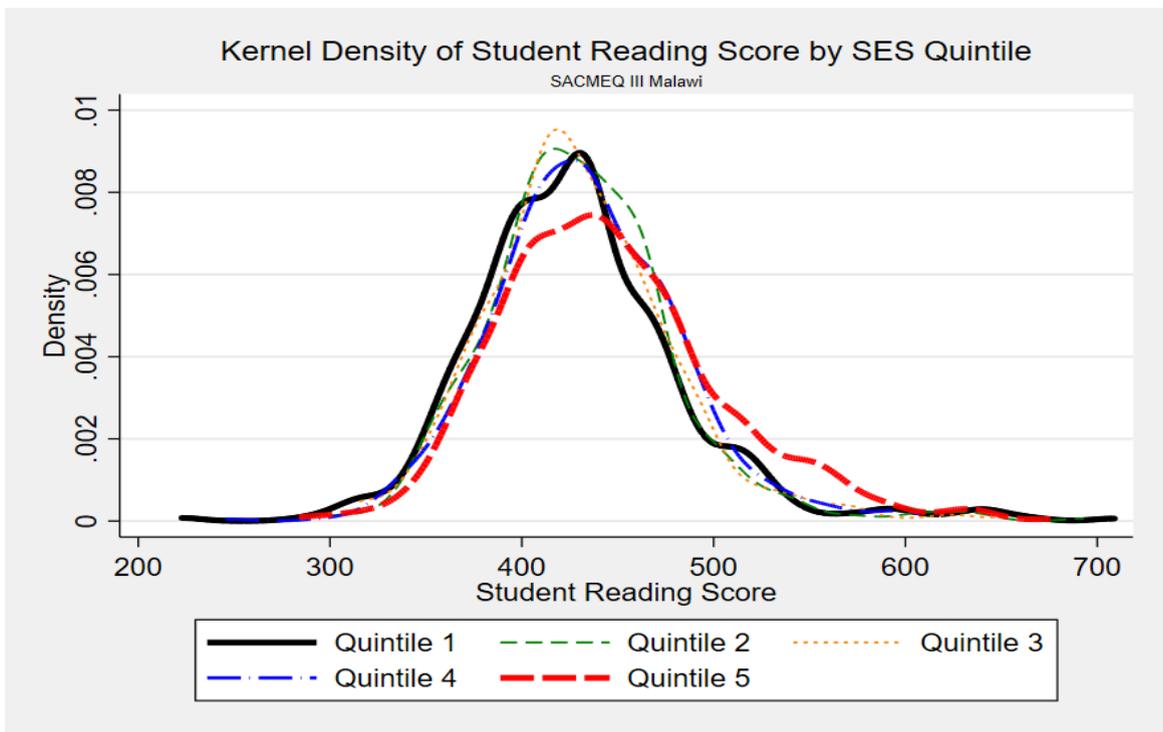
Figure 2.6: Kernel density of student reading score by SES quintile in Malawi

Interestingly, Figure 2.6 for Malawi reinforces the fact that there appears to be little to no advantage of belonging to a wealthier segment of the population. This can be seen by the clear overlapping of the kernel densities for all five SES groups. In fact, even the majority of students from the fifth quintile, representing the richest 20 percent of the population, score below the SACMEQ regional average of 500. Namibia paints a different picture. First, it also reflects accurately and reaffirms the findings from the local polynomial smoothing approach: a steady upwardly increasing relationship between socio-economic status and student performance for the larger part of the SES range, but a steeper slope for the wealthier students.

In fact, the Namibian kernel densities show a top quintile that has a different distribution, characterised by a significant proportion of those students performing well above the SACMEQ average of 500. This resembles a phenomenon researchers use to describe a similar and more pronounced feature of the South African education system referred to as "bimodality" (Fleisch 2008, van der Berg 2008, Spaull 2013). That is, the existence of what appears to be two separate distributions concealed in one overall distribution. In Namibia, quintiles 1 to 4 representing 80 percent of the student population make up that first distribution, while the second is made up of the wealthiest 20 percent of students. Indeed, van der Berg (2008) maintains that each of these two 'subsystems' make up distinct functioning parts of the school system as a whole, and should – in the South African case – be treated with sensitivity as they have significant underlying structural differences. This appears to be the case for Namibia as well.

2.5   Conclusion

This chapter carried out a comparative analysis of the relationship between students' reading performance and their socio-economic background using the third wave of SACMEQ test score data for Malawi and Namibia. Evidence from similar research suggests that this relationship is typically monotonically increasing: in that as students become wealthier, their performance relative to their less advantaged counterparts improves. The mechanisms through which these performance improvements occur was beyond the scope of this study. However, several important findings emerged.

Firstly, while socio-economic background is positively and significantly associated with student

performance in both Malawi and Namibia, it is Namibian students who see higher returns to their literacy performance. In Malawi, academic performance is relatively homogenous regardless of wealth status and seems unresponsive to improvements in socio-economic background. This is reflected in two ways: a small slope and r-squared value for the Malawian socio-economic gradient relative to those of Namibia where socio-economic background explains about 21 percent of the variation in student achievement. The result for Malawi is rather disheartening for education stakeholders such as parents who aspire to upward mobility in the hopes that that would give their child the best chances at performing well. This particular finding goes against the bulk of the literature on the impact of SES on student performance (Caro et al. 2009).

Conducting the same analysis on rural and urban sub-samples in both countries uncovers an interesting finding for Namibia. Socio-economic background appears to only explain variation in student achievement in urban areas, where most of the wealthier students are likely to live. This point is strengthened when one observes the kernel densities of student scores in Namibia, where it is clear that performance is bimodally distributed. That is, there appear to be two distributions: one comprising the bottom 80 percent of the population, and the other consisting of the richest 20 percent of the population, with the latter group performing better than the rest of the other students in the population.

Furthermore, the gradient for Namibia is above the gradient of Malawi along the entire range of SES values, and it has the majority of students (and schools) performing above the SACMEQ average of 500. In addition, even after adjusting for socio-economic status, there is not a big difference between the adjusted and unadjusted mean scores for students in either country. Differences in mean scores emerge when HLM analysis is carried out which show that certain school-specific characteristics and features may affect the extent to which socio-economic background influences performance, more especially in Namibia. That is, depending on the school they attend, the expected performance of students and SES gradient can differ significantly. Unfortunately, for Malawi, this still did not appear to make a difference, since almost all schools perform below the SACMEQ average.

A crucial undertone seen in this chapter, and discussed more in-depth in the next, is that there are inherent differences within sub-samples of any larger sample. More importantly, this het-

erogeneity within the sample is characterised by different data generating processes that require additional care before one makes sample- (and population-) wide generalisations. In the current context, failure to do so could, for example, leave one thinking that SES explains about 20 percent of student variation in Namibia when in reality this statistic is predominantly driven inter-spatially. The knowledge that we are dealing with two very different kinds of school systems—one catering to the relatively wealthy, where SES explains a significant variation in student performance, and another that requires different kinds of interventions to improve the performance of its students—has the potential to affect policy conclusions.

CHAPTER 3

Considerations in Wealth Index Construction in Low-income Countries: A Finite Mixture Model Approach

3.1 Introduction

Since the seminal findings from the Coleman Report (1968) insisted on the importance of a student's socio-economic status (SES) over and above other aspects of the learning process, governments and communities have both come to agree that a student's financial well-being plays a critical role in determining educational outcomes. Stemming from this, a voluminous body of research in the field of economics of education has concerned itself primarily with measuring the impact of SES on student performance in different country contexts[1]. Such research typically uses a household's income or expenditure information to proxy for SES. However, when dealing with survey data from low- and middle-income countries (LMICs), it is not always possible to collect this kind of information. Following the work of Filmer & Pritchett (2001), the use of asset indices has become a widely accepted approach for constructing students' SES in the absence of income or expenditure information. In LMICs specifically, this has grown to become standard practice as the survey designs often include questions on students' possessions at home which provide key information on asset ownership which is required to construct the asset index.

While increased research on the appropriateness of asset indices as proxies of income have led to an overall improvement in their construction techniques, which has led to more accurate measures, there remain some notable challenges. One such challenge discussed at length in Rutstein (2008) has to do with the comparability of the index across different country contexts, regions and datasets. Since these indices are constructed from asset ownership, this challenge is mostly one of weight assignment. Among the problems that arise from this is that depending on how one assigns weights, the ability to distinguish between different socio-economic groups is affected.

---

[1]See for example Sirin (2005) who offers a comprehensive meta-analysis of 58 studies which in summary confirmed a medium to strong relation between SES and student performance

In his paper, Rutstein (2008) discusses possible ways to overcome this limitation. Among these is the construction of a wealth index by sub-samples in the data such as urban and rural classifications. Although this improves overall weight assignment by ensuring that weights are 'localised' to the particular sub-group, one often finds that – especially in LMIC contexts – it becomes increasingly difficult to distinguish between the poorest and the poor population (Rutstein (2008), Vyas & Kumaranayake (2006)). At this point, a common remedy is often to split the wealth index by some further classification, such as quintiles, which are arbitrarily determined cut-off points that classify individuals into one of five equal-sized clusters. However, as Rutstein (2008) demonstrates, this does not deal sufficiently with the problem of distinguishing different socio-economic groups at either national or community levels.

In lieu of these issues, this chapter suggests using a Finite Mixture Model (FMM) approach to identity SES subgroups. FM (finite mixture) modelling allows one to expose any (latent) structure that may underlie empirical data, and is a viable alternative to traditional cluster analysis (Eagle et al. (2010), McLachlan & Chang (2004)). It is viewed as an attempt to allow the data to 'speak for itself', as opposed to techniques that impose arbitrarily defined cut-off points like quintiles. In doing so, the FMM would aid in better differentiating between, for example, urban-rich individuals, from urban-poor individuals in a mathematically robust way. This approach is adopted and applied to data from Malawi — an ideal example of a low-income country.

The broad findings of this chapter show that, irrespective of the sample in consideration, national, urban, or rural, FM modelling prevails at distinguishing between the unique latent socio-economic groups within an individual SES distribution, even among the poorest of households, to other poor households. Specifically, the chapter finds that the rural population is more heterogeneous than the urban population. Using the distinct classes identified by FMM, this chapter identifies the asset mix that best characterises and differentiates each class from another.

The remainder of the chapter proceeds as follows: Section 3.2 presents a country profile for Malawi that seeks to set the context of a typical low-income country; section 3.3 offers a brief discussion on some of the weighting issues one is to expect when using asset indices; section 3.4 outlines the methodology and data to be used in this study; section 3.5 discusses the results of the research findings; and finally, section 3.6 offers a conclusion to the discussion.

3.2   Malawi: A Brief Country Profile

Low-income countries (LICs), such as Malawi, can be defined as those with a GNI of 1005 US dollars or less (Marriott 1974). Due to sustained economic growth, fewer countries than ever before are categorised as falling into this lowest income bracket. In fact, nearly every LIC currently can be found in the sub-Saharan Africa (SSA) region. Not only does Malawi have one of the lowest GDP per capita within this region, but it also remains one of the poorest countries in the world (Brossard et al. 2010).

Geographically, Malawi is made up of 28 districts covering an area of 119 140 square kilometres. Approximately 20 percent of that area is covered by Lake Malawi – the fourth-largest freshwater lake in the world. Unlike most of its natural resource-rich neighbours, Malawi does not boast vast natural resource reserves like copper, gold or diamonds, but is instead largely an agrarian society, depending heavily on agriculture for its livelihood and sustenance. Unfortunately, the agricultural sector itself has not developed to the point where commercial farming is able to contribute significantly to the growth of the economy. Instead, the majority of Malawians rely almost solely on subsistence farming.

For purposes that will become clear later in this chapter, it is of particular importance to note the proportion of the population that is categorised as living either in urban or rural areas. Based on the latest population estimates from the 2018 Malawi Population and Housing Census (PHC) conducted by the National Statistics Office (NSO), Malawi was estimated to have a population of 17 563 749. People living in urban areas made up 16 percent of the total population, and the remaining 84 percent lived in rural areas (National Statistical Office 2019).

While this is undesirable, it is an improvement (though marginal) from the earlier PHC conducted in 2008 that classified 84.7 percent of the population as living in rural areas, and only 15.3 percent in urban areas (National Statistical Office 2019). This urban migration is not surprising. As a country develops one would naturally expect a steadily increasing proportion of the population to be migrating away from rural areas towards urban areas that typically have more opportunities than are present in rural areas. This background provides the context within which the discussions in this chapter are presented.

3.3 Weighting Issues in SES Measures Using Asset Indices

In chapter 2 of this thesis, principal component analysis (PCA) was used to obtain a measure of household wealth for use in the estimation of socio-economic gradients for Malawi and Namibia, This approach has become almost standard within the literature since the seminal work of Filmer & Pritchett (2001). As indicated in Technical Note 1 of Appendix A to this thesis, the construction of a continuous index of SES is obtained from a linear combination of the household assets (coded as binary, 1 = 'Yes, posses' and 0 = 'No, do not posses') and their respective weights estimated from the PCA.

Higher weights are given to assets that are scarce in the population. This loosely suggests that such assets are most likely owned by two groups: the more affluent (since they are the ones that can afford them); and the poorest (since they may be assets which no one else wants or are simply unique to their setting). On the other hand, an asset owned by a larger segment of the population would be assigned a relatively smaller weight. One can therefore easily understand why in LICs that have disproportionately large rural populations, like Malawi, this can cause problems since the weight assignment would primarily be driven by the asset mix and ownership of those living in rural areas, as they make up a significantly large segment of the population.

As demonstrated by Kotzé & Van der Berg (2015) and Vyas & Kumaranayake (2006), a reliable remedy to this shortcoming is to rather create the asset index by subgroup instead of relying on one created at national level. This change would crucially localise weight assignment so that weights take into account the inherent heterogeneity of the community in question. In doing so, the asset-based wealth index is better able to accurately represent not only the wealth of the nation as a whole, but also the communities that comprise it.

As an example, consider again the case of access to electricity. Accessibility and, therefore, price of electricity services is influenced to a large extent by geographical location. Whereas households in urban areas are typically more likely to benefit from state investment in electricity generation and distribution infrastructure, making it more accessible and affordable, the same may not be true for households in rural areas. Indeed, if at all accessible in rural areas, such a 'privilege' might unfortunately only be reserved for the rural-rich. An asset index created by urban/rural

classification would more correctly assign appropriate weights to this asset: a relatively larger weight to those in rural areas, and a smaller weight to those in urban areas.

Depending on the research question at hand, and once the researcher has a localised wealth index that can serve as an appropriate proxy for household SES at both the national and community level, the next thing to consider is how to differentiate and classify different SES groups within those communities. That is, given a rural or urban SES index, how should one go about differentiating the rural-rich from the rural-poor, or the urban-poor from the urban-rich. The norm has typically been to — whether arbitrarily or by a data-driven approach — define cut-offs in the first principal component score in such a way that differentiates and splits the sample into unique and distinguishable socio-economic groups (Gwatkin et al. 2000).

Unfortunately, splitting the PCA SES wealth index in this way crucially assumes that the SES variable is uniformly distributed, which may not always be the case in LICs. In fact, using a selection of developing countries, Vyas & Kumaranayake (2006) showed that the distribution of urban and rural PCA factor scores reveal evidence of 'clumping' and 'truncation' in such low-income areas. Clumping is when households are clustered together in a small number of groups as opposed to filling out the entire distribution, while truncation refers to a more evenly distributed SES, but one that is spread over a narrow range of factor scores, making it difficult to distinguish between different socio-economic groups (McKenzie 2005).

Being able to distinguish between these socio-economic groups is precisely the problem this chapter concerns itself with. This is important because whilst creating an SES variable by urban or rural samples goes a long way in correctly differentiating between segments of the population that are potentially dissimilar in their asset ownership, it is unable to further distinguish between subgroups. In scenarios where, for example, one wants to draft ultra-poor policies, being able to make a distinction between the rural-rich and rural-poor has direct policy relevance as it more accurately informs targeted responses. This not only wastes less resources, but also ensures the beneficiaries of such responses are the intended group.

There are different ways to ensure the SES variable derived from PCA is split into the socio-economic groups that together comprise it. As stated, creating the wealth index at community

level to localise weight assignment already goes a long way to ensure that the wealth index better reflects the true wealth ranking of the population by crucially accounting for this notable difference in where the individual lives, which by itself carries information about possible socioeconomic status. Though the choice to use urban/rural is itself an arbitrary one, it classifies households by some known parameter that distinguishes them by geographical location. In some instances, as in Vyas & Kumaranayake (2006), this is used in combination with cluster analysis to classify households into 'low-', 'middle-', and 'high-' wealth groups.

Clustering is a statistical procedure that assigns objects to groups or clusters in such a way that those in the same group are more similar than those that are not. This method has been used severally for this task to classify households into different wealth groups (Cortinovis et al. 1993). This stands in contrast to Filmer & Pritchett (2001), who used some arbitrarily determined cut-off of the form 40-40-20 to classify the wealth index distribution into unique wealth classes representing the poorest 40 percent, the middle 40 percent, and the richest 20 percent of the population.

While this form of 'clustering' is commonplace and an attempt to accurately disaggregate the wealth index distribution, this approach performs better in parts of the distribution where more affluent members of the population are found. In fact, Vyas & Kumaranayake (2006) importantly note that while such a disaggregation is likely to perform well in more urban areas, it is less likely to give reliable results in low-income areas. This is because the disaggregation "would not reflect the clustered nature of the underlying data" (Vyas & Kumaranayake 2006, p.465). This means that, for LICs like Malawi that are in the top ten poorest countries of the world, such approaches — while insightful — are likely to misclassify the wealth rankings of the households. To classify individuals/households into their appropriate socio-economic clusters after constructing an asset-index whether at national or community level like urban or rural areas, this chapter makes use of a technique called finite mixture modelling.

3.4 Methodology and Data

3.4.1 Finite mixture modelling (FMM)

Broadly, FM models are applied for two main purposes: the modelling of distributions of a wide variety of random phenomena, and as a tool for the clustering of data (McLachlan & Chang 2004, McLachlan & Peel 2004). This chapter specifically uses FMM for the latter. In doing so, FMM can be used as a tool to uncover the existence and nature of underlying latent structure present in data, making it possible to group the data into distinct socio-economic groups. In contrast to arbitrary cut-off choices, FM models thrive at this as they are a model-based clustering technique entirely satisfactory from a mathematical point of view (Hawkins et al. 1982, Marriott 1974).

The main idea in FM modelling is that observed data come from an initially specified number of components, g , in various proportions. A researcher can, therefore, confidently use mixture models to model heterogeneous (sub) populations. To this end, Aitkin et al. (1981) noted that clustering methods based on such models allow analysis to be performed within the framework of standard statistical theory. In this way, mixture modelling provides an objective framework even for model selection (McLachlan & Peel 2004), whilst allowing the researcher to investigate both the existence and nature of underlying natural and latent subgroups or clusters in univariate or multivariate distributions Melnykov et al. (2021), Izenman & Sommer (1988), McLachlan & Basford (1988).

In FM models, each observed data point, y_j , is taken to be a realisation of a mixture density:

$$f\left(y_j\right) = \sum_{i=1}^{g} \pi_i f_i\left(y_j\right) \tag{3.1}$$

where g corresponds to the number of components (or classes), $f_i\left(y_j\right)$ are the component densities of the mixture, and π_i are non-negative quantities (i.e $0 \leq \pi_i \leq 1$) that altogether sum to one, otherwise known as the mixing proportions. These mixing proportions are estimated using a multinomial logistic regression given by:

$$\pi_i = \frac{\exp(\gamma_i)}{\sum_{j=1}^{g} \exp(\gamma_j)} \tag{3.2}$$

where γ_i is the linear prediction for the ith latent class[2]. Following McLachlan & Chang (2004), one can specify the parametric form $f_i(y_j; \theta_i)$ for each component density and thereby fit the parametric mixture model:

$$f(y_j) = \sum_{i=1}^{g} \pi_i f_i(y_j; \theta_i) \tag{3.3}$$

by maximum likelihood (ML) estimation via the expectation maximisation (EM) algorithm of Dempster et al. (1977)[3]. Upon fitting the mixture model, the data is assigned to g clusters by the fitted posterior probabilities of component membership for the data.

This can be contrasted to, for example, a k-means clustering algorithm that involves a hard assignment of data into one and only one cluster.[4] Conversely, the FMM technique of clustering is an example of soft assignment, assigning each data point to the component to which it has the highest estimated posterior probability of belonging (Lee & Shin 2017, Li et al. 2007). These estimated posterior probabilities, though less reliable in small samples, still make for a sufficient assignment of the data.

In using FMM, deciding on the number of components/classes present in the data is a question that not only arises naturally, but is also the subject of a vast amount body of literature due to its methodological complexity (for a recent review see McLachlan & Rathnayake (2014).

In practice, this process can be aided by a researcher's prior information about the number of suspected components present in the distribution. Izenman & Sommer (1988), for example, use their prior knowledge of the Hidalgo stamp of Mexico data to guide their choice of g, leading to a model with three components.

---

[2]By default, the first latent class is the base level, implying that 1=0 and exp1=1. The mixing proportions calculate the probability of a data point belonging to a specific class.

[3]See McLachlan et al. (2019) for a discussion on the EM algorithm.

[4]Note that K-Means clustering and finite mixture modelling have been compared expertly by Lenzenweger et al. (2007).

It is, of course, not always possible to have prior information concerning the potential number of component densities to use as a guide. Among the preferred alternatives are mode counting and hypothesis testing.

Roeder (1994) argued that, in the absence of prior information, one could simply assess the number of modes.[5] Naturally, of course, the success of this approach would depend on the parameters of the distribution in question; for example, if the components of the mixture are not sufficiently wide apart (because of common variances and/or means), then the number of (unique) modes would be difficult to detect (Li et al. 2007). In this way, a mixture distribution can appear to be unimodal even when it is not, leading to a downward bias in the number of observed components/classes[6].

As an aside, an important caveat is that concern arises when the number of components is interpreted as representing the number of unique classes in the population. This has to do with the parametric specification of the component densities. Consider the task of modelling a distribution that has skewed/asymmetrical distributions; certainly, such a distribution can perhaps be easily approximated by a set of finite mixtures of normal densities (McLachlan & Peel 2004, p.197). Therefore, in such a skewed/asymmetrical distribution case, one may not easily count the modes to represent one-to-one correspondence with the distinct number of components. This is because even if just one mixture density were skewed, multiple normal components would be needed to model that distribution more accurately.[7]

Of course, this idea seems to imply that in an attempt to offer a model of best fit to the data, the mixture modelling technique runs the potential risk of continuing to (infinitely) group the data into g components. Empirically, then, it is possible that cases with $g-1$ or $g+1$ components may be indistinguishable (McLachlan & Peel 2004, p.175). To resolve this, the choice of number of components is instead assessed as the smallest number of components compatible with the data, and not merely as all the possible components (McLachlan & Chang 2004). In equations 3.1 to 3.3, g is, therefore, defined to be the smallest value of g such that all components are different,

---

[5]This assumes that each mode represents a single and unique component.

[6]Due to this lack of correspondence between unique modes and clustering, authors Li et al. (2007) proposed a related non-parametric mixture modelling approach that addresses these challenges.

[7]See Gutierrez et al. (1995), McLachlan & Peel (2000) and for a discussion of possible adjustments that can be made in the event of skewness.

and all the associated mixing proportions, π_i , are nonzero.

Apart from mode counting, one can naturally consider using a likelihood ratio test to compare different models. In this case, one has the option to sequentially compare models with different values of g , or to compare constrained and unconstrained models for a given g (Deb & Trivedi 1997, p.321). Unfortunately, the likelihood ratio (LR) test is only appropriate in the latter, where both the constrained and the unconstrained model are hypothesised to have a single g . In the case of the former, the hypothesis relies on the boundary of the parameter space, therefore violating the standard regularity conditions necessary for maximum likelihood (Maitra & Melnykov 2018). In other words, the usual asymptotic null distribution of chi-squared with degrees of freedom equal to the difference between the number of parameters under the null and alternative hypotheses necessary for the likelihood ratio statistic λ to be reliable do not hold (McLachlan et al. 2019). This notwithstanding, some researchers (see for example, Izenman & Sommer (1988)) have made use of the LR test statistic as a guide in their selection of a valid g .

Following McLachlan & Peel (2004), the null and alternate hypothesis are specified as follows:

$$H_0 : g = g_0$$
$$H_1 : g = g_1$$

(3.4)

for some $g_1 > g_0$ (usually $g_1 = g_0 + 1$) using a resampling method in order to produce a p-value (Scrucca et al. 2016).. The choice of g is then based on the simple rule-of-thumb that components are added until the increase in the log likelihood begins to diminish as g goes beyond some threshold (McLachlan & Peel 2004, p.185). This threshold often takes the value of g_0 in H_0 . The bootstrapped samples are generated from the mixture model fitted under the null hypothesis of g_0 components. The value of $-2\log\log\lambda$ is then calculated for each (bootstrapped) sample after fitting mixture models for $g_1 > g_0$. This process is independently repeated iteratively a number of times and used to perform an assessment of the true null distribution of $-2\log\log\lambda$ (McLachlan & Chang 2004).

An alternative, and commonly preferred method of testing the hypothesis above has been to adopt

54

the BIC criterion of (Schwarz 1978). In this particular context, we reject H_0 if twice the increase in the log-likelihood is greater than $d \log \log (n)$, where d denotes the number of free parameters in the model (McLachlan & Chang 2004). Computationally, BIC is given as:

$$BIC = -2 \log \log \lambda + d \log \log (n) \tag{3.5}$$

In addition, this chapter will also report and compare results from a second information criterion often reported in conjunction with BIC, the AIC (Akaike 1974), given as:

$$AIC = -2 \log \log \lambda + 2d \tag{3.6}$$

Although this chapter will only discuss the LR statistic, BIC and AIC for choosing the number of components, the literature is replete with additional information criteria that can be used for the same purpose, including: the Normalized Entropy Criterion (NEC) (Biernacki et al. 1999), Minimum Information Ratio (MIR) (Windham & Cutler 1992), the adjusted BIC (Sclove 1987), and the adjusted LRT (aLRT) (Lo et al. 2001).[8] For a more detailed discussion on the comparative performance of the information criterion (IC) and set of likelihood ratio set of tests for determining the appropriate number of components in a mixture see Nylund et al. (2007).

In the event that the three test results (LR statistic, AIC and BIC) point to drastically different numbers of components being present, this chapter will give preference to the BIC. As shown by Maitra & Melnykov (2018), when a normal mixture is used to estimate density non-parametrically, the BIC performs consistently in the choice of number of components (see also Keribin (2000), Dasgupta & Raftery (1998), Roeder & Wasserman (1997)). Although BIC is also known to underestimate the number of components, this is typically a small sample size problem and, therefore, not relevant for this chapter Melnykov & Maitra (2010). In contrast, AIC tends to overestimate the correct number of components (Celeux & Soromenho 1996, Soromenho 1993). On other related grounds, BIC has also been shown to overestimate the number of clusters themselves (Biernacki

---

[8]Their use in aiding model selection has a formal justification as discussed in Leroux (1992) where he proves approaches such as AIC and BIC lead to a consistent estimator of the true finite mixture model.

et al. 2000) . In response to this, Biernacki et al. (2000) developed the integrated classification (ICL) criterion model that penalises the BIC through an entropy term measuring cluster overlap (Scrucca et al. 2016). In cases where the cluster overlaps are not strong, the ICL approach has shown across a range of cases to perform very well in selecting the number of clusters (Maitra & Melnykov 2018).

3.4.2 Data

SACMEQ is a consortium of education stakeholders in member countries, whose primary mission is to enhance collaborative efforts among member ministries of education by undertaking integrated research and improving capacity by providing training in relevant technical skills. It was officially founded in 1994 in Paris at a gathering of seven initial ministries of education in conjunction with the International Institute for Educational Planning (IIEP). These initial seven member countries are (in alphabetical order): Kenya, Malawi, Mauritius, Namibia, Zambia, Zanzibar, and Zimbabwe. Membership has since increased to include fifteen governments.[9] To date, there have been four education research projects facilitated by SACMEQ, commonly referred to as 'SACMEQ I' (conducted from 1995 to 1999), 'SACMEQ II' (conducted from 2000 to 2004), 'SACMEQ III' (conducted from 2006 to 2011), and 'SACMEQ IV' (conducted from 2012 to 2014).

This study limits itself to only SACMEQ III data for Malawi. In total, 2 781 Grade 6 learners across 139 schools were sampled. In addition to capturing performance on literacy and numeracy assessments, contextual questionnaires were given to Grade 6 students concerning (but not limited to) household possessions, building material, and parental education. These items are used for the construction of the SES index in this chapter.

Since this chapter uses data collected around the 2007 period for Malawi, a reliable reflection of the state of the population size at the time can be found in the PHC of 2008, which reported a national population size of 13 077 160. As mentioned earlier, 84.7 percent of the population lived in rural areas while 15.3 percent lived in urban areas (National Statistical Office 2008). The validity of the representativeness of the data used in this chapter can be verified by confirming if –

---

[9]The member countries – in alphabetical order – are: Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania (Mainland), Tanzania (Zanzibar), Uganda, Zambia and Zimbabwe.

once adjusted for survey design – the rural and urban classifications of the households in the data are approximately similar in proportion to those of the 2008 census.[10] Fortunately, the results affirm this finding.

3.4.3 Data description

As an example of the overall shape of the asset index, Figure 3.1 below shows a histogram of the wealth index constructed with data from the full sample using PCA. Following Wittenberg's(2009) suggestion, PCA is applied on both binary (yes/no) responses asset questions, and also ordinal questions about, for example, the house building material to construct a wealth index (see also Rutstein & Johnson (2004)). I additionally include parental education as in Ross et al. (2005). As can be seen, this distribution is not normally distributed as was shown to be the case for low-income countries by Vyas & Kumaranayake (2006). This exacerbates the difficulty of correctly classifying students into appropriate socio-economic clusters that reflect their true socio-economic background based on their asset-ownership which warrants the use of a mixture model to aid in this exercise.

---

[10]This information is derived from a School Head Survey Questionnaire. Specifically, it was Question 3 which asked the School Head: "Which of the following best describes the location of your school?" . The responses were: (1) Isolated; (2) Rural; (3) In or near a small town; and (3) In or near a large city. Depending on how you manipulate this variable to classify the population into rural and urban, the results you have differ. This chapter combines responses (1) and (2) and classifies those as rural. This results in a rural population of approximately 76 percent, and an urban one of 24. If instead you classify options (1), (2), and (3) as rural, you have a rural population of 85 percent, and an urban population of 15 percent which corresponds to the NSO's 2008 census.
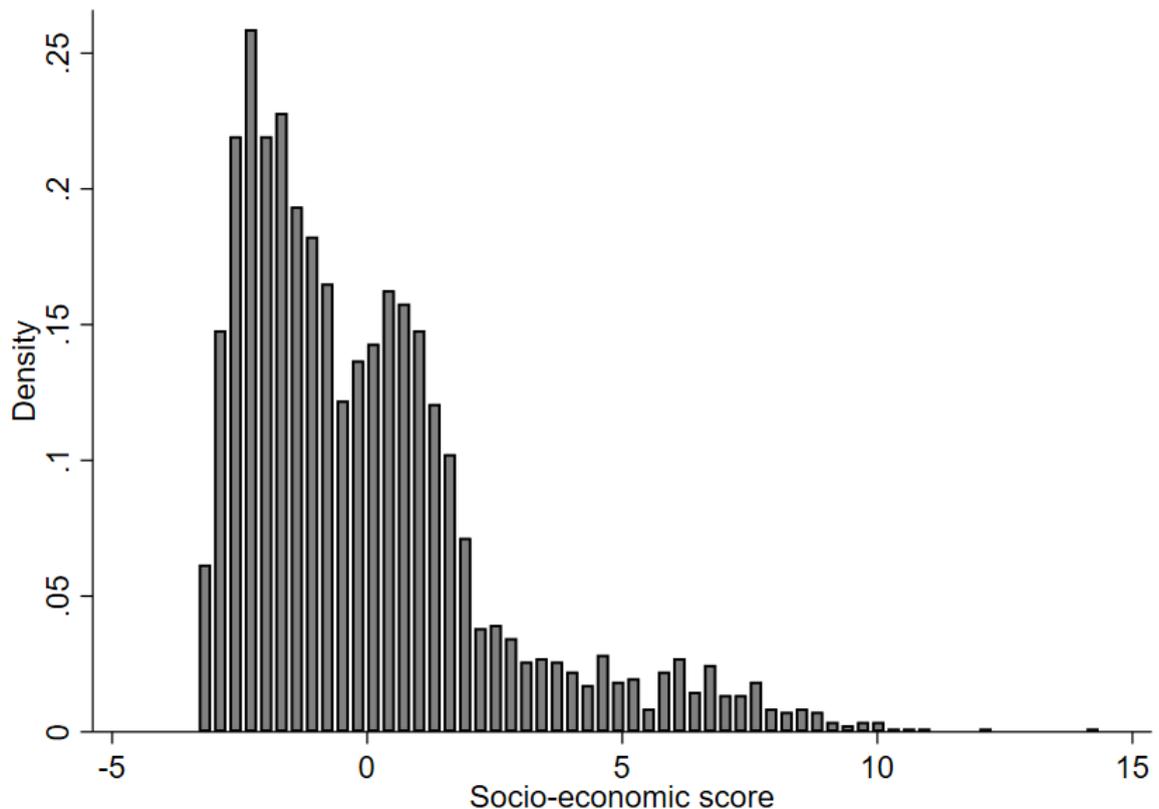
Figure 3.1: Distribution of wealth index for the full sample

3.5 Empirical Results

3.5.1 Finite Mixtures

3.5.1.1 *Choosing the number of components*

Analysis was performed using the two-step, EM-based finite mixture analysis for 1 through *g* number of components. Since the rule-of-thumb as regards BIC is that smaller values of BIC indicate a better fit (Lenzenweger et al. 2007), one conducts this technique on an increasing number of components until the smallest BIC is observed, at which point you terminate the process. The results of this process on the full Malawi SACMEQ III sample are presented in Table 3.1 below.

Table 3.1: Summary of Finite Mixture Model Fits for PCA of Malawi at national level.

g	Log Likelihood	BIC	AIC
1	931.68	12835.3	12823.5
2	423.26	11927.3	11897.8
3	67.17	11527.7	11480.5
4	10.78	11484.3	11419.3
5	3.87	11497.2	11414.6
6	13.33	11517.0	11416.7
7	1.31	11527.4	11409.4
8		11549.8	11414.1

Note: g = number of components (groups/clusters); BIC and AIC are Bayesian and Akaike information criterions respectively.

The results from the BIC contained in Table 3.1 indicate that the PCA constructed at national level contains a minimum of four normal components. Both the AIC and log likelihood ratio (LR) hypothesis test seem to suggest that there are instead five components. Though that is the case, the difference in values between the four- and five-component model in the BIC case is marginal. This gives one confidence to opt for the four-component model. This can be contrasted to the instance where one arbitrarily chooses to split the full sample PCA by, for example, a 40-40-20 split. This would potentially misclassify households into inappropriate wealth groups. In this particular situation, it would run the risk of underreporting the degree of latent socio-economic heterogeneity in the full Malawian sample of students.

FM modelling was additionally used to assess the number of components present in the SES factor scores from a PCA constructed at community level: urban and rural. The results are presented in Table 3.2. In contrast to the previous case with four components, the AIC and BIC both favour the three-component model in the urban case. The LR test instead favours a $g + 1$ component model in

the urban area. For an LIC, three broad wealth clusters within the urban area is plausible: urban areas, assumed to be wealthier than their rural counterparts, are likely to also have less diversity (since they also comprise of a small segment of the population), and in addition have a mix of asset ownership that better allows one to distinguish between classes and overcome the common problems of truncation and clumping that often plague poorer segments of the distribution (Vyas & Kumaranayake 2006). Further, as Table 3.2 below shows, of the three classes present in the urban community, only one has a positive mean SES score, and represents about 34 per cent of the population. It is possible, therefore, that the other two classes — making up 66 per cent of the urban population — are composed of a rural-to-urban migrant population, which would explain why the majority of the urban population still has negative mean scores; that is, asset accumulation remains low.

Table 3.2: Summary of Finite Mixture Model Fits for PCA of Malawi at community level

	Urban			Rural		
g	AIC	BIC	LR	AIC	BIC	LR
1	3,033.79	3,042.70	0.12	9,330.91	9,342.18	634.38
2	2,949.66	2,971.94	22.49	8,702.53	8,730.70	296.62
3	2,933.17	2,968.81	4.31	8,411.91	8,456.98	65.97
4	2,934.86	2,983.87	1.58	8,351.94	8,413.92	18.63
5	2,939.28	3,001.65	3.60	8,339.30	8,418.19	10.18
6	2,941.68	3,017.42	7.48	8,335.12	8,430.90	4.09
7	2,940.20	3,029.30	-0.37	8,337.03	8,449.72	8.36
8	2946.571	3049.04				

Note: g = number of components (groups/clusters); BIC and AIC are Bayesian and Akaike information criterions respectively. Own calculation from SACMEQ 2007 data.

The FMM analysis on the rural subgroup of the population appears to favour a classification of up to a minimum of four normal components if one uses the BIC as a measure of best fit, and goes all the way to six if one refers to the results from the LR and AIC values. If one is to accept the result from the BIC, it would be important to ask if having four wealth groups is plausible in the rural case, and if so, why? This chapter offers two possible explanations. The first is that in LICs, the rural subgroup of the population represents the larger part of the population. In the Malawi case specifically, in the period of data collection (2007–2010), the rural population made

60

up approximately 11 million of Malawi's 13 million national population. In such a large population, it is feasible that there multiple wealth classes exist as opposed to one or two homogeneous subgroups.

Secondly, because individuals in rural areas often own similar assets given limited access (both financially and geographically) to broader markets, it would mean that owning even only one asset that other individuals do not own would classify you as wealthier or poorer. The potential for clumping and truncation is relatively high, because there is a shortage of distinct and unique assets by which we can distinguish one individual's wealth status to another. By the same token, it means that the presence of even one particular asset is able to set individuals apart. It is plausible that there this number of components exists in the rural segment of the population, implying greater heterogeneity among the poor. This data-driven decomposition provides much needed insight into the poor and shows that even in the rural area, there exist rural-elites, rural-poor, and two 'middle' classes. This information is of particular relevance in education data, as differentiated sets of households may have differing impacts on student performance.

3.5.1.2   Class membership probabilities

After resolving the number of components underlying the SES index with FM modelling, the marginal class probabilities are obtained that evaluate the probability of individuals belonging to a particular group/class at both national and community level. In addition, FM modelling allows the researcher to calculate the mean socio-economic score within each class. These are presented in Table 3.3.

Table 3.3: Class membership probability and mean socio-economic score by location

Class	National Proportion	Mean	Urban Proportion	Mean	Rural Proportion	mean
1	0.17	-2.56	0.16	-3.30	0.16	-2.31
		(0.05)		(0.17)		(0.06)
2	0.30	-1.58	0.50	-0.92	0.37	-1.29
		(0.09)		(0.14)		(0.09)
3	0.39	0.54	0.34	2.97	0.39	1.19
		(0.09)		(0.42)		(0.14)
4	0.15	4.71			0.07	5.12
		(0.35)				(0.62)

Source: Own calculation from SACMEQ 2007 data.
Note: Standard errors are reported in parenthesis

For the PCA constructed at national level, approximately 47 percent of the population has a negative mean socio-economic score,[11] although the positive mean score of the third class is marginally above zero. Interestingly, in all three classifications (national, urban, and rural), the mean difference is consistently highest when moving between the relative 'richest' class and the adjoining class. This suggests that within each classification there is a significant difference in SES between the wealthiest segment and the 'next' wealthiest. Furthermore, even when the closest SES counterparts have a positive mean factor score, implying they are non-poor, the difference between them and the richest class is still quite large. This is particularly interesting at the rural community level where one would normally assume that all households are of relatively similar SES. Note here that while comparisons are made between rural, urban, and national SES components and clusters, conclusions are here made cautiously, recognising that an SES derived from assets alone, which in addition is already known to be urban-biased in its construction, would likely misrepresent wealth in a rural setting. This is because wealth/SES in rural settings is more likely reflected in a different mix of assets such as ownership of livestock, inherited ancestral land, and sometimes even the number of children. Unfortunately, this is not information that is captured in the data.

The FM analysis is, therefore, able to clearly make a distinction between the rural-rich, the rural-

---

[11]Interestingly, approximately 40 percent of the Malawian population lived below the poverty line in 2007 which is close (enough) to the 47 percent reported here (National Statistical Office 2008).

middle and the rural-poor. Further, for rural Malawi, one notices that approximately 53 percent of the population has a negative mean score value, implying that even within the rural population, the rural-poor are a dominant class.

3.5.1.3 *Predicted mixture densities*

In addition to calculating class membership probabilities and means, FM modelling also allows one to plot the predicted mixture density of each class. Before doing so, however, it is relevant to show what the distribution of the wealth indices looks like for the different samples— national, urban, and rural. These distributions are plotted as histograms in Figures 3.2, 3.3, and 3.4.
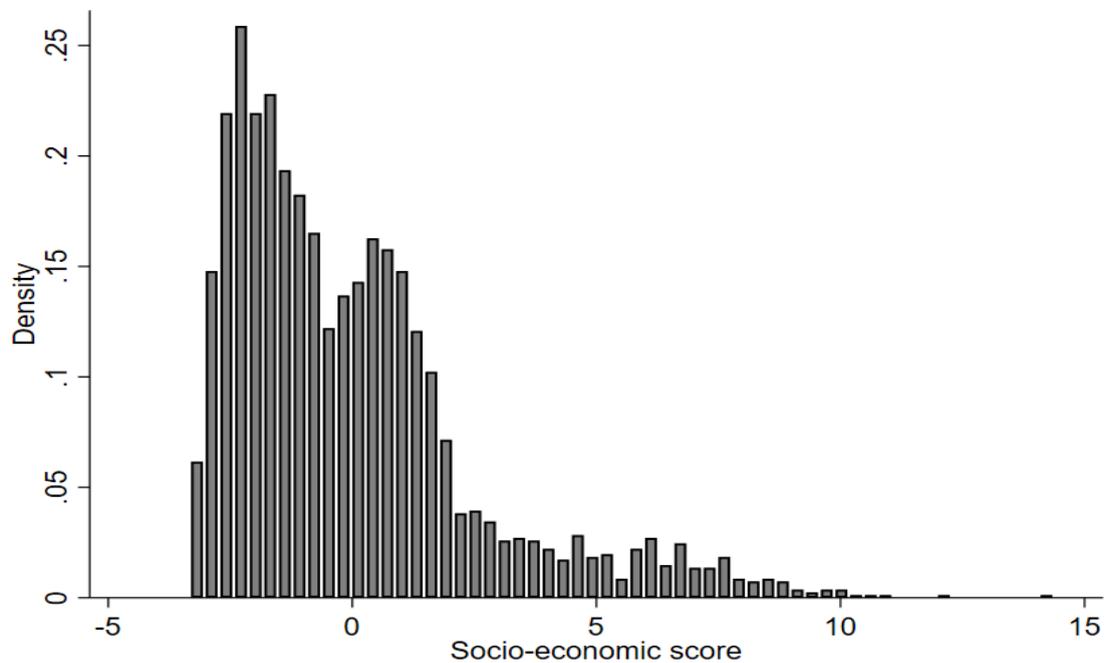
Figure 3.2: Distribution of wealth index for the full sample
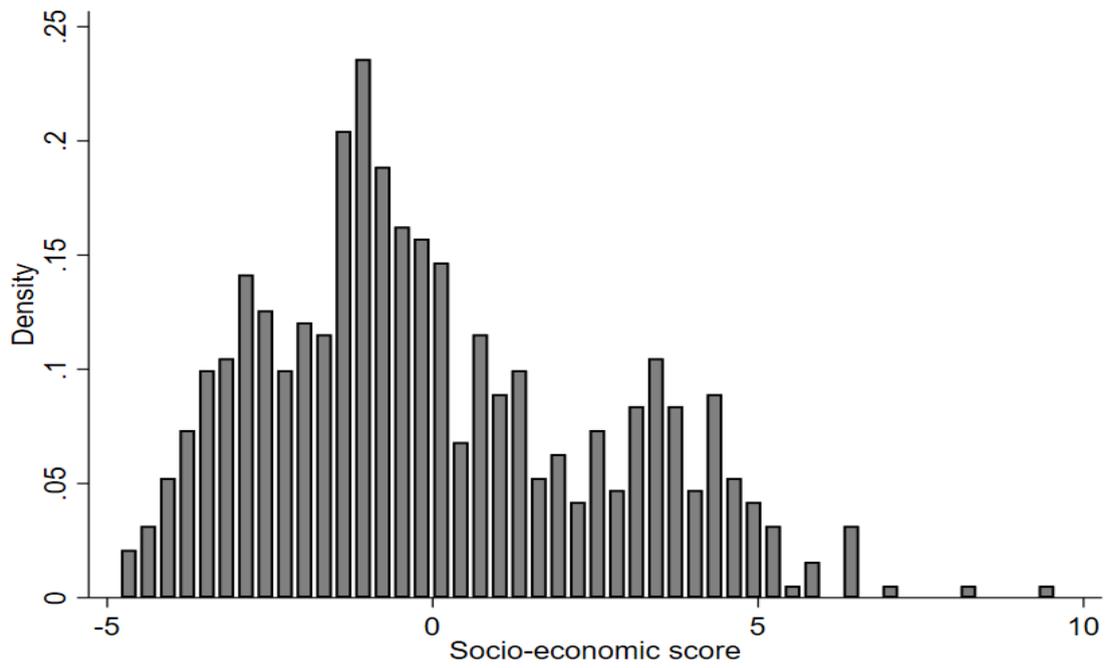
Figure 3.3: Distribution of wealth index for the urban Malawi sample
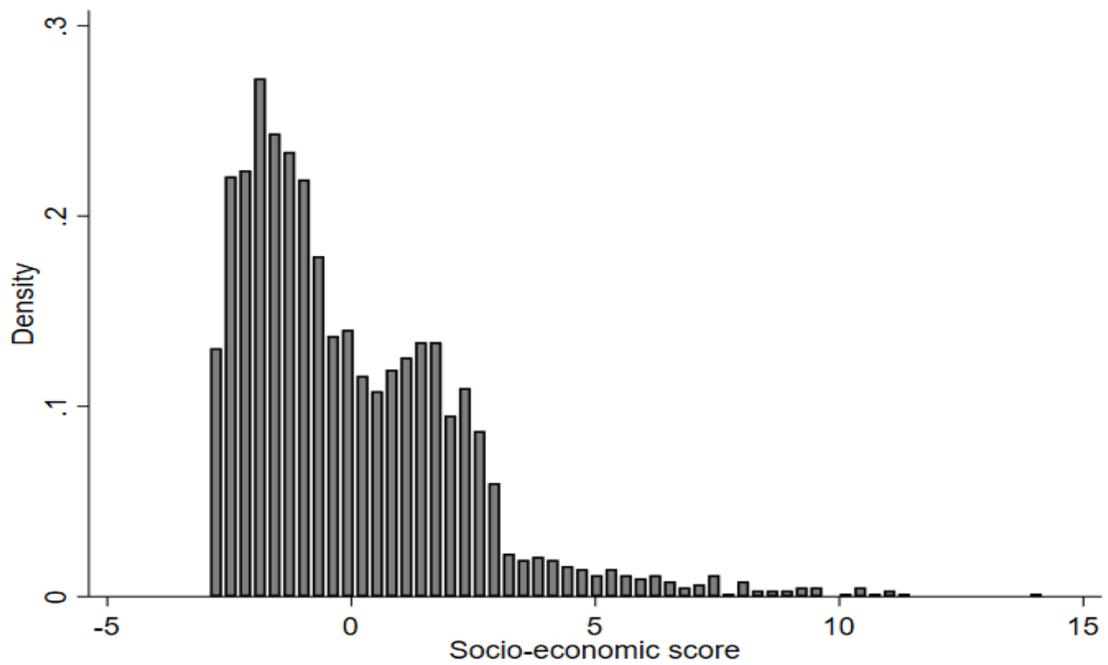
Figure 3.4: Distribution of wealth index for the rural Malawi sample

As expected, the overall shapes of the full and rural samples as demonstrated in the histograms, are similar, attesting to the (disproportionate) influence that the rural segment of the population has when constructing an asset-index in such low-income contexts. To these, FMM allows one to superimpose a plot of the individual latent class density, that is, the mixtures that together make up the histogram. This is possible because output from FMM gives all the necessary parameter estimates that allow one to do so. In the urban Malawi case, were g is equal to three, for example, the output from FMM means one can specifically fit the following model:

$$f(y) = \pi_1 N\left(\mu_1, \sigma_1^2\right) + \pi_2 N\left(\mu_2, \sigma_2^2\right) + \pi_3 N\left(\mu_3, \sigma_3^2\right) \tag{3.7}$$

using estimates of the means (μ), variances (σ), and class probabilities (π) for three distributions that represent the three different types of wealth classes discussed earlier. These distributions are superimposed on the empirical histogram of the data in Figure 3.3. Similar output for the full and rural samples are available in the Appendix B.1.
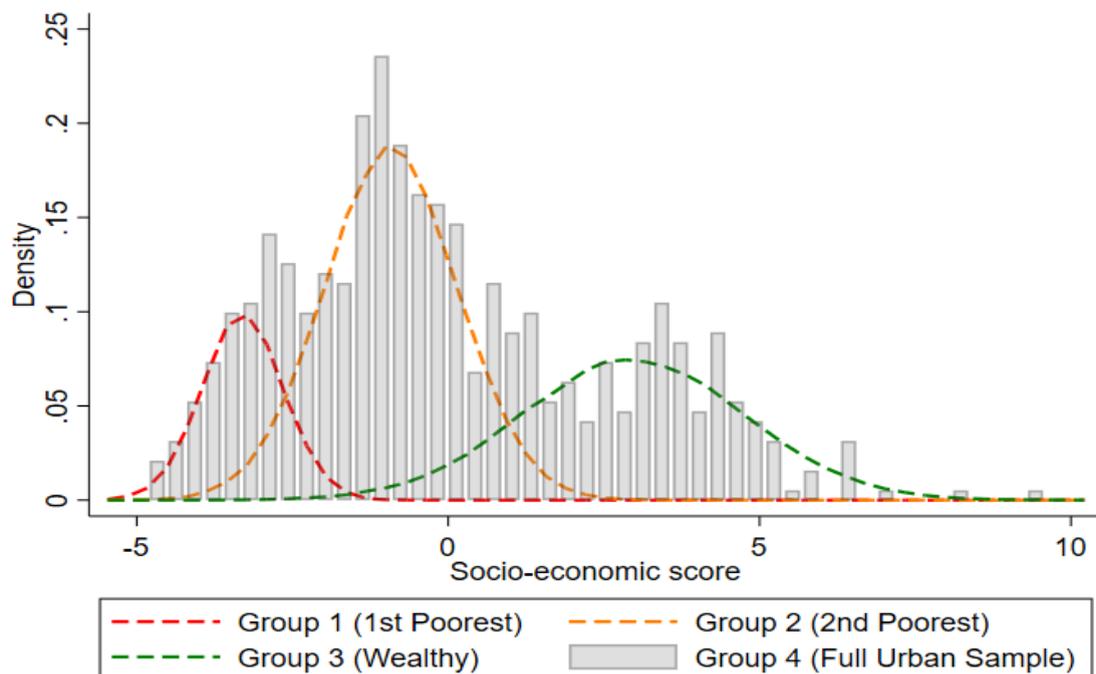
Figure 3.5: Kernel densities of wealth classes and histogram of wealth index for the urban Malawi sample

Figure 3.5 demonstrates that the larger proportion of the urban population belongs to the middle SES group, with the smallest proportion belonging to the wealthiest group. Group 1 (the urban-poor) is less concentrated than both the middle (Group 2) and wealthiest (Group 3) urban groups, as indicated by the larger variance. Unlike what one sees and expects in the histogram of the rural subgroup, the urban distribution shows only slight evidence of clumping and truncation. This implies that the variables chosen for constructing the PCA—at least for the urban sub-population—are able to distinguish between the different socio-economic groups.

3.5.2 A practical example: asset ownership and mean student performance among latent classes of urban and rural sub-populations

To further appreciate the unique socio-economic groups suggested by FMM, as well as link the findings of this chapter to the previous chapter (Chapter 2) of this thesis, this section discusses: (1) the asset ownership of each of the classes in rural and urban Malawi, looking particularly at what mix of assets characterises and differentiates each class and, (2) how average student performance in numeracy and literacy tests differ among the classes.

3.5.2.1 *Asset ownership within urban and rural classes*

The results discussed here compare asset ownership within and across urban and rural classes. Summaries of these can be found in Table B.1 to Table B.5 of Appendix B.2.[12] Recall that the model suggests three urban and four rural socio-economic classes, each characterised by its own asset ownership mix. I begin with a discussion of the three urban classes.

Looking at Tables B.1 to B.5 in Appendix B.2, the urban-poor appear to differ from the urban-middle mainly in their access to utilities — namely, electricity and piped water, where the urban-poor have less than their urban-middle counterparts. Similarly, the urban-rich have even greater access to information and communication (for example, newspapers and phones) when compared to the urban-middle class. In addition, the urban-rich are characterised by greater access and ownership of personal modes of transportation. This is likely due to the fact that cities have more modes of public transportation that people can utilise without personally owning cars or

---

[12]For each class, the reader can also request an urban versus rural comparison from the author. That is, those tables present direct comparisons of asset ownership of the urban and rural wealthy, urban- and rural-poorest, and urban- and rural-middle.

motorcycles (something the rich can afford to do). In addition, the urban-rich appear to have more access to leisure assets such as video cameras, photography cameras and television sets.

As for the four rural classes, the poorest among them, class 1, differs from class 2 in asset ownership primarily through access to information, communication and household furniture (items such as tables, beds and clocks). Since they both have almost negligible access to electricity, both the bottom two rural classes do not seem to own assets that require the use of electricity (e.g. electric fans, computers and refrigerators). In fact, it appears that among all rural classes, only the rural-rich have access to electricity, while in urban areas not even the urban-poor seem to have access to electricity.

While it is beyond the scope of this study to understand why this would be the case, a plausible explanation can be offered for this occurrence. It is important to note that a crucial point excluded from the survey questionnaire, yet which provides justification for some of the anomalies in the asset ownership findings (such as the rural-rich having greater access to electricity than urban-poor), is the lack of information on institutions. One would expect a more pronounced institutional footprint in urban areas in developments concerned with public health, transport and education than in rural areas.

As mentioned earlier, in terms of transport, these could have the effect of lessening the need for urban dwellers to own personal vehicles because of the availability of more affordable modes of public transport. As for the rural-rich having more access to electricity than the urban-poor, institutions that govern urban/rural land planning may provide the rural-rich with more flexibility to choose where they live, which can be around areas that are in close proximity to electricity such as main roads or District Centres (DC).

In contrast, the urban-poor do not have the flexibility to stay where they would like, but instead typically live where they can afford to live, which in a poor country like Malawi, may not always have electricity. This is despite urban areas typically being located closer to the existing grid network (Korkovelos et al. 2019). Keep in mind that as of 2018 only 11 percent of the Malawian population was reported to have access to electricity, with even the richest 20 percent of the population only reporting 31 percent electrification rate (World Bank 2018, 2019).

As for the two middle rural classes, they are largely similar, differing mainly in that the relatively richer class 3 has a greater ownership of furniture items such as tables and beds. Lastly, the rural-rich differ from the rural-middle classes not only in access to electricity, but also in that they enjoy more access to communication, furniture, leisure and entertainment assets.

3.5.2.2   *Mean student performance between classes within urban and rural sub-populations*

Previous research has presented overwhelming evidence that demonstrates that students from advantaged socio-economic backgrounds outperform their less-advantaged counterparts (Coleman 1968). Given the model predictions of the number of latent classes within the urban and rural sub-populations, a logical exercise to perform would be to assess whether the students so classified perform in accordance with theory; that is, do students with a higher SES mean score outperform their lower SES mean score counterparts? The results from this exercise are presented in Table 3.4.

Given that Malawi comprises a largely homogenous population in terms of student performance, having the smallest socio-economic gradient slope among all SACMEQ countries (see Chapter 2 of this thesis), it is no surprise that there does not appear to be much in the way of a noticeable difference in mean performance between the poorest and the richest students within urban/rural contexts. That is, even though students from richer latent classes perform better in both literacy and numeracy test scores, the differences are not large. In summary, what this section shows, is that the previously single SES distribution from performing a PCA on household assets at the nation, urban, and rural levels, is in fact made up of distinct socio-economic classes that are characterised by different asset ownership mix.

Table 3.4: Mean scores for students by classes within national, urban, and rural samples

| | Mean Reading Score | | | | Mean Numeracy Score | | | |
	Class 1	Class 2	Class 3	Class 4	Class 1	Class 2	Class 3	Class 4
National	432	425	433	456***	448	445	445*	457
SE	-5.8	-2.8	-2.6	-3.9	-5.9	-4.1	-2.8	-3.7
N	511	808	1049	336	511	808	1049	336
Urban	447	443**	463		450	457	463	
SE	-5.2	-4.6	-5.1		-10.1	-5.3	-4.6	
N	106	329	201		106	329	201	
Rural	430	422***	431*	447	448	443	442	451
SE	-7.5	-2.6	-3.3	-5.8	-7.6	-4	-3.5	-5.1
N	375	771	805	116	375	771	805	116

Note: statistical significance is indicated relative to the wealthiest class. * significant at the 10% level ** significant at the 5% level *** significant at the 1% level

3.6   Discussion and concluding remarks

The impact of socio-economic status on student learning has been the focus of a considerable body of research in the field of economics of education. Due to the absence or lack of high-quality information on income or consumption particularly in low-income countries, researchers have had to resort to using information on assets to construct wealth indices. Owing to the seminal work by Filmer & Pritchett (2001), the most common approach, although itself not without challenges, has been principal component analysis (PCA). PCA's primary strengths are that it is computationally easier to work with, reduces the dimensionality of the data, while also using variables that are easily collected in household surveys (Jobson 1992).[13]

However, as Rutstein (2008) has shown, the way in which PCA assigns weights, especially in low-income countries (LCIs), tends to bias the wealth index towards urban households in its construction, leading to its inability to differentiate between the poorest of the poor and other poor households. For LICs like Malawi that have up to 81 percent of their population living in rural areas, this is a major challenge, as one of the primary reasons for constructing a wealth index is to be able to distinguish between these distinct shades of relative poverty.

---

[13]Note however, that unfortunately in some cases, ease of collection may not always coincide with alignment to the purpose for which one is collecting the data.

Unlike conventional solutions that attempt to overcome this limitation by finding more efficient weight assignment strategies (as in Rutstein (2008)), this chapter opted to use a clustering technique, — finite mixture modelling (FMM), — to distinguish between the different classes of socio-economic status at national, urban, and rural levels. This is presented as an alternative to the traditional cluster analysis tools that include dividing the wealth index by quintiles or some other arbitrarily chosen division of the data.

Specifically FMM was applied to the 2007 SACMEQ data with Malawi chosen as an example of a low-income country. Unsurprisingly, the results from the PCA show that the index constructed at national level is quite similar to that of the rural sample. This is likely owing to the disproportionately large segment of the population that lives in the rural areas. Further, FMM suggests that both the national and rural samples consist of four latent classes, while the urban sample consists of three. This means that there is more hetereogeneity in rural areas than in urban areas in respect of assets/wealth measures that are more applicable to urban contexts. With such a large rural population, it is not surprising that there is more heterogeneity identified compared to the smaller, urban sample.

As pertains to what mix of assets characterise these different classes, the findings indicate that both the urban- and rural-rich appear to enjoy greater access to assets that allow for entertainment and leisure, in addition to having greater access to the critical utilities of water and electricity. Furthermore, access to information plays a critical role in social mobility, with the poorest in both urban and rural settings also being the ones with the least access to information. Crucially, therefore, FMM succeeds in distinguishing between different wealth groups within urban and rural settings.

This notwithstanding, further research on the same could gain from applying stricter approaches to the choice of number of components, such as the ICL approach of Biernacki et al. (2000). In addition, more complex testing-based methods that typically use variants of the likelihood ratio test have also been recommended, as they circumvent the problems with regularity conditions that hinders the use of the standard LR test (Melnykov & Maitra 2010).

CHAPTER 4

Does technological growth destroy jobs and exacerbate wage inequality in middle-income countries?: Routine Biased Technological Change and the South African labour market

4.1 Introduction

As early as the 1970s research had begun observing an increased demand for high-skilled workers in many labour markets around the world. This observation came to be attributed to the skill-biased technical change hypothesis which postulated that this increased relative demand for high-skilled workers was due to advancements in technology which were biased in favour of skilled workers (see Katz & Autor (1999), Goldin & Katz (2007*b*, 2010)). However, this hypothesis did not predict the phenomenon observed across several developed countries' labour markets of shrinking employment shares and slower wage growth for middle-skilled workers. It has been argued that this was primarily driven by routine-biased technological change (RBTC) that decreased the need for workers to perform the kind of routine tasks that are typically performed by middle-skilled workers. In contrast, low-skilled and high-skilled workers perform a larger share of non-routine manual or cognitive tasks, both of which are harder to automate and replace with technological developments, thus raising their relative demand.

Against this backdrop, this chapter sets out to make the following contributions to the expanding literature on the effects of routine-biased technological change in developing countries. The first is to explicitly investigate the presence of routine-biased technological change in the South African labour market for a period spanning 18 years, from 1997 to 2015. South Africa remains one of the most unequal countries in the world with a Gini coefficient of 0.6 as of 2015 (Stats SA 2019) which in part manifests itself by a surplus of (mainly black) low- and/or unskilled- workers. The South African labour market is also characterised by high unemployment, a sizeable informal

sector, and wages that are influenced by trade unions and minimum wages, which differentiates it from the labour markets in developed countries, where most of the evidence for routine-biased technical change has been found. The second contribution is to investigate the effects of routine-biased technical change across demographic groups, economic sectors and occupations. We are particularly interested to see whether displaced middle-skilled workers are more likely to end up in low- or high-wage jobs, the informal sector, or unemployment, and whether this experience differs for men and women, and the population group of the worker.

The third contribution is to gauge the robustness of the results to using different measures of routine task intensity. So far, the norm in the literature has been to use US occupational schemes such as the Dictionary of Occupational Titles (DOT) to determine the task composition of an occupation which is then used to create a composite measure called the routine task intensity (RTI) (see for example Acemoglu & Autor (2011), Autor et al. (2006, 2003), Autor & Handel (2013)). Obviously, an RTI variable created using such US or other developed country's occupational schemes may not be contextually accurate for use in developing countries such as South Africa. This chapter instead uses data from the World Bank that specifically collects information on job-specific skills in developing countries. Regression analysis is then performed using an RTI created from a developing country context as well as an RTI using US occupational schemes.

The final contribution of this chapter comes from using non-linear systems estimation to determine whether the trends in employment and wages are consistent with labour demand that would be produced by technological progress. Following Klump et al. (2012), non-linear systems estimation is preferred over the standard production function approach for technical reasons that will be discussed in section 4.3 of this chapter. This approach produces estimates that suggest that the model discussed correctly accounts for technical change that leads to a hollowing out of occupations in the middle of the skills distribution, while simultaneously leading to an increase in the share of low-skilled and high-skilled jobs.

The chapter provides several interesting results. In section 4.5 we use descriptive statistics to investigate the trends in employment shares and wages. We observe the same hollowing out of occupational distribution that has been ascribed to RBTC in developed country labour markets. However, South Africa experienced a larger increase in low-wage occupations and a smaller in-

crease in high-wage occupations than is typically observed in developed countries. These occupational shifts are restricted to the private sector; public sector employment appears not to have responded to changes in the relative productivity of workers. The distribution of occupational movement in the private sector is highly correlated to race and gender: upward movement is much more likely for males, and for the white and Indian population groups whereas downward movement is much more likely for females and black workers. We also find that most workers who lose their middle-wage jobs appear to be absorbed in either low-or high-wage occupations, and there is no indication of large displacement into informal employment, unemployment or economic inactivity.

A decomposition of occupational shifts indicates that middle-skilled jobs are disappearing mostly due to within-industry changes in how production occurs, and that these tasks are increasingly performed by lower-skilled workers. However, unlike what has been observed elsewhere, there is no displacement of middle-skilled workers to high-skilled occupations within industries. Instead, the increasing share of high-skilled workers appears to be driven completely by between-industry shifts in which industries that have historically employed a large share of high-skilled workers (e.g. financial or social services) are expanding. Wage trends confirm the reduced demand for middle-skilled occupations and increased demand for high-skilled occupations. Unlike what has been observed in developed country labour markets, the increase in wages of low-skilled occupations has not increased very much, possibly due to the large supply of unemployed workers who can fill these jobs.

In section 4.6, regression analysis is used to investigate whether routine-biased technical change is the most likely explanation for these wage and employment trends. We use a measure of the RTI of different occupations collected from surveys from other developing countries, and confirm that occupations with high RTI experienced slower wage growth and decreasing employment shares. This suggests that the observed trends in wages and employment were indeed driven by the automation of routine tasks. Our analysis also finds that applying a measure of RTI from the US to the South African data – an approach that has been followed in several other developing countries - produces either highly attenuated regression coefficients or misleading conclusions. Finally, we use a non-linear system estimator to jointly model employment and wage trends as equilib-

rium outcomes, assuming that production occurs according to a constant elasticity of substitution function. This model fits the data well, and suggests that middle-skilled occupations experienced slower productivity growth than low- and high-skilled occupations over this period, exactly as predicted by the RBTC model.

The remainder of this chapter is organised as follows. Section 4.2 presents a review of the literature, specifically discussing the two main hypotheses that inform economic thinking about skill composition and migration in the labour market: skills-biased technical change, and routine-biased technical change. Section 4.3 discusses the model and section 4.4 describes the data. Section 4.5 presents the trends that characterise employment and wages for each of the three skill groups between 1997 and 2015. Section 4.6 provides results that verify the model's ability to explain the observed changes in the employment shares across occupations. Finally, Section 4.7 presents the conclusion.

4.2 Literature review

4.2.1 Skills-biased technical change

After several decades during which global income inequality decreased, wage inequality suddenly started rising around the late 1970s (Katz & Autor 1999). Research suggests that this increase was mainly driven by an increase in the wage gap between workers of different skill and occupation classes (Autor et al. 2005, Katz & Autor 1999, Autor et al. 1998). Over time, several hypotheses have been put forward to explain this troubling trend. Wood (1998) emphasised the role of rising international trade that incentivised the offshoring of parts of the production process which decreased demand and hence wages for unskilled workers in developing countries. On the other hand, Katz & Murphy (1992) looked at the slower increase in the relative supply of more educated workers from the 1970s to the 1980s. Additional explanations focused on factors such as wage-setting norms and institutions, the decline in unions, and the real value of the minimum wage (DiNardo et al. 1996). Among the competing explanations, one that garnered particular attention from many labour market analysts viewed wage inequality as a result of increased relative demand and utilisation of more-skilled workers due to greater adoption of technology that was skill-biased (Bound & Johnson 1992). This phenomenon has come to commonly be referred to as

skill-biased technical change (SBTC) (Katz & Autor 1999).[1,2]

The basic idea of SBTC is that technological advancements, primarily in information and communications technology (ICT), favour skilled (e.g. more educated, more experienced) workers compared to unskilled workers, thereby increasing their relative productivity and relative demand (Tinbergen 1974, 1975, Violante 2008). For illustrative purposes, consider the standard aggregate production function presented in equation 4.1. Apart from ignoring capital, it conceptualises output, Y, as being a constant elasticity of substitution (CES) function of skilled and unskilled labour, L_s and L_u, with factor-specific productivities α_s and α_u:

$$Y = [(\alpha_s L_s)^\sigma + (\alpha_u L_u)^\sigma]^{1/\sigma}, \quad \sigma \leq 1 \tag{4.1}$$

As in Violante (2008), it can be shown that the log of the marginal rate of transformation (MRT) between the two labour inputs is given by

$$\ln(MRT_{s,u}) = \sigma \ln\left(\frac{\alpha_s}{\alpha_u}\right) + (1 - \sigma)\ln\left(\frac{L_u}{L_s}\right) \tag{4.2}$$

Technical change is said to be skill-biased if α_s/α_u increases over time.[3] Since skill-biased technology increases the relative productivity of skilled labour, its demand increases relative to that of unskilled labour, *ceteris paribus*. Indeed, part of the appeal of a skilled workforce is its relative complementarity to advancements in ICT (Autor & Handel 2013, Górka et al. 2017). To this end, studies have confirmed that firms that adopt more technologies and invest more in capital are also more inclined to hire more highly skilled workers (Fernandez 2001, Bartel & Lichtenberg 1987, Greenwood & Yorukoglu 1997).[4,5]

One plausible explanation for this is that large investments in ICT often lead to increased worker independence and changes in organisational practices, both of which increase the need for more

---

[1]The same phenomenon is also referred to as skill-biased *technological* change (see for example Goos et al. (2014).
[2]For surveys describing the foundation for SBTC in economic models see Acemoglu (2002) and Hornstein et al. (2005).
[3]This is true under the empirically plausible parametric assumption that $\sigma > 0$.
[4]For example, Krusell et al. (2000) discusses suggestive evidence that rapid increases in the capital stock beginning in the 1960s, together with its strong complementarity to skilled labour, contributed to the observed increase in relative demand for skills.
[5]For a clear and comprehensive early discussion of capital-skill complementarity, see Griliches (1969, 1970).

highly educated workers (Bresnahan et al. 2002). In this way, physical capital and advancements in technology are understood to be relative complements with high-skilled workers. The SBTC model therefore implies that technology has a monotonically upgrading effect on the occupation structure in terms of skills: as new technologies are adopted, labour demand for higher skills also increases and this amplifies wage inequality between the skill types (Raquel & Biagi 2018, Autor & Dorn 2013).

Even though the findings on SBTC are compelling, it is worth bearing in mind that much of this evidence is based on data from developed countries such as the US, Japan and parts of Europe. Similar evidence for low- and middle-income countries is more sparse. There is some evidence that India began to show trends consistent with SBTC in the 1990s (Berman et al. 2005), possibly as a result of a manufacturing boom which increased output and capital-skill complementarity. Similar SBTC-consistent trends have been reported for other middle-income countries like South Korea (Lee & Sim 2016) and Brazil (Nogueira 2015).[6]

In the case of South Africa, a seminal contribution was Bhorat & Hodge (1999). This paper looks at trends from 1970 to 1994 and showed a deepening of capital accompanied by increasing demand for high-skilled workers which is similar to the trends reported in developed economies. Disaggregating these trends further by sector revealed insightful findings. For example, even though the entire economy experienced an increased capital-labour ratio of 142 percent, it was primary sectors like agriculture (168.8 percent) and mining (416 percent) that experienced the most capital deepening. Interestingly, further analysis showed that while overall employment in these primary sectors fell, it was specifically demand for unskilled labour which declined while relative demand for skilled labour rose in the same period (Bhorat & Hodge 1999). Similar trends were observed within both the secondary and tertiary sectors where employment shifts were also characterised by increasing demand for skilled relative to unskilled labour. However, these trends were smaller in magnitude compared with the size of employment shifts experienced in the primary sector.

Even though between-sector shifts in labour demand contributed to the change in employment, a within and between sector decomposition revealed that the observed employment shifts could

---

[6]There are some multi-country studies that find a growing wage gap in several low- and middle-income countries (Meschi & Vivarelli 2009, Conte & Vivarelli 2011). However, the breadth of these studies means that the data for each country is investigated less thoroughly than is typical in papers that have looked at developed countries.

predominantly be explained by changing labour demand within sectors: sectors were employing more skilled and fewer unskilled workers. This finding was likely an artefact of the increased capital deepening which changed production methods in ways that increasingly favoured skilled labour (Bhorat & Hodge 1999).

More recent research shows that there has not only been a continuation but also an intensification of the skill-biased labour demand trajectory established and observed in the pre-1994 period (Bhorat et al. 2014).[7] This is a particularly undesirable finding for the South African labour market for two reasons.

First, South Africa has a surplus of unskilled workers which means even in the absence of SBTC, this pushes down their wages and makes it difficult for them to find jobs. An intensification of SBTC would therefore marginalise a large group of South African workers. Secondly, a continuation and intensification of SBTC would disproportionately harm black workers who were previously already disadvantaged by discriminatory policies enacted during the apartheid era. In particular, education policies which precluded black workers from accessing high quality schools and tertiary institutions led to a surplus of black unskilled workers who — in the face of SBTC — would be unable to adapt to the adoption of new skill-biased technologies.

4.2.2 From skills-biased technical change to routine-biased technical change

As a canonical model, SBTC has proven empirically successful in accounting for growth in the demand of high-skilled relative to low-skilled workers in developed economies. Despite these merits, it was not without its critics (Leamer 1996, Card & DiNardo 2002). Arguably its biggest weakness is that the model does not account for the more recent phenomenon of job polarisation: the shrinking of middle- relative to high- and low-skilled occupations (Autor et al. 2003, Goos & Manning 2007).

In response, Autor et al. (2003) put forward a more nuanced and refined version of SBTC called routine-biased technological change.[8] The novel contribution of this theory is that it views the

---

[7]In fact, in addition to trade liberalisation, Dunne & Edwards (2006) credit a majority of the job losses between the 1994 to 2003 period to skill-biased technological changes.

[8]Noteworthy revisions to the model have since been made by Goos & Manning (2007), and later by Acemoglu & Autor (2011).

production process as consisting of a number of different tasks. Technological progress therefore allows firms to perform some tasks with less worker inputs, and for other tasks to be completely automated. The model then hypothesises that recent technological change is biased towards replacing labour in routine tasks as such tasks are uniquely susceptible to replacement by computer based systems. Such tasks can be classified as "codifiable".

According to Autor & Handel (2013) tasks are considered to be of three main types: manual (typically requiring physical exertion and direct personal interaction); abstract (e.g. soft skills or tasks requiring intuition, reasoning or some form of higher-level cognitive exertion); and routine.[9] Manual tasks are typically performed by low-skilled workers, routine tasks by middle-skilled workers, and abstract tasks by high-skilled workers. Several studies have confirmed that the shrinking of the share of middle-skilled occupations is indeed driven by the disappearance of occupations with a high RTI in developed economies, including the US (Autor & Handel 2013, Autor & Dorn 2013, Autor et al. 2006), UK (Goos & Manning 2007), and parts of Europe (Spitz-Oener 2006, Matthes et al. 2014, Fernández-Macías & Bisello 2016).

While early evidence of RBTC was mostly based on studies and data from developed economies, similar, more recent studies, have also been carried out in a number of developing country contexts, though notably not on the same scale. Indeed, there is limited empirical evidence consistent with RBTC in developing countries (World Trade Organization 2017) despite the fact that developing countries are on average significantly less exposed to routinisation than their developed country counterparts (Das & Hilgenstock 2018). One such study found that the occupational distribution in South Korea was characterised by a decline in middle-skilled occupations accompanied by increases in low- and high-skilled occupations (Kim et al. 2019) as predicted by RBTC.[10] Similar findings have also been reported for Brazil between 1996 and 2006 (Almeida et al. 2017). In both these cases, the authors used RTI index measures linked to US-based occupational schemes to confirm that occupations in routine-intensive occupations were more vulnerable to job loss.[11] In a broader overview, different cross-country studies have produced mixed findings, some con-

---

[9]Initial work on the tasks framework by Autor et al. (2003) presented five categories of occupations, and it is only in more recent work that that has been scaled down to three, as in Goos et al. (2014), Acemoglu & Autor (2011), and Autor & Dorn (2013).

[10]Note that even though South Korea gave up its 'developing' country status towards the end of 2019, it was still classified as a developing country for the total number of years included in the study, 1993 to 2015.

[11]Additional evidence is also presented for Colombia and Mexico (Ariza & Bara 2020, Medina & Suárez 2010).

firming the presence of employment polarisation trends that are consistent with RBTC and others not (Maggie Fu et al. 2020, Longmuir et al. 2020). In fact, in some of these developing countries, it appears that adoption of technologies disproportionately benefits those with middle to advanced levels of education at the expense of those with less education.

Evidence of RBTC for South Africa itself is limited. This does not imply, however, that the features and consequences consistent with RBTC itself have not been reported. Indeed, even though the actual term routine-biased technical change has not been used in the South African literature, the trend of a hollowing-out of the skills distribution has been documented (Bhorat & Hodge 1999, Bhorat et al. 2014).

A notable contribution in the South African literature that actually looks at the task composition of jobs is Bhorat et al. (2014). Though the study looked at a shorter time period (2001-2012), the authors find evidence of a shift in labour demand towards higher- and middle-skilled workers within the tertiary sector, while the share of low- and middle-skilled workers within the primary and secondary sectors declined. Further, as in the pre-1994 period, they also show that these shifts were driven disproportionately by within-sector changes in labour demand. In addition, they use regression analysis to show that on average, jobs that consist of routine tasks have experienced a decline in wages over the 2001-12 period. In doing so, they measured the task content of work by identifying five task categories using occupation codes in South African Labour Force Survey data. Although an important contribution, this paper suffers from a few shortcomings. Firstly, the way in which the RTI of occupations is identified is not as reliable. The authors link every occupation to one or more of five task categories that best describe the kind of work that occupation actually does.[12] This is problematic because not only can an occupation fall into two or more categories, but also, the categories are based on the work of Firpo et al. (2011) who constructed them to specifically describe work in a US setting and not in a developing country. Secondly, it does not explicitly test whether observed trends in wages and employment are consistent with RBTC. Lastly, it does not attempt to specifically estimate the growth in productivity of the three types of labour (manual, routine and abstract).

---

[12]These categories are: information and communication technology (ICT), automation/routinisation, face-to-face, on-site, and decision-making/analytic.

79

As alluded to above in the case of Bhorat et al. (2014), it is also worth discussing the data that different studies have used to measure the RTI of different occupations. The norm in the literature has been to use occupation schemes from one country, often the US-based DOT, or its predecessor the Occupational Information Network (O*NET) to calculate a composite routine task measure for different occupations. It is this US-based measure that was used to test for the existence of RBTC in the US (Autor et al. 2003, Autor & Handel 2013) and at least 16 European countries (Goos & Manning 2007, Fonseca et al. 2018, Anghel et al. 2014).[13]

Using this measure of RTI makes sense when studying the US labour market. It may also provide a plausible approximation of RTI in other developed labour markets, provided that the task content of occupations is quite similar to those in the US. However, such an extrapolation is less appropriate when applying an RTI measure calculated using occupation schemes from a rich country such as the US to investigate effects of RBTC in a relatively low- to middle-income country like South Africa. Indeed, this US-based measure may provide a very misleading indication of the routine-task intensity of occupations. In fact, this approach has been critiqued by Raquel & Biagi (2018), who hypothesised that this may lead to results that are driven by the choice of data used to calculate the RTI and may consequently introduce measurement error that biases the estimates. One of the contributions of this chapter will be to gauge the robustness of results to using different RTI measures.

4.3 A framework for understanding job-polarisation: The CES production function

The choice of a production function that models how profit-maximising firms combine a mix of inputs is based on theoretical properties such as consistency, flexibility, factual conformity and computational facility. The three functional forms most commonly used in the literature, Cobb Douglas, CES (constant elasticity of substitution) and Translog, do not respectively meet all these properties. Consequently, in selecting a production function, there has to be a trade-off among

---

[13]Although DOT/O*NET is used by the majority of empirical studies in this literature, there are various other less-widely used occupation classification schemes as well. Examples would be the Princeton Data Improvement Initiative Survey (PDII) (Autor & Handel 2013), the Programme for the International Assessment of Adult Competencies (PI-AAC) (Marcolin et al. 2016) and the European Working Condition Survey (EWCS) (Fernández-Macías & Hurley 2017, Sebastian 2018).

these properties. Following Klump et al. (2007) and León-Ledesma et al. (2010), this paper uses the CES production function, which has become the workhorse for both theoretical and empirical research on RBTC.

Unlike the Cobb Douglas production function, the CES and Translog functions have some flexibility in the estimation of elasticities by allowing for substitutability among factors of production. However in empirical modelling, the Translog function may produce concerns of multicollinearity in the inputs and their quadratic transformations. By normalising the CES production function as discussed in section 4.3.2 of this chaper, concerns of potential shortcomings that may arise when jointly estimating factor productivities and elasticities are addressed. The ability to leverage its desirable properties once normalised sets the CES production function apart as a preferred modelling tool as most of the literature has motivated (Saam 2014, Klump et al. 2007).

4.3.1 The general CES production function

The first formal formulation of the CES production function owes itself to seminal work by Arrow et al. (1961).[14] This formalised how economists understood firms to combine factors of production capital (K) and labour (L) to produce output (Y). The standard CES production function can be expressed as

$$Y_t = F(K_t, L_t) = \left[\pi K_t^{\frac{\sigma-1}{\sigma}} + (1-\pi) L_t^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{4.3}$$

where $\pi \in (0,1)$ is the distribution parameter and represents capital's share of total output and σ is the elasticity of substitution between capital and labour. As in all standard CES production functions, Equation 4.3 above is Cobb Douglas when $\sigma \to 1$; Leontif when $\sigma = 0$; and linear when $\sigma \to \infty$.

There are various reasons why σ, the elasticity of substitution, is of particular significance to economists.[15] In the two-factor case as in Equation 4.3, σ is defined as the change in the ratio of capital and labour (K/L), with respect to the marginal rate of substitution between K and L. That is, σ is the percentage change in the proportions of the factors, owing to a change in the marginal rate of technical substitution along an isoquant (Jehle 2010, Klump et al. 2012). Formally, we can

---

[14]For a comprehensive discussion of the CES production function, see De La Grandville (1989).

[15]This is in part due to its importance in generating perpetual economic growth as demonstrated in neoclassical growth models for values of elasticity above unity (Klump et al. 2012).

define the elasticity of substitution as:

$$\sigma \in [0, \infty] = \frac{d(K/L) \, / \, (K/L)}{d(F_L/F_K) \, / \, (F_L/F_K)} = \frac{d\log\,(K/L)}{d\log\,(F_L/F_K)} = \frac{d\log\,(K/L)}{d\log\,(w/r)} \tag{4.4}$$

As mentioned in Klump et al. (2012), the elasticity of substitution is implicitly always represented and defined as a point elasticity, meaning that it relates to a particular baseline point on an individual isoquant. When the point elasticity changes, the old and new isoquants are non-intersecting but tangent at the old elasticity. At this point, the old and new CES production functions still possess identical factors of production and a marginal rate of technical substitution even though the isoquants themselves are non-intersecting. A higher measure of elasticity represents the ease with which one factor of production can be substituted for another, and a low measure represents the difficulty of substituting one factor of production for another.

This chaper builds on an extension of the standard CES production function presented in Equation 4.3 by assuming that physical output Y is produced using a combination of low-skilled (or manual) labour, L_l; middle-skilled (or routine) labour, L_m; and high-skilled (or abstract) labour, L_h. All goods and services are produced by firms in a perfectly competitive market, and each industry uses only labour as its input in production with the following production function:

$$Y_t = \left[\pi_l (L_{l,t})^{\frac{\sigma-1}{\sigma}} + \pi_m (L_{m,t})^{\frac{\sigma-1}{\sigma}} + \pi_h (L_{h,t})^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{4.5}$$

where π_l, π_m, and π_h are the distribution parameters for low-, middle-, and high-skilled labour respectively.[16] The CES functional form has several benefits over alternative specifications when modelling the production process, and has become the standard in the theoretical literature on skill-biased and routine-biased technical change (Saam 2014). However, the CES poses some well-documented challenges in empirical analysis, especially when attempting to simultaneously estimate factor productivities and elasticities. Fortunately, using Monte Carlo simulations, it has been shown that a normalised systems approach is superior to other methods in identifying highly non-linear CES functions (León-Ledesma et al. 2010).

---

[16]This labour-only form of the CES production function has been used in the literature on several occasions (Harrigan et al. 2018, Bøler 2015, Bárány & Siegel 2019).

4.3.2 Normalisation of CES function with factor augmenting technical progress

The non-normalised CES production function tends to yield results that have no economic meaning. As Klump et al. (2012) explains, the reason for this is that the production function estimates are themselves dependent on both the normalisation point and the elasticity of substitution. This feature crucially undermines estimation.

Essentially, normalisation is the fixing of a baseline point, thought of as a point at which $t = t_0$, which is characterised by specific values of not only all the factors of production (L_l, L_m, and L_h) but also the distribution parameters. At this point, the isoquants of CES functions with different elasticities of substitution but with all other parameter values equal are tangent. In this way, normalisation yields a realisation of a family of CES production functions whose only difference is the elasticity of substitution but share the same baseline point. Additionally, as Klump et al. (2012) points out, normalisation represents the production function in consistent indexed number form which crucially implies that the variables present in the production function become of the same dimension. This consistent indexing of the variables also importantly produces estimates that are invariant to changes in units of measurement. Following Klump et al. (2007, 2012) and León-Ledesma et al. (2010) the normalised form of Equation 4.5 above is as follows:

$$\frac{Y_t}{Y_0} = \left[\pi_{l,0} \left(\frac{L_{l,t}}{L_{l,0}} \right)^{\frac{\sigma-1}{\sigma}} + \pi_{m,0} \left(\frac{L_{m,t}}{L_{m,0}} \right)^{\frac{\sigma-1}{\sigma}} + \pi_{h,0} \left(\frac{L_{h,t}}{L_{h,0}} \right)^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{4.6}$$

for values of $L_{i,0} \in \{l, m, h\}$; $\pi_{i,0} \in \{l, m, h\}$, $t = t_0$, and Y_0 which are the values used to normalise Equation 4.5. It is common in the literature to choose the point of normalisation as the mean values over the full sample, so that $Y_0 = \overline{Y}$, $L_{l,0} = \overline{L_l}$, $L_{m,0} = \overline{L_m}$, $L_{h,0} = \overline{L_h}$, $t_0 = \overline{t}$ and $\pi_{i,0} = \overline{\pi_i}$; where the bar values refer to averages (geometric means for growing variables such as the factors of production, and arithmetic means for approximately stationary variables such as the labour shares of each labour type). As in Klump et al. (2012), the actual normalised log-transformed CES production function with factor-augmenting technical progress to be estimated takes the following form:

$$\log\left(\frac{Y_t}{Y_0} \right) = \frac{\sigma}{\sigma-1} \log\left[\pi_{l,0} \left(e^{\gamma_l(t-t_0)} \frac{L_{l,t}}{L_{l,0}} \right)^{\frac{\sigma-1}{\sigma}} + \pi_{m,0} \left(e^{\gamma_m(t-t_0)} \frac{L_{m,t}}{L_{m,0}} \right)^{\frac{\sigma-1}{\sigma}} + \pi_{h,0} \left(e^{\gamma_h(t-t_0)} \frac{L_{h,t}}{L_{h,0}} \right)^{\frac{\sigma-1}{\sigma}} \right] \tag{4.7}$$

83

where in addition to the variables presented in Equation 4.6, γ_l, γ_m and γ_h denote growth in productivity of the three types of labour: low/manual, middle/routine and high/abstract. From this, one can derive the profit-maximisation conditions for the three types of labour as follows:

$$\log(w_{l,t}) = \log\left(\pi_{l,0}\frac{Y_0}{L_{l,0}}\right) + \frac{1}{\sigma}\log\left(\frac{L_{l,0}}{Y_0}\right) + \frac{1}{\sigma}\log\left(\frac{Y_t}{L_{l,t}}\right) + \frac{\sigma-1}{\sigma}(\gamma_l(t-t_0)) \tag{4.8}$$

$$\log(w_{m,t}) = \log\left(\pi_{m,0}\frac{Y_0}{L_{m,0}}\right) + \frac{1}{\sigma}\log\left(\frac{L_{m,0}}{Y_0}\right) + \frac{1}{\sigma}\log\left(\frac{Y_t}{L_{m,t}}\right) + \frac{\sigma-1}{\sigma}(\gamma_m(t-t_0)) \tag{4.9}$$

$$\log(w_{h,t}) = \log\left(\pi_{h,0}\frac{Y_0}{L_{h,0}}\right) + \frac{1}{\sigma}\log\left(\frac{L_{h,0}}{Y_0}\right) + \frac{1}{\sigma}\log\left(\frac{Y_t}{L_{h,t}}\right) + \frac{\sigma-1}{\sigma}(\gamma_h(t-t_0)) \tag{4.10}$$

Following the literature, it is assumed that factors of production (types of labour in this case) are paid their marginal revenue (or wage) which allows the distribution parameters π_l, π_m, and π_h to be replaced by an estimate of the labour income share.[17] This assumption is crucial for the identification of the parameters in the model as it allows for the systems approach to estimation.[18]

As in Klump et al. (2007), León-Ledesma et al. (2010) and Saam (2014), the importance of Equations 4.8-4.10 is in allowing the estimation procedure to take advantage of a normalised systems approach which allows for simultaneous identification of factor-augmenting technical change and the elasticity of substitution. This also helps in identification of σ because it provides additional degrees of freedom in estimation, as well as the application of cross-equation parameter restrictions (León-Ledesma et al. 2010).[19] This chapter estimates the system of equations using Stata's non-linear seemingly unrelated regressions command which employs a two-step feasible generalised least squares estimator (StataCorp 2013). This approach is used to estimate the model's

---

[17]This replacement is usually justified on the assumption that competitively determined wages will equal the marginal revenue product of workers. This is a particularly heroic assumption to make in the South African case where unemployment is high and wages are affected by institutional factors such as trade unions, minimum wages and social norms. Although wages are unlikely to be exactly equal to worker productivity in South Africa, wages are still a very important indicator of the relative productivity of workers. This information is particularly useful when attempting to estimate the growth in productivity of different worker categories using a systems estimation approach. We therefore follow the theoretical and empirical literature that commonly uses the profit-maximisation conditions when modelling RBTC.

[18]By construction, a further "restriction" imposed on the model, is that the distribution parameter for all three types of labour sum to unity, that is; $\pi_{l,0} + \pi_{m,0} + \pi_{h,0} = 1$.

[19] León-Ledesma et al. (2010) shows that the profit maximising first order conditions can be used directly, while Klump et al. (2007) makes a case for the joint identification of σ and other parameters being improved through use of factor shares in estimating production functions.

three parameters of interest: γ_l, γ_m and γ_h.

4.4   Data Sources

4.4.1   Employment and wages

This study uses three sources of data. The first is the PALMS dataset created by DataFirst at UCT. The PALMS was created by harmonising nationally representative microdata from 61 household surveys conducted by Statistics South Africa (StatsSA) between 1995 and 2017 (Kerr et al. 2013). It consists of the October Household Surveys (OHS) (1994-1999), the bi-annual Labour Force Surveys (2000-2007) — including the smaller LFS pilot survey from February 2000 — and the 2008 to 2017 Quarterly Labour Force Surveys. These surveys include individual responses to questions regarding wages, employment status, and industry and occupation of employment. Occupations are largely classified using the internationally recognised four-digit method prescribed by the United Nations' International Standard Classification of Occupations; the ISCO-88.[20] Although the PALMS data has been harmonised with respect to most labour market variables, the codes used to classify occupations are internally inconsistent and therefore require some data wrangling. In order to use this variable to compare occupations over time it had to be reclassified to be consistent across all surveys.

As a starting point, while the occupation variable for the OHSs from 1997 to 1999 was consistently classified using ISCO-88, from 2000, StatsSA officially migrated from using the United Nations' ISCO-88 classification to their own method, the South African Standard Classification of Occupations (SASCO). Motivated by the structural changes in the South African labour market, this naming convention was created to serve as the basis for all future occupational classification work.

Unlike later iterations of SASCO such as SASCO 2012, the first version of SASCO (released in 2003) on which StatsSA's LFS and QLFS data were based, was developed using a similar conceptual basis to that of the United Nations' ISCO-88. As a result, the initial SASCO 2003 version

---

[20]The ISCO was developed by the International Labour Office (ILO) in Geneva in 1957 and has had several iterations: ISCO-58, ISCO-68, ISCO-88, and the most recent 2007 version, the ISCO-08. The main reason for the different versions is to have a classification of occupations that takes into account developments in the world of work.

is very similar to ISCO-88 (StatsSA 2012). This study uses the OHS's 1997 to 1999 occupation variable as the base category. In that way, the remaining occupation variables in the sample are reclassified to resemble the ISCO-88 format. Creating this harmonised variable involved a side-by-side comparison of more than 1 300 individual four-digit occupations in the OHS and matching them to the SASCO 2003 occupations in the LFS/QLFS series. In the vast majority of cases, both the occupational label and occupational code remained the same across SASCO and ISCO codes. However, in some cases the occupational labels or codes changed over time. This yields two kinds of inconsistencies.

The first has to do with codes that refer to new or different occupations in SASCO-03 and ISCO-88. For those that refer to new occupations, in some cases this required making informed judgements about what such occupations included in the SASCO handbook could have been in the old ISCO method of classifying occupations. In other cases, the same occupation exists in both SASCO and ISCO and is merely classified by a different code in SASCO-03 and ISCO-88. For example, the occupation "bootlegger" is classified with the code 9998 in SASCO, but with the code 5230 using the ISCO-88 classification. Making such occupations consistent is then simply achieved by recoding SASCO occupations with ISCO's codes.

The second type of inconsistency arose within the OHS ISCO-88 classification itself by having occupation codes that were not present in SASCO. For example, ISCO-88 had two codes that were not present in the SASCO-03 classification: 3500 (such as lieutenant, or captain under the associate professionals in armed forces and civil service minor group), and 1500 (such as brigadier and lieutenant general under the managers in armed forces and civil service minor group). These are then both collapsed into ISCO-88's 1120 code which is designated for senior government officials. The motivation is that in merging the LFS's codes to those of the OHS, a lot of occupations under ISCO-88's 3500 are in SASCOs 1120 code.[21] Dealing with these two kinds of inconsistencies led to changes in four occupation codes, thus creating a consistently defined occupation variable for all the survey data used in this chapter.

The sample used for the empirical analysis in this chapter is limited to employed individuals of

---

[21]Possibly, SASCO did this because such occupations are indeed senior government officials though 3500 is solely for armed forces and civil service.

working age with wage values greater than zero and non-missing data for hours worked. Total (or aggregate) employment is calculated from individual responses to questions pertaining to employment and their respective survey weights.[22] In addition, informal sector workers were omitted. This omission is motivated by research that looks at potential areas of measurement error that may arise due to the comparison of StatsSA household surveys that underwent modifications in questionnaire design and sampling methodology over time (Kingdon & Knight 2007, Altman 2008, Casale et al. 2005, Burger & Yu 2007). The research shows that most of the inconsistencies that arise from comparing these StatsSA surveys are limited to the informal sector and therefore removing it from the sample mitigates this potential for measurement error significantly. The implication of this omission is that our analysis is restricted to the effect of RBTC on the formal labour market.

4.4.2 Output

For industry level output data, this study follows Burger & Teal (2015) in using output data taken from the South African Reserve Bank (SARB) and merging it with PALMS employment data to construct a balanced South African industry panel. This output data is collated and made publicly available by Quantec.[23,24] Included in the data are the variables for "gross value added by kind of economic activity" which are used as the measure of industry output. This variable is categorised using the one-digit Standardised Industrial Classification (SIC) method which leads to a categorisation of nine industries.[25] Output was measured at 2010 constant prices.

4.4.3 Task content

The literature on RBTC is relatively young and continually expanding, so it is not surprising that some unresolved conceptual issues remain.[26] One contentious issue regards how to measure the routine task intensity of occupations. This issue arises partly from the fact that informa-

---

[22]PALMS provides cross-entropy weights which are more consistent over time than the sampling weights provided by Stats SA (Branson & Wittenberg 2014).

[23]The data can be accessed from their website (https://www.quantec.co.za).

[24]The output data series can also be obtained from StatsSA's P0441 — Gross Domestic Product (GDP) series.

[25]These are (1) agriculture, forestry, and fishing, (2) mining and quarrying, (3) manufacturing, (4) electricity, gas, and water, (5) construction, (6) wholesale and retail trade, catering and accommodation, (7) transport, storage, and communication, (8) finance, insurance, real estate, and business services, and (9) community, social, and personal services.

[26]For a more extensive discussion of these issues, see Raquel & Biagi (2018).

tion on tasks is not commonly collected by representative surveys, which means that researchers face binding data limitations when choosing the best measure of routine task intensity. What researchers have typically done is to map the occupation variable codes to a database that acts as a source of information on task content of occupations. An early example of this approach is Autor et al. (2003) who created an RTI from task measures obtained from the DOT and applied this measure to US labour market data.[27] Given the success of this approach, other researchers proceeded to use this same data on tasks to study the labour markets of countries other than the US which is potentially problematic especially when applying this measure in the context of a developing country.

In 2012, the World Bank launched an initiative called the Skills Towards Employability and Productivity (STEP) Measurement Program. The STEP survey is a first-of-its-kind attempt to collect and systematically measure information on job-relevant skills in developing countries. In addition to conceptually drawing inspiration from the very influential DOT, the STEP survey items are drawn from the survey of Skills, Technology, and Management Practices (STAMP) which is a two-wave nationally representative survey of US wage and salary workers (Pierre et al. 2014). Items were then chosen and adapted to developing country contexts in various ways. For example, the survey design team made sure to include skills that are relevant to rural agriculture, which is an important economic activity in developing countries. This program was rolled out in 12 low- and middle-income country (LIC and MIC) contexts.[28] By its unique design, STEP provides a dataset that can be used more directly as a measure of skill and task content in developing countries. Specifically, this is achieved by three innovative modules aimed at collecting data on cognitive, technical, and non-cognitive skills.

Using information from these modules, researchers from Wageningen University meticulously created four task groups that can be used to understand and aggregate the type of occupations workers engage in. These are: routine (including cognitive and manual); non-routine manual; non-routine analytic; and non-routine interactive. For each of these four task groups, they created

---

[27]One important feature of the DOT is that it divides job-relevant skills according to their level of involvement with "data, people, and things" which conveniently correspond to cognitive, interpersonal, and manual skills making it easier to categorise skills accordingly (Pierre et al. 2014). Further, note also that more recent work tends to opt for the O*NET, which is a successor to the DOT.

[28]These 12 countries are : Armenia, Azerbaijan, Bolivia, Colombia, Georgia, Ghana, Kenya, Lao PDR, Macedonia, Ukraine, Vietnam, and Yunnan Province in China.

two versions of routine task intensity. One is an occupational mean z-score across all individuals in the STEP data, which includes all 12 LIC and MIC countries. The other is an occupational mean z-score across individuals in Colombia and Macedonia, which are the two richest countries in the data and the only two with GDP per capita levels close to that of South Africa. Following Autor & Handel (2013), these variables are used to create a composite measure of routine, the Routine Task Index (RTI). This data is then merged with South African occupational codes in PALMS at the two-digit level, making it possible to analyse the impact/return of routine on wages. In addition, this chapter also presents wage regression results that used an RTI index developed by Goos et al. (2014) based on the widely used U.S DOT task measures.[29] This dataset was also merged with the South African employment data by matching the 2-digit ISCO-88 occupation codes. In section 4.6.1, all three of these measures will be used to estimate the effect of RBTC on South African wage trends, which allows us to gauge the effect of using RTI variables derived from countries with dissimilar labour markets.

4.5 Descriptive Statistics

The model of RBTC makes a very clear prediction about the evolution of employment and wages for the different occupational classes. Individuals who engage in highly routine occupations are expected to experience a decrease in wages and in their employment share due to declines in their relative productivity as their tasks become increasingly performed by technological advancements. Since these highly routine occupations are typically found in the middle of the wage and skills distribution, this predicts a decrease in wages and employment for occupations in the middle of the wage distribution. By the same token, the model predicts that low- and high-skilled workers should see an increase in wages and employment share. Finding descriptive results that are in favour of these trends is prima facie evidence of the existence of RBTC in the South African labour market. In section 4.5.1 this chapter discusses the employment trends of the skill classes, followed by a discussion of the wage trends in section 4.5.2.

---

[29]It is in fact identical to the index used in such papers as Autor & Handel (2013), Autor & Dorn (2013), Mahutga et al. (2018)

4.5.1 Trends in employment

4.5.1.1 *General occupation trends*

This section provides an overview of the overall changes in employment shares. As in Bhorat et al. (2014), skills are of three types: low, middle and high. In order to arrive at these three, this chapter follows Goos et al. (2014) who classify the occupations by ordering the 27 unique two-digit occupations by their mean wage. Ranking them in this way leads to high-skilled/paying occupations that are made up of: the top three high-earning managerial occupations (11 to 13), professional occupations (31 to 34), and associate professionals (41 and 42); middle-paying occupations that consist of clerical occupations (41 and 42), crafts and other related occupations (71 to 74), and plant and machine operators and assemblers (81 to 83); and lastly, low-paying occupations that are made up of sales, personal, protective, and agricultural workers (51,52,61 and 62), labourers in agricultural, mining, construction, and manufacturing related work (92,93) and other low-skilled sales and services occupations (91). Coincidentally, this classification matches that of Goos et al. (2014).

The descriptive analysis starts by looking at the changing composition of employment in the private sector—where market forces are most likely to quickly reflect changes in labour demand— between 1997 and 2015. Figure 4.1 specifically reports the trends in employment shifts for private sector workers.

90
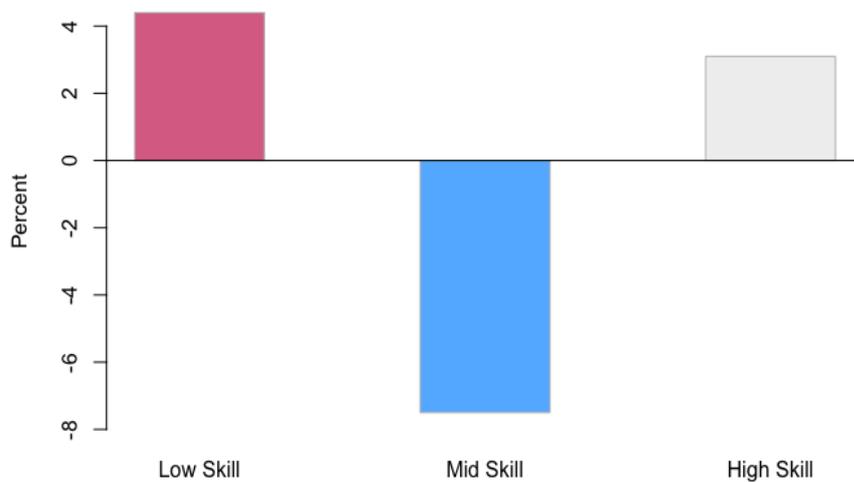
Figure 4.1: Private sector changes in occupational employment shares among working age adults (1997-2015)

The changes in employment shares in the private sector follow the standard U-shaped trend reported in developed countries (Goos et al. 2014). This is marked by a significant decline in the share of middle-skilled occupations, while the relative shares of low- and high-skilled occupations increased. This is exactly what we would expect to see if RBTC is causing the automation of routine tasks previously performed predominantly by workers in middle-skilled occupations. A feature that distinguishes the shape observed in South Africa from those observed in most developed economies is that the job losses in middle-skilled occupations are offset mainly by relative increases in the share of low- rather than middle-skilled occupations. That is, whereas the disappearance of middle-skilled occupations is primarily associated with a migration up the skills ladder for most workers in developed country labour markets, in South Africa most workers migrate downward. High-skilled occupations increased their share of total employment from 20.5 percent in 1997 to 23.5 percent in 2015, while low-skilled occupations increased from 33.2 to 38.7 per cent in the same period.

As a way of highlighting how differently the employment shares of the private sector evolved over

the 1997 to 2015 period relative to the public sector, Figure 4.2 below juxtaposes the changes in employment shares in the private and public sectors.
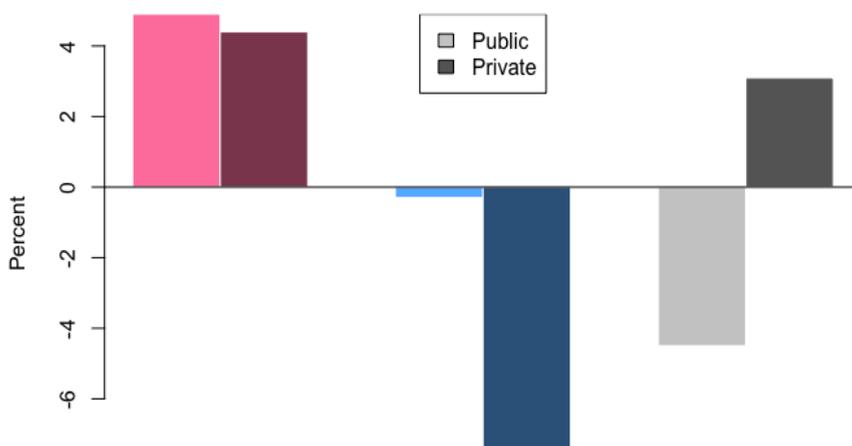
Figure 4.2: Changes in employment shares in the private and public sector (1997 - 2015)

Figure 4.2 shows that the nature of employment shifts among the different kinds of workers differs by sector. Specifically, the effects of RBTC appear to be limited to the private sector where employment decisions depend more on worker productivity than in the public sector. There is, however, no evidence that RBTC has influenced and affected employment outcomes in the public sector.

4.5.1.2   *Occupational trends across demographic subgroups*

The changes in employment shares demonstrate that South Africa experienced the same hollowing-out of the skills distribution reported elsewhere, and that the decline in the share of middle-skilled jobs coincided with a larger increase in the share of lower- than higher-skilled jobs. However, this overall trend does not tell us how the benefits and burdens of adjustment were distributed across racial and gender groups. For example, due to the legacy of racial discrimination under apartheid, one may be particularly interested to analyse whether RBTC is exacerbating existing

racial inequalities in the labour market. Figure 4.3 shows the change in the shares of workers of different race groups employed in the three occupational groups between 1997 and 2015.
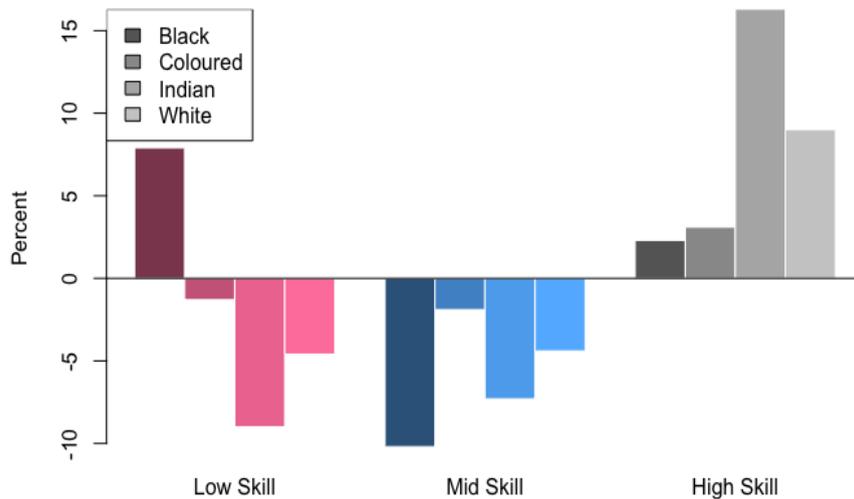
Figure 4.3: Private sector changes in occupational employment shares by race (1997-2015)

Disaggregating the employment share changes by race shows that black workers experienced the largest decrease in middle-skill employment which was predominantly offset by increases in low-skill employment, and minimal increases in high-skilled occupations. This differs strikingly from the three other race groups—white, Indian, and coloured—who consistently experienced a decline in both low- and middle-skill employment shares having them offset solely by gains in high skilled work. While it is beyond the scope of this chapter to determine specifically what channels led to the apparent disadvantage of the black worker, it is probable that the increases in low-skilled occupations for black workers may point to some enduring effects of apartheid. This may have disadvantaged the black worker in several ways including: leading black learners to dysfunctional schools which do not transfer adequate skills to students thereby having long-term productivity effects; entrenching discriminatory hiring and firing practices[30], and creating and reinforcing so-

---

[30]Although this period coincided with legislative improvements in South Africa like the Labour Relations Act (LRA) (1995), the Basic Conditions of Employment Act (BCEA) (1997) and the Employment Equity Act (EEA) (1998), there is little indication that these laws reduced the extent of discrimination in South Africa (Rulof Burger et al. 2016).

cial spheres where white workers have access to powerful networks that are not available to black workers. Consequently, as the nature of work became increasingly skills-biased leading to a sharp decline in the share of middle-skilled jobs, these low-skill black workers find themselves being relegated to low-skilled occupations as opposed to being able to adjust to the demands and requirements of high-skilled occupations. Note also however, that an equally compelling reason for these observed increases in the share of black workers at the bottom of the skills distribution may be due to increases in women's labour market participation in South Africa, which also coincided with the abolition of apartheid. There was therefore a situation where more "new entrants" were entering the labour market, however in low and unskilled occupations.

In addition to employment share changes disaggregated by race, Figure 4.4 reports employment share shifts for both genders.
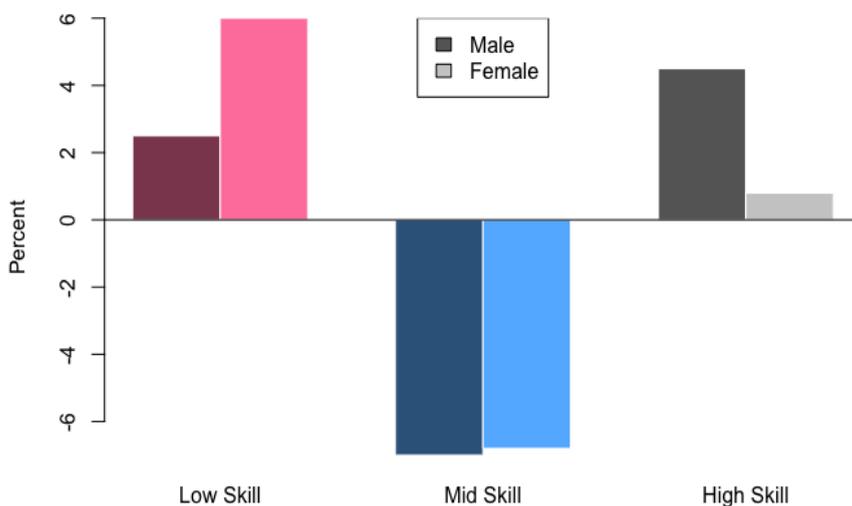
Figure 4.4: Private sector changes in occupation employment shares by gender (1997-2015)

We can observe the same U-shaped pattern we saw in Figures 4.1 and 4.3, signifying a hollowing-out of the skills distribution for both male and female workers. However, the share of female workers employed in low-skill occupations grew by more than the share of male workers. In

94

contrast, the share of male workers increased more among high-skilled occupations.

Furthermore, given the unsettling reality of South Africa's high unemployment, it is only befitting to conclude this section by ensuring that the trends so far discussed for the private sector hold even after taking into account the entire labour force which includes a broader group of working-age adults apart from those in the private sector. We therefore recalculate the shares of workers in low, middle and high occupations in private sector employment, but as a share of the working-age population. The remaining working-age population is classified into either public employment, or a residual "Outside formal employment category" which is made up of workers from the informal sector, as well as those that were unemployed, and not economically active.[31] This achieves two goals. First, if large changes in the share of workers in the formal private sector occurred during this period, then this may create the misleading impression of shrinking or growing occupational shares because we are scaling by a very different denominator. Secondly, we can see whether the decrease in the share of middle-skilled employment coincided with increases in public employment or those outside formal employment, which may suggest that these workers are displaced to other sectors rather than just other occupations with the same sector.

---

[31] The period under investigation was marked by a large increase in labour force participation, and changes in survey questionnaire design which results in improved capturing of informal sector workers, so accurately distinguishing the trends in informal sector employment, labour force participation and unemployment is not possible. However, since we are primarily concerned that RBTC may have resulted in a lower share of formal sector employment and that those who would in previous generations find employment among middle-skilled jobs wound up in either the informal sector, unemployment or economically inactive, our "Outside of formal employment" category allows us to test that hypothesis.
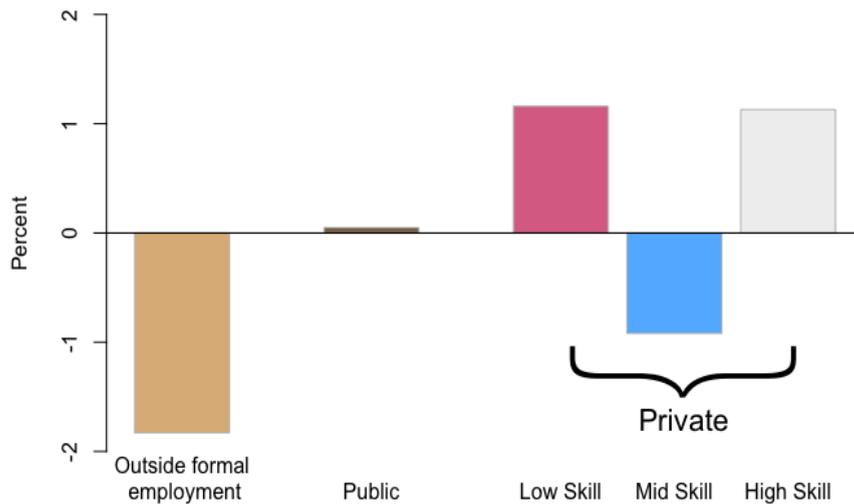
Figure 4.5: Private sector changes in occupational employment shares among working age adults (1997-2015)

Indeed, Figure 4.5 affirms that the trends discussed up to this point persist even when expressing these as a share of the entire working age population. Most workers who lose their middle-wage jobs appear to be either absorbed in low- or high-wage occupations within the formal private sector, and there is no indication of large displacement into informal employment, unemployment or economic inactivity.

The discussion has so far shown two features of the South African private sector in the context of RBTC. Firstly, just as the model predicted, the share of middle-skilled jobs declined relative to the share of low- and high-skilled jobs, which are harder to automate and therefore replace with technological advancements. Secondly, and more uniquely to South Africa, these dynamic effects are differentiated by race and gender where the advantaged groups appear to be moving to higher-skilled occupations, while the disadvantaged groups are moving to the less-skilled occupations. Finally, even though South Africa has a high share of workers in either informal employment or unemployment, there is little evidence that RBTC contributed to this problem due to displaced middle-wage workers ending up outside of formal employment.

4.5.1.3 Detailed occupational analysis

While classifying skills into these three broad types does a lot by way of observing the general employment shift trends, there still remains the question of which occupations actually drive these trends, and how those differ by demographic features. Tables 4.1 and 4.2 attempt to answer these questions. Table 4.1 begins by reporting the initial 1997 proportions of the 27 unique two-digit occupations by race and gender. This sets the context by showing the demographic composition of each occupation. Table 4.2 reports only those occupations that had employment share gains or drops of four or more percentage points where the average percentage point change for all occupations was 1.6.[32]

---

[32]The full version of these changes is available in Appendix C.1

Table 4.1: Employment shares by race and gender, 1997

Occupations ranked by mean earnings	ISCO-88 code	Male	Female	Black	White	Indian	Coloured
High paying occupations							
Legislators and senior officials	11	0.39	0.84	0.32	1.24	0.53	0.25
Corporate managers	12	2.43	1.77	0.57	7.34	3.73	0.98
General managers	13	6.8	5.14	3.65	13.6	12.26	4.46
Physical, mathematical, and engineering professionals	21	2.06	0.43	0.71	4.82	0.7	0.68
Life science and health professionals	22	0.14	0.29	0.12	0.46	0.25	0.02
Teaching professionals	23	0.36	0.49	0.23	1.06	0.1	0.2
Other professionals	24	2.23	2.67	1.01	6.31	4.45	1.3
Natural and engineering science associate professionals	31	2.49	2.53	1.34	5.19	3	3.03
Life science and health associate professionals	32	0.1	0.5	0.2	0.25	0.1	0.24
Teaching associate professionals	33	0.1	0.24	0.15	0.1	0.15	0.15
Other associate professionals	34	3.52	6.35	2.65	9.98	5.35	2.59
Middling occupations							
Office clerks	41	4.29	14.44	4.17	13.73	14.71	7.44
Customer service clerks	42	1.64	11.32	3.86	5.41	7.87	4.24
Extraction and building trades workers	71	10.69	1.35	10.13	4.11	3.23	6.72
Metal, machinery, and related trade workers	72	8.01	0.91	6.02	7.21	4.26	4.46
Precision, handicraft, craft printing, and related trade workers	73	0.9	0.81	0.85	0.64	1.62	1.06
Other craft and related trade workers	74	2.17	7.15	3.94	0.85	6.37	5.37
Stationary plant and related operators	81	1.67	0.38	1.72	0.72	0.7	0.67
Machine operators and assemblers	82	4	4.26	4.79	0.85	4.54	5.72
Drivers and mobile plant operators	83	13.71	1.12	13.93	3.17	4.31	6.75
Low-paying occupations							
Personal and protective service workers	51	5.78	5.92	7.35	3.79	1.26	4.26
Models, salespersons and demonstrators	52	3.74	8.23	4.56	5.07	11.27	4.88
Skilled agricultural and fishery workers	61	2.22	2.01	2.88	0.51	0.19	2.35
Subsistence agricultural and fishery workers	62	0.04		0.05			
Sales and service elementary occupations	91	3.09	5.17	5.03	1	1.43	3.05
Agricultural, fishery and related labourers	92	4.95	5.89	4.91	0.18	0.24	15.11
Labourers in mining, construction, manufacturing, and transport	93	12.47	9.8	14.86	2.41	7.38	14.01

Source: Own calculations

Table 4.1 shows that white and Indian workers had a relatively large proportion of their workers working in high-paying jobs such as managerial (12, 13) and professional (24, 34) occupations compared to black and coloured workers in 1997. This is likely an artefact of the apartheid regime which disadvantaged black and coloured workers by excluding them from high quality schools, tertiary education, and powerful social networks that are often key in obtaining employment in

highly skilled occupations. Among the middle-paid workers, occupations seem to be split more along gender lines than was the case with high-paying occupations. Jobs in office and customer service clerk occupations (41,42) appear to be predominantly performed by white and Indian females.

As expected, while all occupations experienced some level of growth or decline in employment share during the 1997 to 2015 period, the observed trends were not driven by all the occupations. Indeed, a handful of occupations can be singled out as being particularly important in driving the employment share changes.[33] Table 4.2 presents these occupations.

---

[33]To be easily distinguishable, the demographic group within these selected occupations that experienced a percentage point gain or drop of more than four percentage points is reported in bold.

Table 4.2: Levels & Changes in selected employment shares by race and gender, 1997-2015

Skill level	Occupations ranked by mean earnings	ISCO-88 code	Sub-group	Percent in 1997	Perecent in 2015	Percent point change
H	Corporate managers	12	Black	0.57	3.63	3.06
			White	**7.34**	**24.28**	**16.94**
			Indian	**3.73**	**17.69**	**13.96**
			Coloured	**0.98**	**6.27**	**5.29**
			Male	**2.43**	**9.27**	**6.84**
			Female	**1.77**	**7.06**	**5.29**
M	Other craft and related trade workers	74	Black	3.94	2.14	-1.8
			White	0.85	0.36	-0.49
			Indian	**6.37**	**0.53**	**-5.84**
			Coloured	5.37	1.57	-3.8
			Male	2.17	1.41	-0.76
			Female	**7.15**	**2.07**	**-5.08**
	Drivers and mobile plant operators	83	Black	**13.93**	**7.49**	**-6.44**
			White	3.17	1.06	-2.11
			Indian	4.31	3.37	-0.94
			Coloured	6.75	5.68	-1.07
			Male	**13.71**	**9.3**	**-4.41**
			Female	1.12	0.55	-0.57
L	Sales and service elementary occupations	91	Black	**5.03**	**10.28**	**5.25**
			White	1	0.76	-0.24
			Indian	1.43	1.39	-0.04
			Coloured	**3.05**	**7.67**	**4.62**
			Male	3.09	4.67	1.58
			Female	**5.17**	**12.72**	**7.55**
	Labourers in mining, construction etc	93	Black	14.86	12.29	-2.57
			White	2.41	1.47	-0.94
			Indian	**7.38**	**1.87**	**-5.51**
			Coloured	**14.01**	**9.59**	**-4.42**
			Male	12.47	10.51	-1.96
			Female	9.8	7.91	-1.89
	Personal and protective service workers	51	Black	**7.35**	**13.22**	**5.87**
			White	3.79	3.03	-0.76
			Indian	1.26	2.34	1.08
			Coloured	4.26	6.96	2.7
			Male	5.78	9.14	3.36
			Female	**5.92**	**11.55**	**5.63**

Source: Own calculations

Table 4.2 shows that the increase in employment shares in high-skilled workers seen in Figures 4.1 to 4.4 is predominantly due to large shifts in white and Indian workers specifically into the cor-

porate manager occupation. This increase was also more pronounced for males than for females. Among low-paying occupations, even though there was an overall increase in their employment share, one also sees that among the Indian and coloured sub-population, employment shares fell drastically in cases where they worked as labourers in mining, construction and manufacturing. On the other hand, the increases within the low-paying occupations appear to be predominantly driven by increases in employment shares among black and coloured female workers in sales and service elementary occupation and black female workers in personal and protective services. Unsurprisingly, given what we know to expect in the presence of RBTC, there are two main drivers for the decline in the middling-occupations: drivers and mobile plant operators and craft and related workers. These include jobs like wood and leather cutting, clothes-sorting, knitting, cement and concrete mixing, ice cream and cheesemaking, and various types of mining and construction work like sorting and excavating. Such occupations are jobs with high routine intensity and are therefore prone to replacement by advancements in ICT. These are occupations predominantly performed by black and Indian male and female workers.

In summary, it appears that among the highly skilled workers, there has been a relative increase in labour demand for corporate managers which includes jobs like production, operation, and department managers; school principals and deans; chief executive officers; and bank managers. At the lower end of the skills distribution, relative labour demand has increased for jobs like security personnel, barbers, waiters, maids, cleaners, and nannies — all jobs that cannot be automated and are low in routine task intensity. Furthermore, these jobs are predominantly performed by black and coloured workers.

4.5.1.4   *Decomposition of changes in occupational shares*

There are two reasons why middle-skilled occupations are disappearing. First, it could be driven by a general trend, observed across all industries, in which production is increasingly performed with more high- and low-skilled workers, and with fewer middle-skilled workers. Secondly, it could be that industries that use middle-skilled workers more intensively have experienced a decrease in their share of production, while all industries continue to produce with the same mix of low-, middle- and high-skilled workers. We can decompose the change in occupation shares into a *within* component (the part that is driven by changes in occupational shares within

industries, while keeping industry shares constant) and a *between* industry component (due to changing industry shares, while keeping occupational shares constant). Following Autor et al. (1998), this is achieved using the standard equation given by

$$\Delta P_{jt} = \sum_k \left(\Delta E_{kt} \gamma_{jk} \right) + \sum_k \left(\Delta \gamma_{jkt} E_k \right) = \Delta P_{jt}^b + \Delta P_{jt}^w \tag{4.11}$$

where ΔP_{jt} is the change in the aggregate share of total employment for occupation j between years t and τ, E_{jkt} is the employment of occupation j in year t and industry k as a share of total employment in year t, $E_{kt} = \sum_j E_{jkt}$ is total employment in industry k in year t, $\gamma_{jkt} = E_{jkt}/E_{kt}$ is the group j share of employment in industry k in year t, $\gamma_{jk} = (\gamma_{jkt} + \gamma_{jk\tau})/2$, and $E_k = (E_{kt} + E_{k\tau})/2$. The resulting first term (ΔP_{jt}^b) represents the changes in the aggregate proportion of workers due to changes in employment shares *between* industries, and the second term (ΔP_{jt}^w) reflects the changes in the aggregate proportion of workers due to changes in employment shares *within* industries. Table 4.3 below reports the results of Equation 4.11 above.

Table 4.3: Between and within decomposition

	Overall	Between	Within
Low	5.22	-0.49	5.71
Middle	-5.82	-2.56	-3.26
High	0.60	3.05	-2.45

Source: Own calculations

Table 4.3 shows that for low- and middle-occupations, within-industry forces explain relative labour demand shifts more than between-industry forces. In their work, Bhorat et al. (2014) find a similar result, showing further that the within-industry component constituted between 86 to 96 percent of aggregate demand shifts in the 2001 to 2012 period. Among other competing reasons, they suspect that this may be due to technological change that has a preference for a specific skill composition which in turn altered the labour demand practices of firms.

Regarding the middle-skill occupations, both the between and within components are negative. What this shows is that for these groups of workers, not only is there a relative demand shift away from their kind of jobs within their industry, but also that that phenomenon is happening in industries outside their own. Practically, this means a middle-skilled worker cannot easily

leave their industry and hope to find work that requires similar skills in another industry. Table 4.3 further shows that whereas both low- and middle-skilled employment shifts are driven by within-industry forces, that is not the case for high-skilled occupations. For these occupations, the between-industry component dominates the within-industry component. This means that high-skilled jobs are growing despite the fact that some industries are not increasing in their share of high-skilled workers. Instead, industries like finance and services that employ a large proportion of high-skilled workers are expanding.[34] The results from these decompositions match well with the descriptive employment shifts discussed earlier.

4.5.2 Trends in wages

The theory of RBTC makes strong predictions about trends in employment shares of the different occupation groups, which was shown to also characterise the South African labour market between 1997 and 2015. Since this theory suggests that the labour demand for different types of workers is affected, we may also see its effects in wage trends. More specifically, if the relative demand for middle-skilled workers is decreasing then we would expect to see a decrease in the relative wages paid to these workers in labour markets where wages are determined mainly by market forces. However, if non-market forces—like minimum wages or trade union bargaining—are crucial factors in determining wages, then the effect of such factors may drown out the effect of labour demand so that the effect of technological growth is only observable in employment shares. To this end, Figure 4.6 plots the average annualised growth rate (AAGR) of real earnings across percentiles in the private sector of the South African labour market between 1997 and 2015. In addition, Figure 4.6 can be used to assess the increasing and decreasing patterns of inequality at different points in the wage distribution, where a negative slope signifies decreasing inequality, and an upward slope signifies inequality increasing.

---

[34]See Table C.5 in Appendix C.2 that reports the detailed between and within decomposition for each industry.
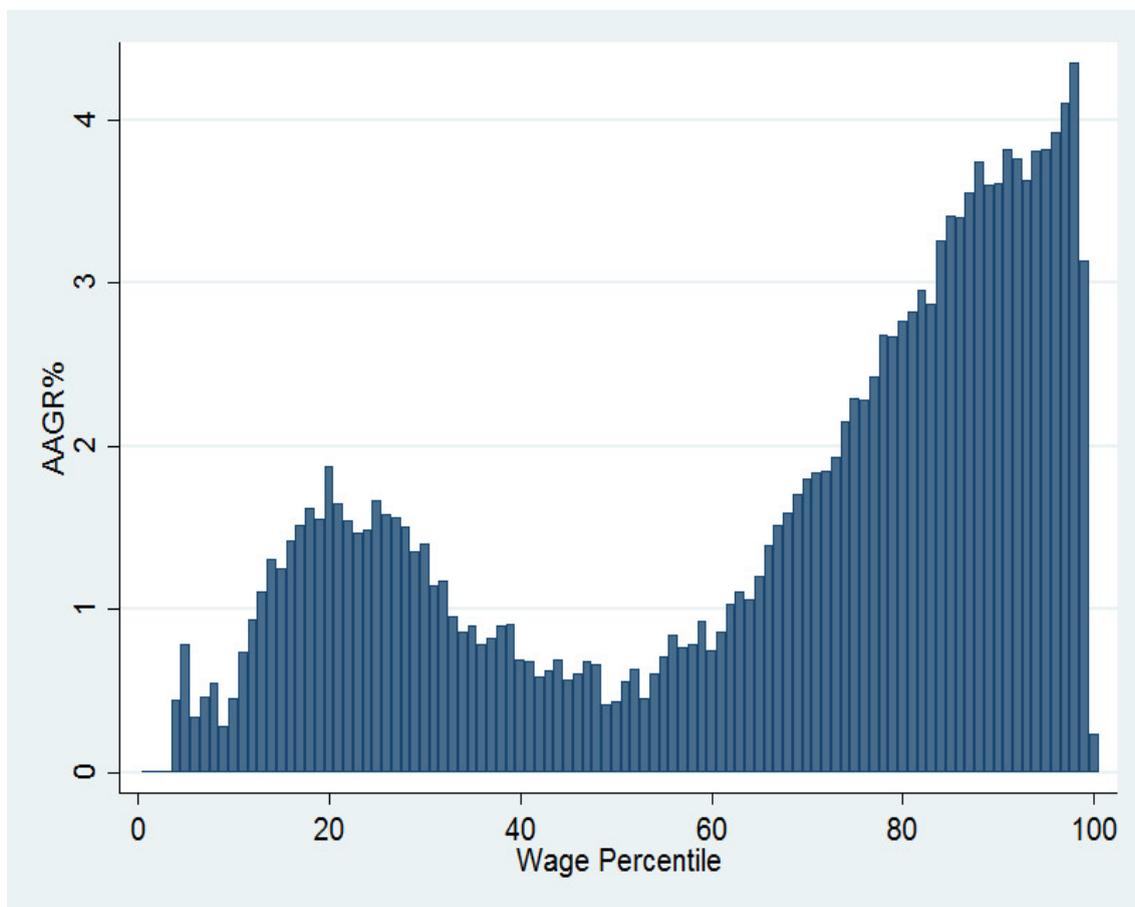
Figure 4.6: Annual average growth rate of mean real earnings (South Africa, 1997-2015)

As can be observed, the change in wages clearly differs by position in the wage distribution. Workers who were initially in the poorest decile experience almost no wage growth while for deciles 2 and 3, there was quite high wage growth. This is followed by 3 deciles of consecutively low wage growth: 3, 4 and 5. Finally, workers in deciles 7 to 10 all experience increasing wage growth. Specifically, the wage growth in decile 10 was higher than that of decile 9, the one for decile 9, higher than decile 8, and the one for decile 8 higher than that of decile 7. What all this implies is rapidly increasing inequality at the top end and at the very bottom end of the wages distribution, but decreasing inequality in the middle of the distribution.

Another way of looking at how wages for each skill category evolved over the 18-year sample period would be to report what the actual mean wage for each of them was in 1997 and 2015

respectively. From this, one can calculate the AAGR for each occupational category. Table 4.4 presents these results.

Table 4.4: Mean and annual average growth rate per skill type

	Mean		AAGR(%)
	1997	2015	
Low	4302.2	5536.6	1.4
	(220.9)	(608.6)	(0.007)
Middle	6310.9	8457.8	1.6
	(307.9)	(607.7)	(0.005)
High	16100.4	33690.4	4.1
	(1294.8)	(2513.0)	(0.006)

Source: Own calculations
Note: Standard errors in parenthesis

Unsurprisingly, the AAGR for high-skilled workers is the largest, showing in a different way that on average, the highly skilled workers experienced the largest growth rate in their wages. The fact that high-skilled wages grew much more rapidly than middle-skilled wages is exactly in line with the predictions of RBTC which suggests that middle-skilled workers experienced slower productivity growth than highly skilled workers during this period. In a labour market with very low unemployment and competitively determined wages, we would also have expected to see the wages of low-skilled workers expand more rapidly than those of their middle-skilled counterparts. However, Table 4.4 suggests that in the South African context, where there is an oversupply of low-skilled workers and the wages for low-skilled workers are partly determined by institutional factors, firms need not raise wages to increase employment of low-skilled workers.

4.6 Estimation Results

This section moves beyond using descriptive evidence to investigate whether the predictions of RBTC can help us understand South African labour market trends between 1997 and 2015. In section 4.6.1 we start by using a multivariate regression to explore the trends in wages by occupational group. Although we have already demonstrated that employment trends across broad occupational groups - and to a lesser extent, also wage trends across the same groups - support the hypothesis that RBTC is affecting South African labour market trends, the discussion in section

4.4.3 noted that this occupational grouping is not the ideal measure of the routine task intensity. We may be concerned that these trends are driven by some other factors that vary across occupations. In order to address this concern, we use regression analysis and the recently gathered STEPS data to confirm that the trends observed for the low-, middle- and high-skilled occupation groups are driven by the routine task intensity of individual occupations, rather than by some other differences.

In section 4.6.2 we go one step further and use non-linear, system estimation techniques to estimate a fully specified RBTC model. This model leverages all of the available wage and employment data to investigate whether RBTC can provide a coherent account of observed wage and employment trends as equilibrium outcomes in a labour market where technological growth decreased the relative productivity of workers in middle-skilled occupations.

4.6.1   Returns to routine - OLS regression

To investigate the relationship between either being a low-, middle-, or high-skilled worker on wages, this section estimates the expected change in wages owing to what skill level a worker is employed in. One way of doing this is by running a reduced form OLS regression that allows us to model the relationship between a dependent variable (log of wages) and in this case three predictor variables: a dummy for the type of labour, a linear time trend variable to account for the expected year-on-year changes, and an interaction term between the type of labour dummy and linear time trend variables.[35] We are primarily interested in the interaction between labour type (or RTI) and time. RBTC predicts that middle skilled workers, or those in occupations that require a high share of routine tasks, should experience slower than average wage growth. The results from this regression analysis are presented in Table 4.5.

---

[35]This follows work by Firpo et al. (2011), Goos et al. (2014).

Table 4.5: Wage regression with RTI time trend interactions

| | *Dependent Variable:* Log of Wages | | | |
	(1)	(2)	(3)	(4)
Trend	0.00615***	0.0135***	0.0118***	0.00939***
	(0.000381)	(0.000436)	(0.000474)	(0.000261)
RTI		-1.343***	-1.412***	-0.108***
		(0.0101)	(0.00956)	(0.00265)
RTI*Year		-0.0206***	-0.0125***	-0.00264***
		(0.000940)	(0.000889)	(0.000245)
High Skill	1.328***			
	(0.00647)			
Middle Skill	0.511***			
	(0.00570)			
High Skill*Year	0.00116*			
	((0.000584)			
Middle Skill*Year	-0.00159**			
	(0.000533)			
N	494660	467039	467009	493124
R^2	0.248	0.150	0.163	0.020

Source: Own calculations

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results in column 1 take low-skill workers as the base group, meaning the results for middle-skill and high-skill are interpreted relative to low-skill workers. The results suggest that, compared to a low-skill worker, being a middle- and high-skill worker is estimated to raise your actual wage by a factor of 66.7 and 277.3 percent respectively. More importantly, and consistent with the descriptive results discussed in section 4.5, the interactions between year and occupation groups indicate that both high- and low-skill occupations experienced wage growth relative to middle-skilled workers during this period. This is consistent with the predictions of the RBTC model.

Table 4.5 also presents results for three more models that replace occupation groups (which have

been used as an imprecise proxy for routine task intensity in the early empirical literature) with an RTI variable. The motivation for this change is that the measure of RTI will more accurately capture the task intensity of specific occupations, especially where this variable was constructed using information on occupations from labour markets that more closely resemble that of South Africa. This allows us to explicitly test whether the wage and employment trends observed across occupation groups in section 4.5 are driven by the routine task intensity of these occupations (as predicted by the RBTC model) or whether they are driven by other factors. As discussed in section 4.4, there are three measures of RTI that we could conceivably use in our analysis. The one used in column 2 of Table 4.5 calculates the RTI variable as a composite index for all 12 countries included in the STEPS data. The variable used in column 3 does so only for workers in Colombia and Macedonia which have GDP per capita that are similar to South Africa and could therefore potentially have RTI values that better measure task content of occupations in South Africa. Finally, the measure in column 4 is from Goos et al. (2014) which is derived from US-based DOT task measures as is the norm to do in this literature.

The literature on measurement error suggests that variables that contain more noise and less informative variation usually cause attenuation bias in regression coefficients, provided that the measurement error is roughly classical in nature. This provides us with a way of assessing the appropriateness of the different RTI measures to the South African case: measures containing more measurement error should be more strongly biased towards zero. Comparing the models in columns 2, 3 and 4 all show that occupations that require the execution of more routine tasks have experienced slower wage growth over this period than those that perform more non-routine tasks. This is exactly what is predicted by the RBTC model, and serves to confirm that the wage trends observed across occupation groups were (at least partly) driven by the routine task intensity of these occupations. Furthermore, we can see that the RTI measure derived from US data (in column 4) produces much smaller coefficients and a lower R-squared than those obtained with the STEPS measures. This is consistent with the hypothesis that extrapolating the routine task intensity of occupations in developed countries to developing country data induces severe measurement error problems. In a smaller dataset, these small effects may well have been insignificant, which could have resulted in an erroneous conclusion that RBTC has not affected the South African labour market. Regarding which of the two STEPS measures - used in columns 2 and 3 -

is more appropriate, the evidence is mixed. The measure used in column 3 produces a higher R-squared and RTI coefficient, but a lower RTI-time interaction coefficient than the measure used in column 2. We are inclined to conclude that the measure that only uses information from comparable middle-income countries produces a less noisy RTI index. However, the fact that this measure only uses two countries to compile the measure may in itself contribute to measurement error, which could explain the attenuated RTI-time interaction coefficient. Regardless of which of these measures is more appropriate, the results clearly show that wage trends operate via a reduced demand for occupations that have a high RTI, and that this result is easier to identify when using an RTI measure derived from data from other developing countries. In reality, the true estimate would likely be higher than that produced by the model in column 2 as even that RTI is based on a basket of countries and not South Africa specifically. Overall, these findings are consistent with those reported in Bhorat et al. (2014) for the 2001 to 2012 period, that found a negative impact on wages as occupations became more routine in their processes, giving credence to the RBTC theory.

Finally, to further scrutinise the descriptive analysis on employment share changes discussed in section 4.5, this section also includes regression output on the same outcome. The model specifications are exactly the same as those for Table 4.5 with two exceptions. Firstly, the dependent variable is no longer the log of wages. Instead, we use a variable which is calculated as the share of each two-digit ISCO-88 occupation in that year. In so doing we are particularly interested in the time interaction term which shows us how the employment shares were evolving over time. Secondly, the base category in column 1 is no longer the low-skill category but the middle-skill category.

Table 4.6: Employment share regression with RTI time trend interactions

| | *Dependent Variable:* Share of Workers | | | |
	(1)	(2)	(3)	(4)
Trend	-0.000451	0.000433	0.000451	0.0000297
	(0.000262)	(0.000244)	(0.000253)	(0.000173)
RTI		0.0465***	0.0491***	0.00336
		(0.00555)	(0.00510)	(0.00181)
RTI*Year		-0.00133*	-0.00124*	0.000167
		(0.000534)	(0.000491)	(0.000174)
High Skill	-0.0329***			
	(0.00369)			
Low Skill	0.0135***			
	(0.00406)			
High Skill*Year	0.00110**			
	(0.000354)			
Low Skill*Year	0.000164			
	(0.000391)			
N	1173	1088	1076	1130
R^2	0.229	0.112	0.152	0.021

Source: Own calculations

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Even though the coefficient for low-skill occupations was not significant, both the interactions terms in column 1 suggest an upward trend of high- and low-skilled share of workers relative to those in the middle of the skills distribution. This is again consistent with the predictions of RBTC. Further, columns 2 and 3 show that occupations that require performing more routine tasks have experienced a decline in the share of workers over time. This is in contrast to the findings reported in column 4 that are based on the US occupation schemes which seem to suggest that occupations that involve a lot of routine tasks were trending upwards during this time. Not only is this finding inconsistent with the descriptive analysis presented earlier in section 4.5, but the RTI-time interaction coefficient is also insignificant at all significance levels. This further goes to

show that occupation schemes based on developed country labour markets may not be best suited for use in developing country contexts.

4.6.2 Estimates of the elasticity of substitution and labour specific growth rates

Section 4.5 demonstrated that South Africa experienced wage and employment trends across broad occupational groups that followed the same pattern observed in developed country labour markets and ascribed to RBTC. Section 4.6.1 then confirmed that these trends were indeed driven by the routine task intensity of these occupations, which provided further support for the hypothesis that RBTC has affected labour demand in South Africa. In this section, we now investigate whether the RBTC model provides a coherent account of wage and employment trends across industries. We do this by using a non-linear system estimator to model wage and employment levels as equilibrium outcomes of a labour market in which production occurs according to a CES production function, and the productivity of workers is allowed to vary over time. The RBTC model hypothesises that workers in middle-skilled occupations are experiencing reduced productivity due to the automation of routine tasks, and that firms are responding to this by replacing these workers with high- and low-skilled workers. This substitution affects employment shares and wage levels. If RBTC is truly the driving force behind the observed South African labour market trends, then we would expect to see that our model fits the data well, and that the estimated productivity growth should be low for middle-skilled workers, and high for low- and high-skilled workers. This analysis estimates the system of equations 4.7 to 4.10 from section 4.3. In doing so, the system of equations backs out estimates for the growth rates of low-, middle- and high-skilled augmenting technical progress.

The current literature provides little by way of a consensus concerning what the value of the point elasticity of substitution for the different labour types should be. Bárány & Siegel (2019), who model a production function with three types of labour, settle on an elasticity of substitution estimate of 0.6 after trying a series of calibrated values. Further, depending on the model specification and sample of countries they were looking at, Goos et al. (2014) recover elasticities σ, of 0.53, 0.66, and 0.9. Lee & Shin (2017), who use a different set of occupations settle on a value of 0.7, while Duernecker & Herrendorf (2017) estimate a point elasticity of 0.56. Interestingly, all the proposed and estimated values of the elasticity of substitution parameter are less than one, i.e.

111

$\sigma < 1$. Crucially, this assumes that the different labour types are complements in the production process, which may be true for developed economies where most of this literature is based but may not be for South Africa. To the best of our knowledge, no research has been carried out for the South African labour market specifically aiming to recover the elasticity of substitution for the different labour types.[36]

Since there is no consensus in the literature, and no clear guidance from South African research itself, we follow Bárány & Siegel (2019) to run the system of equations 4.7 to 4.10 for a wide range of possible values, $\sigma \in [0.1, 1.9]$ (Bárány & Siegel 2019). From this analysis, we settle on a point elasticity of substitution value of 1.4 which gives the best fit for the data. Table 4.7 reports the estimates for the growth rates of low-, middle- and high-skilled workers during the period from 1997 to 2015 for an elasticity value of 1.4.

---

[36]Notwithstanding this, Kreuser et al. (2015), estimates a factor augmenting technical change elasticity of substitution value of 1.32. This finding was however for the standard production function with two standard input variables: labour and capital.

Table 4.7: Estimation output for system equations

System Equations Estimates	
γ_l	0.0182*
	(0.00735)
γ_m	0.0008
	(0.00550)
γ_h	0.0475***
	(0.00592)
R^2 for equations 7-10	
R^2_{Eq7}	0.5974
R^2_{Eq8}	0.9927
R^2_{Eq9}	0.9935
R^2_{Eq10}	0.9944
Wald Hypothesis Test	
$\gamma_l = \gamma_m$	0.0720
$\gamma_h = \gamma_m$	0.0066
$\gamma_h = \gamma_l$	< 0.000
N	153

Source: Own calculations

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4.7 shows that at the point elasticity of substitution of 1.4, the respective productivity growth rates of low-skilled, middle-skilled and high-skilled workers are consistent with the existing theory for RBTC and indicate general factor-augmenting technical progress (i.e. $\gamma_l > 0 \neq \gamma_m > 0 \neq \gamma_h > 0$, while $\gamma_m < \gamma_l$ and $\gamma_m < \gamma_h$). Specifically, for a unit increase in the log of output, the model suggests that high-skilled workers experience the highest productivity growth during the 18 year period of 4.8 percent, followed by low-skilled workers with 1.8 percent. Both of these estimates are significant at the 1 and 10 percent level of significance respectively. Meanwhile, middle-skilled workers appear to have experienced the slowest productivity growth of 0.08 percent during the same period. However, this estimate appears not to be significantly different from zero productivity growth. That is, the model finds that technical change in South Africa is biased

113

in a way that increases the relative productivity (and consequently shares) of low- and high-skilled workers over time, which is exactly in line with the prediction of the RBTC model.

Other interesting findings included in Table 4.7 are results from a Wald hypothesis test. From this test, we can reject the hypothesis that high-skilled workers experienced similar productivity growth to middle-skilled workers at a 1% level of significance, and that low-skilled workers experienced similar productivity growth to middle-skilled workers at a 10% level of growth. Therefore, the productivity growth estimates for both low-skilled and high-skilled workers are statistically different from that of middle-skilled workers. In being consistent with the literature, this chapter reports additional estimates from a select number of elasticity of substitution values other than 1.4. These are presented in Table 4.8.

Table 4.8: Estimation output for system equations varying elasticity of substitution

| | $\sigma = 1.3$ | $\sigma = 1.4$ | $\sigma = 1.5$ | $\sigma = 1.6$ | $\sigma = 1.7$ | $\sigma = 1.8$ | $\sigma = 1.9$ |
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
γ_l	0.0178*	0.0182*	0.0181**	0.0178**	0.0175**	0.0171**	0.0168**
	(0.00795)	(0.00735)	(0.00682)	(0.00637)	(0.00601)	(0.00570)	(0.00545)
γ_m	0.000381	0.000833	0.00170	0.00268	0.00365	0.00455	0.00537
	(0.00593)	(0.00550)	(0.00516)	(0.00487)	(0.00465)	(0.00446)	(0.00430)
γ_h	0.0472***	0.0475***	0.0474***	0.0471***	0.0468***	0.0464***	0.0461***
	(0.00614)	(0.00592)	(0.00569)	(0.00548)	(0.00530)	(0.00515)	(0.00501)
N	153	153	153	153	153	153	153

Source: Own calculations

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

It appears the productivity growth among high-skilled workers remains significant at the 1% level and continues to be the largest among the types of workers irrespective of the value of the elasticity of substitution chosen. Notably, γ_l appears to be even more significant when choosing higher elasticity of substitution values. Overall, the results appear to be robust within this range of values of the elasticity of substitution.

4.7   Conclusion

This chapter used various sources of representative data to explore whether the evidence of RBTC reported in several developed countries is also present in a large, middle-income developing country; South Africa. In line with the predictions of RBTC, the findings point to job polarisation in the South African labour market characterised by increases in the share of low- and high-skilled workers and a decline in the share of middle-skilled workers. Further, these movements appear to be differentiated by race and gender with the advantaged groups (male, white, and Indian) predominantly moving into high-skilled occupations and the disadvantaged groups mainly moving into less-skilled occupations. Furthermore, regression analysis shows that jobs that involve a high share of routine tasks—whose workers happen to be those in the middle of the skills distribution—experienced a decline in employment and wages over time. Fortunately, there is no indication that RBTC is currently contributing to the problem of unemployment.

What this means is that if the policy goal is merely keeping unemployment in check, then the immediate, short-term implication of these trends may be minimal since all you have is the migration of workers from an occupation with one skill requirement to the next skill category while unemployment is left (largely) unaffected. However, if reducing inequality is the policy goal then, if the current trends persist, it is more than likely that this would further entrench and exacerbate wage inequality, especially by race and gender in the medium to long term. To navigate away from this impending predicament, it may be necessary to up-skill middle-skilled workers from disadvantaged groups via, for example, improved schooling quality or training programmes to facilitate movement up the skills distribution. In doing so, the public sector could potentially play an important role to ameliorate employment shocks that are a result of RBTC. Expanding public sector employment could therefore protect workers from unemployment. The government could also specifically look at incentivising private sector investment into sectors that are known to hire a large number of middle-skilled workers. This would have the benefit of protecting an otherwise eroding middle-class which serves as a good middle-ground between low- and high-skilled occupations. If middle-skilled jobs continue to disappear, it more than likely locks low-skilled individuals to the bottom of the skills (and wages) distribution simply because the leap from low to high-skill occupations is too daunting to overcome.

Despite its contributions, this chapter was not able to address other equally important questions which are left for future research. This includes not being able to determine the mechanisms behind why there was no upward movement of the displaced middle-skilled workers. Potential reasons could be that low- and middle-skilled workers are more substitutable than middle- and high-skilled workers. However, answering these questions requires more flexible functional forms other than that used in this chapter. Further, although this chapter used repeated cross-sectional data and found no evidence of net increases in unemployment, it is not clear that workers employed in new low-skilled jobs are the same ones who lost their middle-skilled jobs. Follow-up studies can therefore use panel data to investigate such employment dynamics. Additionally, this chapter did not investigate the role of physical capital by excluding it in the production function. Future work could incorporate it as well as allow productivity trends to vary by subgroup in the structural model. However, since RBTC has been found in international studies to mainly operate through technology rather than the capital stock, we do not expect this to alter the results by much.

CHAPTER 5

Conclusion

Inequality in much of the developing world continues not only to persist but to increase, despite large-scale economic growth both globally and sometimes even in developing countries in particular. These inequalities are present from the early years of an individual's life and continue on to adulthood, often culminating in, and manifesting through differentiated labour market outcomes: where individuals from high socio-economic backgrounds not only outperform their poorer counterparts, but continue on to acquire the necessary skills, social acumen and access to networks that firmly bestow on them an unfair advantage in the labour market. This thesis explored these broad issues in three main ways: (1) by analysing performance differentials in two developing countries, Malawi and Namibia; (2) exploring whether it is possible to identify different socio-economic groups in low-income country contexts; and (3) assessing whether technological growth is destroying jobs and exacerbating wage inequality in South Africa.

Doing so, contributed to the current literature in two main ways. First, by providing evidence of a flat socio-economic gradient in Malawi and rural Namibia which strays from the typical finding reported for most of the world where student performance is responsive to increases in family SES. This included discussing an alternative approach at grouping households in low-income areas especially, into their respective wealth groups. Secondly, for the first time in South Africa, this thesis provides evidence of routine-biased technological change in its labour market over a period of 18 years, 1997 to 2015. A brief summary of the chapters is offered below.

5.1 Chapters 2 and 3: Socio-economic status and student reading performance

In light of the overwhelming and (seemingly) conclusive finding of the positive impact of socio-economic status on student performance in developed countries, Chapter 2 sought to ascertain if this finding carries over to low- and middle-income countries in the selfsame way. This was done using Malawi as an example of a low-income country and Namibia as an example of a middle-

117

income country. The motivation here was to investigate the reading performance differentials between students with high socio-economic status from those with low socio-economic status, based on the suspicion that the factors that affect student performance in a typical developed or developing country may not be the same in a country that is very poor (Malawi) or one with that is highly unequal (Namibia).

The findings affirm this hypothesis. In the case of Malawi, Chapter 2 points to there being other contributing factors that influence student performance to a much greater extent than socio-economic background. This is both problematic and insightful: Problematic because it goes against the bulk of the evidence that has shown this to be true, and insightful because knowing that socio-economic status does not in itself hold the key to better performance allows education stakeholders to look elsewhere for the solution. This means that further research is required to ascertain whether less emphasis should be placed on socio-economic status in low-income countries overall, or perhaps instead, on finding other factors that may be holding back learning amongst large segments of the school population. It is likely that the benefits of socio-economic status in such areas are best realised when socio-economic status is paired with other factors such as improved teacher quality, improved teaching and learning facilities, and overall poverty reduction—all of which are problems common to low-income countries.

Namibia paints a strikingly different picture in that student performance responds positively to socio-economic status. Further analysis, however, shows that these performance differentials are primarily a feature of urban centres while rural Namibia, like Malawi, has an approximately flat socio-economic gradient. Again, this seems to suggest that the mechanisms that influence student performance—at least in richer areas—do not carry over nor retain their potency when applied to poorer areas. To put it loosely, one size does not fit all as you have on the one hand a "fairer" education system in Malawi, but one that performs below acceptable levels for the entire range of the SES distribution, and another less-equitable education system in Namibia whose high SES students far outperform their poorer counterparts. As alluded to in Chapter 2, this brings to the fore (Willms et al. 2006, p.11) point that "the central question facing most schools and countries is: 'How can we raise and level the learning bar?' ". Indeed, while Malawi is in need of raising of their learning bar, Namibia requires both the raising and levelling of theirs.

Taken together, this amplifies the need to identify who the poor are, and what features likely characterise them so as to aid researchers and policymakers further in coming up with interventions that are adequately adapted to improving student outcomes in these areas that typically pull the learning bars down. This serves as the main motivation for Chapter 3, which uses finite mixture modelling applied to 2007 SACMEQ III data to decompose the SES index into its latent distributions that together make up the single SES distribution. This reveals latent heterogeneity within the samples in the data, crucially making it possible to distinguish between the distinct shades of relative poverty that underly the data. Taken together, Chapters 2 and 3 highlight the need for analysis that is not blind to the idiosyncrasies that distinguish one area from another.

5.2   Chapter 4: Does technological growth destroy jobs and exacerbate wage inequality in middle-income countries? Routine Biased Technological Change and the South African Labour Market

Chapter 4 examined employment and wage trends among working age adults in South Africa for a period spanning 18 years (1997–2015). The analysis aimed to answer three main questions. First, does the South African labour market show evidence of routine-biased technological change? Second, if routine-biased technological change does exist, how has it affected employment and wage trends across demographic groups, economic sectors and occupations? Third, are the trends in employment and wages consistent with labour demand that would be produced by technological progress? And, fourth, are the results robust to changes in how routine is measured?

To answer these questions the analysis used several sources of data. Employment and wage information was retrieved from DataFirst's PALMS data, output information from the SARB output data series, and task content information from the World Bank's STEPS data and the US-based DOT task measures. Merging this task content information to the PALMS data made it possible—for the first time in a South African study in as far as can be ascertained, and one of the very first times in a developing country context—to explicitly test whether the wage and employment trends observed across occupation groups were driven by routine task intensity using a measure of routine that is specifically constructed for use in a developing country context.

The findings show evidence of employment polarisation characterised by increases in the share of high- and low-skilled workers, with a corresponding decline in the share of middle-skilled workers. Unfortunately, the increases in the share of low-skilled workers are almost solely driven by black and coloured workers, whereas Indian and white workers move up the skills distribution and find work in jobs requiring more skills. Using OLS regressions and non-linear systems estimation techniques the analysis is further able to show that the declines in employment and wages are occurring primarily among those occupations whose work involves a high share of routine tasks, as their type of work is easily replaceable by advancements in ICT. These happen to be workers in the middle of the skills distribution. The inclusion of three different measures of routine makes it possible to verify that measures specifically constructed with developing country contexts in mind outperform their developed country counterparts. This should not be surprising, seeing that developing countries are uniquely different from developed countries.

For a country with high inequality like South Africa, the findings in Chapter 3 are quite disheartening. While it is commendable that parts of the population are finding better jobs, it is particularly unfortunate that these advantages are differentiated by race, and that previously disadvantaged groups, like black and female workers, predominantly end up in jobs that require less skill and are therefore also paid less. This differs from similar trends in developed countries where displaced middle-skilled workers are typically absorbed into high-skilled occupations. If these trends were to continue unabated, South Africa would be heading towards a situation where a large and poorer majority were "serving" an elite minority. You would then have an economy with more or less two classes, the rich and the poor, with the rich primarily engaged in jobs that over time perpetuate their status quo and continue to exploit current and arising economic opportunities. The fact that the education system in South Africa is also known to be differentiated by race, means that it is up to the government to find ways of up-skilling current and previously middle-skilled workers while also making schools more equitable in terms of educational quality. While such policy action may show little to no fruit in the short run, they most likely would in the medium to long term.

5.3 Further research and final remarks

This thesis has sought to reduce the information gap in developing countries by analysing cross-country reading differentials and labour market employment and wage trends in a select number of developing countries: Malawi, Namibia, and South Africa. Though the findings discussed here are informative and insightful, one issue that clearly stands out is the need for research that is both more quantitively comprehensive as regards availability of data for historically disadvantaged and poor groups particularly, but also nuanced in its appreciation of contextual differences that influence the mechanisms that are known to affect outcomes.

Sadly, this is not always possible for a number of reasons. Among them is the fact that while developed countries have long participated in large-scale international assessment programmes that have given them much needed access to comparative data about their education systems and student outcomes, this has not been the norm among developing countries. Despite agencies like SACMEQ making significant progress to fill this gap, there is a need for more to be done if individual governments are to have access to that kind of comparative information. Related to this is the lack of the necessary technical skills that are needed to exploit this data in meaningful ways once it has been made available. In this regard, there is a need for widespread training in order to develop teams with required competences whose availability is critical to the success of these international assessments. In additional to their data management and analysis roles, having a competent workforce also means you have individuals familiar with the local contexts to guide questionnaire designs in ways that lead to the collection of the most essential information. It may well be, for example, that the flat socio-economic gradient in Malawi, being a very poor country, may in part be due to measurement error owing to the use of an asset ownership list whose items are not best suited to distinguish between households of different socio-economic rank. In the worst case, both the data and the technical skills may be absent, in which case these information gaps will continue to impede the development and implementation of interventions that can potentially make a difference and improve student outcomes in parts of the world that really need improvement.

Despite its contribution, the thesis also highlights the need for further research concerned with

the impact on developing country labour markets of RBTC specifically, and ICT adoption more generally. To this end, an upper-middle income country like South Africa that has aspects of both a developed and a developing country and is further characterised by a labour market with high unemployment and a sizeable informal sector models well what could be in store for other developing countries. This is especially true for developing countries in Africa, who are usually net-importers of advancements in ICT (which are known to alter production processes in favour of skilled workers) while also boasting sub-par education systems that do not produce workers that are flexible and skilled enough to gainfully progress in the modern economy. On the far end of the scale, low-income countries like Malawi may be too homogenous, lacking the required data, and too underdeveloped to show evidence similar to the trends reported in more developed economies. Therefore, future research in the South African case that extends the structural model discussed in this thesis to include such important factors as capital and unemployment would go a long way in furthering our understanding of the underlying dynamics of continued and inevitable adoption of ICT into the production process. This may hold further insights for other current upper-middle-income countries, and for countries that currently have a less developed and differentiated labour market.

REFERENCES

Acemoglu, D. (2002), 'Technical Change, Inequality, and the Labor Market', *Journal of Economic Literature* **40**(1), 7–72.

Acemoglu, D. & Autor, D. (2010), 'Skills, tasks and technologies: Implications for employment and earnings'.

Acemoglu, D. & Autor, D. (2011), Chapter 12 - Skills, Tasks and Technologies: Implications for Employment and Earnings, *in* D. Card & O. Ashenfelter, eds, 'Handbook of Labor Economics', Vol. 4, Elsevier, pp. 1043–1171.

Adler, N. E., Boyce, T., Chesney, M. A., Cohen, S., Folkman, S., Kahn, R. L. & Syme, S. L. (1994), 'Socioeconomic status and health: The challenge of the gradient', *American Psychologist* **49**(1), 15–24.

Aitkin, M., Anderson, D. & Hinde, J. (1981), 'Statistical Modelling of Data on Teaching Styles', *Journal of the Royal Statistical Society: Series A (General)* **144**(4), 419–448.

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**(6), 716–723.

Almeida, R. K., Corseuil, C. H. L. & Poole, J. P. (2017), *The Impact of Digital Technologies on Routine Tasks: Do Labor Policies Matter?*, Policy Research Working Papers, The World Bank.

Altman, M. (2008), 'Revisiting South African Employment Trends in the 1990s', *South African Journal of Economics* **76**(s2), S126–S147.

Alvaredo, F., Chancel, L., Piketty, T., Saez, E. & Zucman, G. (2018), World inequality report 2018., World inequality report 2018., Belknap Press.

Anghel, B., De la Rica, S. & Lacuesta, A. (2014), 'The impact of the great recession on employment polarization in Spain', *SERIEs* **5**(2), 143–171.

Ariza, J. & Bara, J. L. R. (2020), 'Technological change and employment in Brazil, Colombia and Mexico: Which workers are most affected?', *International Labour Review* **159**(2), 137–159.

Arrow, K. J., Chenery, H. B., Minhas, B. S. & Solow, R. M. (1961), 'Capital-Labor Substitution and Economic Efficiency', *The Review of Economics and Statistics* **43**(3), 225–250.

Autor, D. H. & Dorn, D. (2013), 'The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market', *The American Economic Review* **103**(5), 1553–1597.

Autor, D. H. & Handel, M. J. (2013), 'Putting Tasks to the Test: Human Capital, Job Tasks, and Wages', *Journal of Labor Economics* **31**(S1), 59–96.

Autor, D. H., Katz, L. F. & Kearney, M. S. (2005), Rising Wage Inequality: The Role of Composition and Prices, Working Paper 11628, National Bureau of Economic Research.

Autor, D. H., Katz, L. F. & Kearney, M. S. (2006), 'The Polarization of the U.S. Labor Market', *American Economic Review* **96**(2), 189–194.

Autor, D. H., Katz, L. F. & Krueger, A. B. (1998), 'Computing Inequality: Have Computers Changed the Labor Market?', *The Quarterly Journal of Economics* **113**(4), 1169–1213.

Autor, D. H., Levy, F. & Murnane, R. J. (2003), 'The Skill Content of Recent Technological Change: An Empirical Exploration', *The Quarterly Journal of Economics* **118**(4), 1279–1333.

Bailey, M. & Dynarski, S. (2011), *Inequality in Postsecondary Education. in Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*, Russell Sage Foundation, New York.

Bárány, Z. L. & Siegel, C. (2019), 'Job Polarization, Structural Transformation and Biased Technological Change', *Travail et emploi* **n° 157**(1), 25–44.

Barro, R. (2001), 'Human capital and growth', *American Economic Review* **91**(2), 12–17.

Barro, R. (2013), 'Education and economic growth', *Annals of Economics and Finance* **14**(2), 301–28.

Bartel, A. P. & Lichtenberg, F. R. (1987), 'The Comparative Advantage of Educated Workers in Implementing New Technology', *The Review of Economics and Statistics* **69**(1), 1–11.

Berman, E., Somanathan, R. & Tan, H. W. (2005), *Is Skill-Biased Technological Change Here Yet ? Evidence From Indian Manufacturing In The 1990*, Policy Research Working Papers, The World Bank.

Bhorat, H., Goga, S. & Stanwix, B. (2014), Skills-biased labour demand and the pursuit of inclusive growth in South Africa, Working Paper 2014/130, Helsinki: The United Nations University World Institute for Development Economics Research (UNU-WIDER).

Bhorat, H. & Hodge, J. (1999), 'Decomposing Shifts in Labour Demand in South Africa', *South African Journal of Economics* **67**(3), 155–168.

Biernacki, C., Celeux, G. & Govaert, G. (1999), 'An improvement of the NEC criterion for assessing the number of clusters in a mixture model', *Pattern Recognition Letters* **20**(3), 267–272.

Biernacki, C., Celeux, G. & Govaert, G. (2000), 'Assessing a mixture model for clustering with the integrated completed likelihood', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(7), 719–725.

Bøler, E. A. (2015), 'Technology-skill complementarity in a globalized world', *University of Oslo mimeo* .

Bollen, K. A., Guilkey, D. K. & Mroz, T. A. (1995), 'Binary outcomes and endogenous explanatory variables: Tests and solutions with an application to the demand for contraceptive use in tunisia', *Demography* **32**(1), 111–131.

Bornstein, M. & Bradley, R. (2002), *Socioeconomic Status, Parenting, and Child Development*, 1st edn, Routledge, New York.

Bound, J. & Johnson, G. (1992), 'Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations', *The American Economic Review* **82**(3), 371–392.

Branson, N. & Wittenberg, M. (2014), 'Reweighting South African National Household Survey Data to Create a Consistent Series Over Time: A Cross-Entropy Estimation Approach', *South African Journal of Economics* **82**(1), 19–38.

Bresnahan, T. F., Brynjolfsson, E. & Hitt, L. M. (2002), 'Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence', *The Quarterly Journal of Economics* **117**(1), 339–376.

Brossard, M., Coury, D. & Mambo, M., eds (2010), *The Education System in Malawi*, The World Bank.

Burchfield, S. A. & Kadzamira, E. C. (1996), 'Malawi GABLE Social Mobilisation Campaign Activities: A review of research and reports on findings of KAP follow-up', *Washington*, *DC: Creative Associates International Inc. and USAID* .

Burger, R. P. & Teal, F. J. (2015), 'The Effect of Schooling on Worker Productivity: Evidence from a South African Industry Panel', *Journal of African Economies* **24**(5), 629–644.

Burger, R. & Yu, D. (2007), Wage Trends in Post-Apartheid South Africa: Constructing an Earnings Series from Household Survey Data, SSRN Scholarly Paper ID 966118, Social Science Research Network, Rochester, NY.

Card, D. & DiNardo, J. E. (2002), 'Skill-Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles', *Journal of Labor Economics* **20**(4), 733–783.

Caro, D. H., McDonald, J. T. & Willms, J. D. (2009), 'Socio-economic Status and Academic Achievement Trajectories from Childhood to Adolescence', *Canadian Journal of Education* **32**(3), 558–590.

Casale, D., Muller, C. & Posel, D. (2005), 'Two Million Net New Jobs': A Reconsideration of the Rise in Employment in South Africa, 1995 - 2003, Working Paper, University of Cape Town.

Celeux, G. & Soromenho, G. (1996), 'An entropy criterion for assessing the number of clusters in a mixture model', *Journal of Classification* **13**(2), 195–212.

Chimombo, J. (2009), 'Changing patterns of access to basic education in Malawi: A story of a mixed bag?', *Comparative Education* **45**(2), 297–312.

Coady, D. & Dizioli, A. (2018), 'Income inequality and education revisited: Persistence, endogeneity and heterogeneity', *Applied Economics* **50**(25), 2747–2761.

Colclough, C., Kingdon, G. & Patrinos, H. (2010), 'The Changing Pattern of Wage Returns to Education and its Implications', *Development Policy Review* **28**(6), 733–747.

Coleman, J. S. (1968), 'Equality of Educational Opportunity', *Equity & Excellence in Education* **6**(5), 19–28.

Conte, A. & Vivarelli, M. (2011), 'Imported Skill-Biased Technological Change in Developing Countries', *The Developing Economies* **49**(1), 36–65.

Corak, M. (2013), 'Income inequality, equality of opportunity, and intergenerational mobility', *Journal of Economic Perspectives* **27**(3), 79–102.

Cortinovis, I., Vella, V. & Ndiku, J. (1993), 'Construction of a socio-economic index to facilitate analysis of health data in developing countries', *Social Science & Medicine* **36**(8), 1087–1097.

Crane, J. (1991), 'The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing', *American Journal of Sociology* **96**(5), 1226–1259.

Crouch, L. A. & Vinjevold, P. (2006), 'South Africa: Access before quality, and what to do now?', *ResearchGate* **10**(1).

Das, M. & Hilgenstock, B. (2018), The Exposure to Routinization: Labor Market Implications for Developed and Developing Economies, SSRN Scholarly Paper ID 3221220, Social Science Research Network, Rochester, NY.

Dasgupta, A. & Raftery, A. E. (1998), 'Detecting Features in Spatial Point Processes with Clutter via Model-Based Clustering', *Journal of the American Statistical Association* **93**(441), 294–302.

De La Grandville, O. (1989), 'In Quest of the Slutsky Diamond', *The American Economic Review* **79**(3), 468–481.

Deaton, A. & Grosh, M. (2000), 'Consumption, in designing household survey questionnaires for developing countries: Lessons from 15 years of living standards measurement study', *Washington, DC: The World Bank* .

Deb, P. & Trivedi, P. K. (1997), 'Demand for Medical Care by the Elderly: A Finite Mixture Approach', *Journal of Applied Econometrics* **12**(3), 313–336.

Dempster, A. P., Laird, N. M. & Rubin, D. B. (1977), 'Maximum Likelihood from Incomplete Data Via the EM Algorithm', *Journal of the Royal Statistical Society: Series B (Methodological)* **39**(1), 1–22.

DiNardo, J., Fortin, N. M. & Lemieux, T. (1996), 'Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach', *Econometrica* **64**(5), 1001–1044.

Dorn, D. & Autor, D. (2013), 'The growth of low-skill service jobs and the polarization of the US labor market', *American Economic Review* **103**(5), 1553–1597.

Duernecker, G. & Herrendorf, B. (2017), Structural Transformation of Occupation Employment, Technical Report 1239, Society for Economic Dynamics.

Dunne, J. & Edwards, L. (2006), Trade Technology and Employment: A case Study of South Africa, Working Paper 0602, Department of Accounting, Economics and Finance, Bristol Business School, University of the West of England, Bristol.

Eagle, R. F., Romanczyk, R. G. & Lenzenweger, M. F. (2010), 'Classification of children with autism spectrum disorders: A finite mixture modeling approach to heterogeneity', *Research in Autism Spectrum Disorders* **4**(4), 772–781.

Falkingham, J. & Namazie, C. (2002), *Measuring Health and Poverty: A Review of Approaches to Identifying the Poor*, DFID Health Systems Resource Centre, London.

Fantom, N. & Serajuddin, U. (2016), *The World Bank's Classification of Countries by Income*, Policy Research Working Papers, The World Bank.

Fernández-Macías, E. & Bisello, M. (2016), 'A framework for measuring tasks across occupations'.

Fernández-Macías, E. & Hurley, J. (2017), 'Routine-biased technical change and job polarization in Europe', *Socio-Economic Review* **15**(3), 563–585.

Fernandez, R. M. (2001), 'Skill-Biased Technological Change and Wage Inequality: Evidence from a Plant Retooling', *American Journal of Sociology* **107**(2), 273–320.

Fields, G. (1989), 'CHANGES IN POVERTY AND INEQUALITY IN DEVELOPING COUNTRIES', *The World Bank Research Observer* **4**(2), 167–185.

Fields, G. (2011), 'Labor market analysis for developing countries', *Labour Economics* **18:S16–S22**.

Filmer, D. & 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.

Firpo, S., Fortin, N. M. & Lemieux, T. (2011), Occupational Tasks and Changes in the Wage Structure, SSRN Scholarly Paper ID 1778886, Social Science Research Network, Rochester, NY.

Fleisch, B. (2008), *Primary Education in Crisis: Why South African Schoolchildren Underachieve in Reading and Mathematics*, Juta and Company Ltd.

Fonseca, T., Lima, F. & Pereira, S. C. (2018), 'Job polarization, technological change and routinization: Evidence for Portugal', *Labour Economics* **51**, 317–339.

Glewwe, P., Maïga, E. & Zheng, H. (2014), 'The Contribution of Education to Economic Growth: A Review of the Evidence, with Special Attention and an Application to Sub-Saharan Africa', *World Development* **59**, 379–393.

Goldin, C. & Katz, L. (2007*a*), 'The race between education and technology: The evolution of U.S. educational wage differentials, 1890 to 2005'.

Goldin, C. & Katz, L. F. (2007*b*), The Race between Education and Technology: The Evolution of U.S. Educational Wage Differentials, 1890 to 2005, Working Paper 12984, National Bureau of Economic Research.

Goldin, C. & Katz, L. F. (2010), *The Race between Education and Technology*, 6390th edition edn, Belknap Press: An Imprint of Harvard University Press, Cambridge, Massachusetts London, England.

Goos, M. & Manning, A. (2007), 'Lousy and Lovely Jobs: The Rising Polarization of Work in Britain', *The Review of Economics and Statistics* **89**(1), 118–133.

Goos, M., Manning, A. & Salomons, A. (2014), 'Explaining Job Polarization: Routine-Biased Technological Change and Offshoring', *The American Economic Review* **104**(8), 2509–2526.

Górka, S., Hardy, W., Keister, R. & Lewandowski, P. (2017), Tasks and skills in European labour markets. Background paper for the World Bank report "Growing United: Upgrading Europe's Convergence Machine", IBS Research Reports 03/2017, Instytut Badan Strukturalnych.

Government of Malawi, . (1994), Poverty Alleviation Programme, Technical report, Ministry of Economic Planning and Development, Lilongwe.

Government of Malawi, . (1998), Vision 2020: National Long Term Development Perspective for Malawi, Technical report, Ministry of Economic Planning and Development, Lilongwe.

Government of Malawi, . (2002), Malawi Poverty Reduction Strategy Paper, Technical report, Ministry of Economic Planning and Development, Lilongwe.

Government of Malawi, . (2012), Malawi Growth and Development Strategy, Technical report, Ministry of Economic Planning and Development, Lilongwe.

Government of Malawi, . (2016), Malawi Growth Development Strategy (MGDS III) – 2017-2022, Technical report, Malawi Government.

Green, F. (2012), 'Employee involvement, technology and evolution in job skills: A task-based analysis', *ILR Review* **65**(1), 36–67.

Greenwood, J. & Yorukoglu, M. (1997), '1974', *Carnegie-Rochester Conference Series on Public Policy* **46**, 49–95.

Griliches, Z. (1969), 'Capital-Skill Complementarity', *The Review of Economics and Statistics* **51**(4), 465–468.

Griliches, Z. (1970), 'Notes on the Role of Education in Production Functions and Growth Accounting', *Education, Income, and Human Capital* pp. 71–127.

Gutierrez, R. G., Carroll, R. J., Wang, N., Lee, G.-H. & Taylor, B. H. (1995), 'Analysis of Tomato Root Initiation Using a Normal Mixture Distribution', *Biometrics* **51**(4), 1461–1468.

Gwatkin, D., Rustein, S., Johnson, K., Pande, R. P. & Wagstaff, A. (2000), 'Socio-economic differences in health, nutrition, and population in Brazil', *Country reports on HNP and poverty Washington, D.C. : World Bank Group. .*

Hanushek, E. (2013), 'Economic growth in developing countries: The role of human capital', *Economics of Education Review* **37**, 204–212.

Hanushek, E. A. & Woessmann, L. (2008), 'The Role of Cognitive Skills in Economic Development', *Journal of Economic Literature* **46**(3), 607–668.

Harrigan, J., Reshef, A. & Toubal, F. (2018), Techies, Trade, and Skill-Biased Productivity, Working Paper 25295, National Bureau of Economic Research.

Hawkins, D. M., Muller, M. W. & Krooden, J. A. (1982), Cluster analysis, *in* 'Topics in Applied Multivariate Analysis', Cambridge University Press, Cambridge.

Hertzman, C. & Wiens, M. (1996), 'Child development and long-term outcomes: A population health perspective and summary of successful interventions', *Social Science & Medicine* **43**(7), 1083–1095.

Hornstein, A., Krusell, P. & Violante, G. L. (2005), 'The Replacement Problem in Frictional Economies: A near-Equivalence Result', *Journal of the European Economic Association* **3**(5), 1007–1057.

Howe, L. D., Hargreaves, J. R. & Huttly, S. R. (2008), 'Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries', *Emerging Themes in Epidemiology* **5**(1), 3.

Hox, J. J., Moerbeek, M. & van de Schoot, R. (2017), *Multilevel Analysis: Techniques and Applications, Third Edition*, 3rd edition edn, Routledge, New York, NY.

Izenman, A. J. & Sommer, C. J. (1988), 'Philatelic Mixtures and Multimodal Densities', *Journal of the American Statistical Association* **83**(404), 941–953.

Janosz, M., LeBlanc, M., Boulerice, B. & Tremblay, R. (1997), 'Disentangling the weight of school dropout predictors: A test on two longitudinal samples', *Journal of Youth and Adolescence* **26**(6), 733–762.

Jehle, G. (2010), *Advanced Microeconomic Theory*, 3 edition edn, Pearson.

Jobson, J. (1992), *Applied Multivariate Data Analysis: Volume II: Categorical and Multivariate Methods*, Springer Texts in Statistics, Applied Multivariate Data Analysis, 1st edn, Springer-Verlag, New York.

Kadzamira, E. & 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.

Kattan, R. B. (2006), 'Implementation of free basic education policy', *Education Working Paper Series*, *World Bank. 133.* .

Katz, L. & Autor, D. (1999), Changes in the wage structure and earnings inequality, *in* O. Ashenfelter & D. Card, eds, 'Handbook of Labor Economics', first edn, Vol. 3, Part A, Elsevier, pp. 1463–1555.

Katz, L. F. & Murphy, K. M. (1992), 'Changes in Relative Wages, 1963–1987: Supply and Demand Factors', *The Quarterly Journal of Economics* **107**(1), 35–78.

Keribin, C. (2000), 'Consistent Estimation of the Order of Mixture Models', *Sankhyā: The Indian Journal of Statistics*, *Series A (1961-2002)* **62**(1), 49–66.

Kerr, A., Lam, D. & Wittenberg, M. (2013), 'Post-apartheid labour market series 1994-2012.[Dataset]', *Datafirst*, *Cape Town* **Version 2.1**, 1994–2012.

Kim, E., Hong, A. & Hwang, J. (2019), 'Polarized labor demand owing to routine-biased technological change: The case of Korea from 1993 to 2015', *Telematics and Informatics* **39**, 1–10.

Kingdon, G. & Knight, J. (2007), 'Unemployment in South Africa, 1995–2003: Causes, Problems and Policies', *Journal of African Economies* **16**(5), 813–848.

Klump, R., McAdam, P. & Willman, A. (2007), 'Factor Substitution and Factor-Augmenting Technical Progress in the United States: A Normalized Supply-Side System Approach', *The Review of Economics and Statistics* **89**(1), 183–192.

Klump, R., McAdam, P. & Willman, A. (2012), 'The Normalized CES Production Function: Theory and Empirics', *Journal of Economic Surveys* **26**(5), 769–799.

Korkovelos, A., Khavari, B., Sahlberg, A., Howells, M. & Arderne, C. (2019), 'The Role of Open Access Data in Geospatial Electrification Planning and the Achievement of SDG7. An OnSSET-Based Case Study for Malawi', *Energies* **12**(7), 1395.

Kotzé, J. & Van der Berg, S. (2015), 'Investigating cognitive performance differentials by socio-economic status: Comparing Sub-Saharan Africa and Latin America', *Lisbon* p. 34.

Kreuser, F., Burger, R. & Rankin, N. (2015), *The Elasticity of Substitution and Labour-Displacing Technical Change in Post-Apartheid South Africa*, UNU-WIDER.

Krusell, P., Ohanian, L. E., Ríos-Rull, J.-V. & Violante, G. L. (2000), 'Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis', *Econometrica* **68**(5), 1029–1053.

Kunje, D. (2002), 'The Malawi integrated in-service teacher education programme: An experiment with mixed-mode training', *International Journal of Educational Development* **22**(3), 305–320.

Leamer, E. E. (1996), In Search of Stolper-Samuelson Effects on U.S. Wages, Working Paper 5427, National Bureau of Economic Research.

Lee, H. & Sim, S. (2016), 'Global Outsourcing and Wage Inequality in Middle-Income Countries: Evidence from South Korea', *Global Economic Review* **45**(1), 19–41.

Lee, S. Y. T. & Shin, Y. (2017), Horizontal and Vertical Polarization: Task-Specific Technological Change in a Multi-Sector Economy, Working Paper 23283, National Bureau of Economic Research.

Lee, V. E. & Zuze, T. L. (2011), 'School Resources and Academic Performance in Sub-Saharan Africa', *Comparative Education Review* **55**(3), 369–397.

Lenzenweger, M. F., McLachlan, G. & Rubin, D. B. (2007), 'Resolving the latent structure of schizophrenia endophenotypes using expectation-maximization-based finite mixture modeling', *Journal of Abnormal Psychology* **116**(1), 16–29.

León-Ledesma, M. A., McAdam, P. & Willman, A. (2010), 'Identifying the Elasticity of Substitution with Biased Technical Change', *American Economic Review* **100**(4), 1330–1357.

Leroux, B. G. (1992), 'Consistent Estimation of a Mixing Distribution', *The Annals of Statistics* **20**(3), 1350–1360.

Lewin, K. M. (2009), 'Access to education in sub-Saharan Africa: Patterns, problems and possibilities', *Comparative Education* **45**(2), 151–174.

133

Li, J., Ray, S. & Lindsay, B. G. (2007), 'A nonparametric statistical approach to clustering via mode identification', *Journal of Machine Learning Research* **8**(59), 1687–1723.

Ligomeka, B. (2002), 'Free primary education backfires', http://africa.peacelink.org/newsfromafrica/articles/

Lo, Y., Mendell, N. R. & Rubin, D. B. (2001), 'Testing the number of components in a normal mixture', *Biometrika* **88**(3), 767–778.

Longmuir, M., Schröder, C. & Targa, M. (2020), De-routinization of Jobs and Polarization of Earnings – Evidence from 35 Countries, Technical Report 796, LIS Cross-National Data Center in Luxembourg.

Maggie Fu, X., Bao, Q., Xie, H. & Fu, X. (2020), 'Diffusion of industrial robotics and inclusive growth: Labour market evidence from cross country data', *Journal of Business Research* .

Mahutga, M. C., Curran, M. & Roberts, A. (2018), 'Job tasks and the comparative structure of income and employment: Routine task intensity and offshorability for the LIS*:', *International Journal of Comparative Sociology* .

Maitra, R. & Melnykov, V. (2018), 'Assessing significance in finite mixture models', *Statistics Publications* .

Majgaard, K. & Mingat, A. (2012), *Education in Sub-Saharan Africa: A Comparative Analysis*, World Bank Publications.

Manda, M. & Ben Dhaou, S. (2019), Responding to the challenges and opportunities in the 4th Industrial revolution in developing countries, *in* 'Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance - ICEGOV2019', ACM Press, New York, New York, USA, pp. 244–253.

Marcolin, L., Miroudot, S. & Squicciarini, M. (2016), 'The Routine Content Of Occupations: New Cross-Country Measures Based On PIAAC'.

Marriott, F. H. C. (1974), *The Interpretation of Multiple Observations*, Academic Press, London.

Matthes, B., Christoph, B., Janik, F. & Ruland, M. (2014), 'Collecting information on job tasks—an instrument to measure tasks required at the workplace in a multi-topic survey', *Journal for Labour Market Research* **47**(4), 273–297.

McKenzie, D. J. (2005), 'Measuring inequality with asset indicators', *Journal of Population Economics* **18**(2), 229–260.

McLachlan, G. J. & Basford, K. E. (1988), *Mixture Models: Inference and Applications to Clustering*, New York: Marcel Dekker.

McLachlan, G. J. & Chang, S. U. (2004), 'Mixture modelling for cluster analysis', *Statistical Methods in Medical Research* **13**(5), 347–361.

McLachlan, G. J., Lee, S. X. & Rathnayake, S. I. (2019), 'Finite Mixture Models', *Annual Review of Statistics and Its Application* **6**(1), 355–378.

McLachlan, G. J. & Peel, D. (2004), *Finite Mixture Models*, John Wiley & Sons, New York.

McLachlan, G. J. & Rathnayake, S. (2014), 'On the number of components in a Gaussian mixture model', *WIREs Data Mining and Knowledge Discovery* **4**(5), 341–355.

McLachlan, G. & Peel, D. (2000), *Finite Mixture Models*, John Wiley & Sons Inc, New York: New York.

Medina, C. & Suárez, C. M. P. (2010), Technical Change and Polarization of the Labor Market: Evidence for Brazil, Colombia and Mexico, Technical Report 007269, BANCO DE LA REPÚBLICA.

Melnykov, V. & Maitra, R. (2010), 'Finite mixture models and model-based clustering', *Statistics Surveys* **4**, 80–116.

Melnykov, V., Sarkar, S. & Melnykov, Y. (2021), 'On finite mixture modeling and model-based clustering of directed weighted multilayer networks', *Pattern Recognition* **112**, 107641.

Meschi, E. & Vivarelli, M. (2009), 'Trade and Income Inequality in Developing Countries', *World Development* **37**(2), 287–302.

Milner, G., Chimombo, J., Banda, T. & Mchikoma, C. (2001), 'The quality of primary education in Malawi:(an interim report) Some policy suggestions based on a survey of schools', *Internation Instititute of Planning - UNESCO* .

Ministry of Education (1994), Education Statistics., Technical report, Malawi Government, Lilongwe.

Ministry of Education (1997), Education Statistics, Technical report, Malawi Government, Lilongwe.

Moloi, M. Q. & Strauss, J. (2005), 'The SACMEQ II Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education', *SACMEQ* **26**, 2009.

Montgomery, M. R., Gragnolati, M., Burke, K. A. & Paredes, E. (2000), 'Measuring living standards with proxy variables', *Demography* **37**(2), 155–174.

Mueller, C. W. & Parcel, T. L. (1981), 'Measures of Socioeconomic Status: Alternatives and Recommendations', *Child Development* **52**(1), 13–30.

Murimba, S. (2005), 'The Southern and Eastern Africa Consortium for Monitoring Educational Quality (Sacmeq): Mission, Approach and Projects', *Prospects* **35**(1), 75–89.

Mussa, R. (2017), 'Poverty and Inequality in Malawi: Trends, Prospects, and Policy Simulations', https://mpra.ub.uni-muenchen.de/75979/.

Namibia Statistics Agency & World Bank (2017), *Does Fiscal Policy Benefit the Poor and Reduce Inequality in Namibia?*, World Bank, Washington, DC.

National Statistical Office, . (2008), *Welfare Monitoring Survey 2008*, NSO, Zomba.

National Statistical Office, . (2012), *Integrated Household Survey 2010/11 (IHS3*, NSO, Zomba.

National Statistical Office, . (2017), *Integrated Household Survey 2016/17 (IHS4*, NSO, Zomba.

National Statistical Office, . (2019), Malawi Population and Housing Census (PHC), Technical report, Malawi Government, Zomba.

Nogueira, V. d. A. (2015), *Is That Where You Work Or What You Do? Understanding Job Polarization in Brazil*, Simon Fraser University.

Nylund, K. L., Asparouhov, T. & Muthén, B. O. (2007), 'Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study', *Structural Equation Modeling: A Multidisciplinary Journal* **14**(4), 535–569.

OECD, . (2011).

Onwujekwe, O., Hanson, K. & Fox-Rushby, J. (2006), 'Some indicators of socio-economic status may not be reliable and use of indices with these data could worsen equity', *Health Economics* **15**(6), 639–644.

Pierre, G., Sanchez Puerta, M. L., Valerio, A. & Rajadel, T. (2014), 'STEP skills measurement surveys: Innovative tools for assessing skills'.

Polidano, C., Hanel, B. & Buddelmeyer, H. (2013), 'Explaining the socio-economic status school completion gap', *Education Economics* **21**(3), 230–247.

Pritchett, L. (1999), *Where Has All the Education Gone?*, Policy Research Working Papers). The World Bank.

Raquel, S. L. & Biagi, F. (2018), The Routine Biased Technical Change hypothesis: A critical review, Technical Report JRC113174, Joint Research Centre (Seville site).

Raudenbush, S. W. & Bryk, A. S. (2001), *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edition edn, SAGE Publications, Inc, Thousand Oaks.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F. & Jr, R. T. C. (2004), *HLM 6: Hierarchical Linear and Nonlinear Modeling*, Scientific Software International, Inc., Lincolnwood, Ill.

Raudenbush, S. W. & Willms, J. (1995), 'The Estimation of School Effects', *Journal of Educational and Behavioral Statistics* **20**(4), 307–335.

Ravallion, M. (2006), 'Looking beyond averages in the trade and poverty debate', *World Development* **34**(8), 1374–1392.

Ravallion, M. (2014), 'Income inequality in the developing world', *Science* **344**(6186), 851–855.

Reardon, S. F. (2013), 'The Widening Income Achievement Gap', *Educational Leadership* **70**(8), 10–16.

Roeder, K. (1994), 'A Graphical Technique for Determining the Number of Components in a Mixture of Normals', *Journal of the American Statistical Association* **89**(426), 487–495.

Roeder, K. & Wasserman, L. (1997), 'Practical Bayesian Density Estimation Using Mixtures of Normals', *Journal of the American Statistical Association* **92**(439), 894–902.

Ross, K. N., Dolata, S., Zuze, L. & Murimba, S. (2005), *The Conduct of the SACMEQ II Project in Kenya.*, Southern and Eastern Africa Consortium for Monitoring Educational Quality., Harare.

Ross, K. & Saito, M. (2008), SACMEQ III project international report: The condition of schooling and the quality of education in Southern and Eastern Africa (2000-2007), Technical Report Volume 1: Technical Report.

Ross, K. & Zuze, L. (2004), 'Traditional and alternative view of school system performance', *IIEP Newsletter). IIEP Newsletter* .

Rulof Burger, Rachel Jafta & Dieter von Fintel (2016), 'Affirmative action policies and the evolution of post-apartheid South Africa's racial wage gap', **2016**(66).

Rumberger, R. W. (1995), 'Dropping out of Middle School: A Multilevel Analysis of Students and Schools', *American Educational Research Journal* **32**(3), 583–625.

Rutstein, S. O. (2008), *The DHS Wealth Index: Approaches for Rural and Urban Areas*, Macro International, Calverton, Maryland, USA.

Rutstein, S. O. & Johnson, K. (2004), *The DHS Wealth Index*, ORC Macro, Calverton, Maryland, USA.

Saam, M. (2014), The identification of directed technical change revisited, ZEW Discussion Paper 14-127, ZEW - Leibniz Centre for European Economic Research.

Sahn, D. E. & Stifel, D. (2003), 'Exploring Alternative Measures of Welfare in the Absence of Expenditure Data', *Review of Income and Wealth* **49**(4), 463–489.

Schwarz, G. (1978), 'Estimating the Dimension of a Model', *Annals of Statistics* **6**(2), 461–464.

Sclove, S. L. (1987), 'Application of model-selection criteria to some problems in multivariate analysis', *Psychometrika* **52**(3), 333–343.

Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. (2016), 'Mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models', *The R journal* **8**(1), 289–317.

Sebastian, R. (2018), 'Explaining job polarisation in Spain from a task perspective', *SERIEs* **9**(2), 215–248.

Setel, P., Abeyasekera, S., Ward, P., Hemed, Y., Whiting, D., Mswia, R., Antoninis, M. & Kitange, H. (2003), 'Development, validation, and performance of a rapid consumption expenditure proxy for measuring income poverty in Tanzania: Experience from AMMP demographic surveillance sites.', *Development, validation, and performance of a rapid consumption expenditure proxy for measuring income poverty in Tanzania: Experience from AMMP demographic surveillance sites.* pp. 169–183.

Sewell, W. H. & Hauser, R. M. (1975), *Education, Occupation, and Earnings. Achievement in the Early Career*, Academic Press Inc.

Sirin, S. R. (2005), 'Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research', *Review of Educational Research* **75**(3), 417–453.

Soromenho, G. (1993), 'Comparing approaches for testing the number of components in a finite mixture model', *Computational Statistics* **9**(1), 65–78.

Spaull, N. (2013), 'Poverty & privilege: Primary school inequality in South Africa', *International Journal of Educational Development* **33**(5), 436–447.

Spitz-Oener, A. (2006), 'Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure', *Journal of Labor Economics* **24**(2), 235–270.

StataCorp (2013), *Stata 13 Base Reference Manual*, Stata Press., College Station, TX.

Stats SA, . (2019), Inequality Trends in South Africa: A multidimensional diagnostic of inequality, Technical Report 03-10-19, Statistics South Africa, Pretoria.

StatsSA (2012), South African Standard Classification of Occupations (SASCO), Technical report, Statistics South Africa, Pretoria.

Stiglitz, J. (2015), '8. Inequality and economic growth', *The Political Quarterly* **86**, 134–155.

Taylor, S. & Spaull, N. (2015), 'Measuring access to learning over a period of increased access to schooling: The case of Southern and Eastern Africa since 2000', *International Journal of Educational Development* **41**, 47–59.

Taylor, S. & Yu, D. (2009), The importance of socio-economic status in determining educational achievement in South Africa, Technical Report 01/2009, Stellenbosch University, Department of Economics.

Tinbergen, J. (1974), 'Substitution of Graduate by Other Labour*', *Kyklos* **27**(2), 217–226.

Tinbergen, J. (1975), *Income Differences: Recent Research*, North-Holland Publishing Company, Amsterdam.

United Nations SDSN, . (2016), 'United nations sustainable development solutions Network,(2016)', *Data for Development. A Needs Assessment for SDG monitoring and Statistical Capacity Development* .

USAID, . (2017), USAID Malawi National Reading Strategy Implementation Phase II Report 2017, Technical report, USAID Malawi.

Van der Berg, S. (2002), Education, poverty and inequality in South Africa, *in* 'In Conference of the Centre for the Study of African Economies on Economic Growth and Poverty in Africa'.

van der Berg, S. (2008), 'How effective are poor schools? Poverty and educational outcomes in South Africa', *Studies in Educational Evaluation* **34**(3), 145–154.

Van der Berg, S., Ferreira, T., Zuze, L., Van Wyk, C. & Smith, A. (2020), School Examination Outcomes and Trends in Namibia: Report on Two Linked Studies for the Review of the sNamibian National Examination and Assessment System, Technical report, Ministry of Education,Arts and Culture.

van Ewijk, R. & Sleegers, P. (2010), 'The effect of peer socioeconomic status on student achievement: A meta-analysis', *Educational Research Review* **5**(2), 134–150.

Violante, G. L. (2008), 'Skill-biased technical change', *The new Palgrave dictionary of economics* **2**.

Vyas, S. & Kumaranayake, L. (2006), 'Constructing socio-economic status indices: How to use principal components analysis', *Health Policy and Planning* **21**(6), 459–468.

Walker, J., Pearce, C., Boe, K. & Lawson, M. (2019), 'The Power of Education to Fight Inequality: How increasing educational equality and quality is crucial to fighting economic and gender inequality'.

White, K. R. (1982), 'The relation between socioeconomic status and academic achievement', *Psychological Bulletin* **91**(3), 461–481.

Williams Shanks, T. R. & Robinson, C. (2013), 'Assets, economic opportunity and toxic stress: A framework for understanding child and educational outcomes', *Economics of Education Review* **33**, 154–170.

Willms, J. (2004), *Reading Achievement in Canada and the United States: Findings from the OECD*, Programme of International Student Assessment.

Willms, J. D. (2003), *Ten Hypotheses about Socioeconomic Gradients and Community Differences in Children's Developmental Outcomes. Final Report*, Human Resources Development Canada.

Willms, J. D., Smith, T. M., Zhang, Y. & Tramonte, L. (2006), 'Raising and levelling the learning bar in central and Eastern Europe', *PROSPECTS* **36**(4), 411–418.

Windham, M. P. & Cutler, A. (1992), 'Information Ratios for Validating Mixture Analyses', *Journal of the American Statistical Association* **87**(420), 1188–1192.

Wittenberg, M. (2009), Weighing the value of Asset Proxies: The case of the Body Mass Index in South Africa, Technical Report 39, Southern Africa Labour and Development Research Unit, University of Cape Town.

Wood, A. (1998), 'Globalisation and the Rise in Labour Market Inequalities', *The Economic Journal* **108**(450), 1463–1482.

World Bank, . (2010), 'The Education System in Malawi.', *Africa education country status report*, *World Bank working paper ; no. 182. World Bank* .

World Bank, . (2017), Land for Inclusive development, Technical report, World Bank.

World Bank, . (2018), 'DevelopmentAid', https://www.developmentaid.org/.

World Bank, . (2019), 'Malawi - Electricity Access Project', https://projects.worldbank.org/en/projects-operations/project-detail/P164331.

World Trade Organization, . (2017), *World Trade Report 2017*.

Xiaofeng, M., Wenhui, D. & Aibao, Z. (2018), 'The Link Between Parental Absence and Poor Reading Comprehension: Evidence From the Left-Behind Children in Rural China', *Frontiers in Education* **3**.

APPENDIX A

Chapter 2 Appendix

A.1 PCA for SES Construction?

PCA is a data reduction method used to re-express multivariate data in fewer dimensions (Howe et al. 2008). The goal of this method is to reorient data so that numerous variables can be represented with relatively few 'components' that capture the maximum possible information and variation from the original variables. In simpler terms, PCA finds linearly weighted combinations of the asset variables that maximise the variance of the linear sum (Vyas & Kumaranayake 2006). Mathematically, the objective of PCA is to find components $w = (w_1, w_2, \ldots w_m)$ that are a linear combination $v = (v_1, v_2, \ldots v_m)'$ of the original variables $x = x_1, x_2, \ldots x_m$ that achieve maximum variance.

Consider, for example, a set of variables x_1 through x_m ,

$$
\begin{aligned}
w_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_m \\
&\vdots \\
w_m &= a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_m
\end{aligned}
\tag{A.1}
$$

where a_{mn} represents the weight for the mth principal component and the nth variable. In this way, the first component, w_1, is given by the linear combination of the original variables x_1 through x_m , each with its own respective weight. The second component, w_2 , would represent the second most information (variance) not captured by the first component.

Critically, these subsequent components, w_1 and w_2 , are uncorrelated, which in practice ensures that the two components do not capture the same information (Vyas & Kumaranayake 2006). This feature follows through to all subsequent components, ensuring that no two components represent the same information. Each additional component, therefore, summarises and captures

additional dimensions of the data, while explaining increasingly less proportions of the variation present in the original variables (Vyas & Kumaranayake 2006). Highly correlated original variables would therefore require fewer components to capture all the available information, since the first principal component would already have captured the vast majority of it (Howe et al. 2008). When a wealth index is constructed in this way, the first principal component is assumed to represent the household's wealth and, by extension, their socio-economic status.

APPENDIX B

Chapter 3 Appendix

B.1 Kernel densities of latent classes within wealth index distributions by location
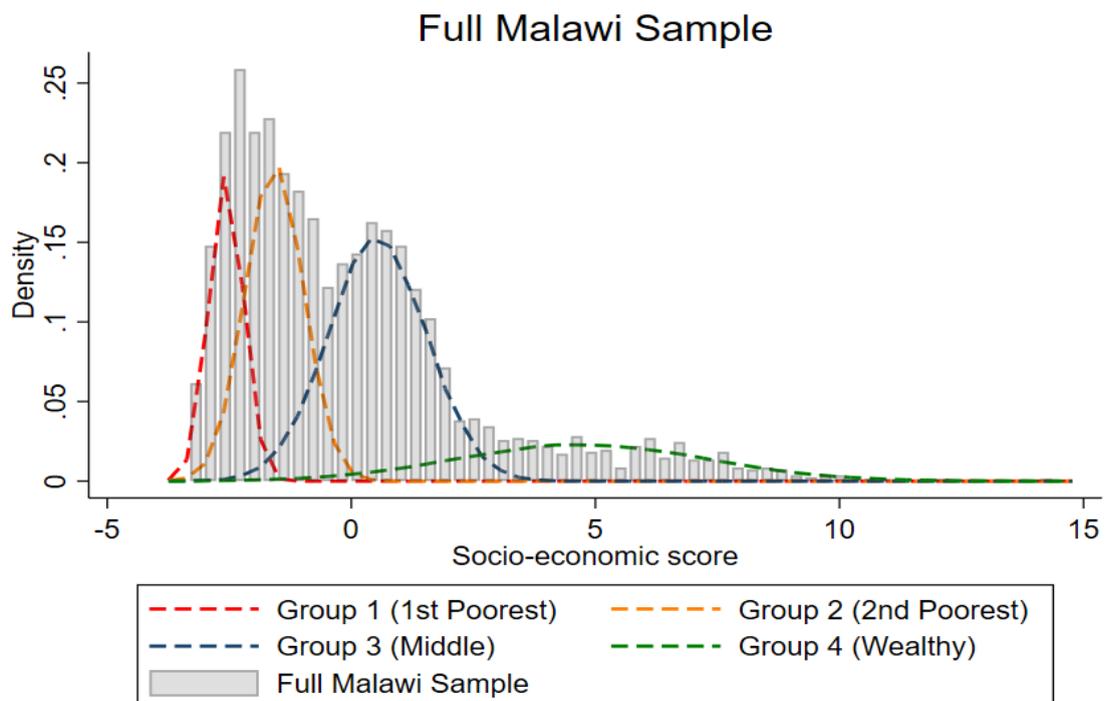
Figure B.1: Kernel densities of wealth classes and histogram of wealth index for the full Malawi sample
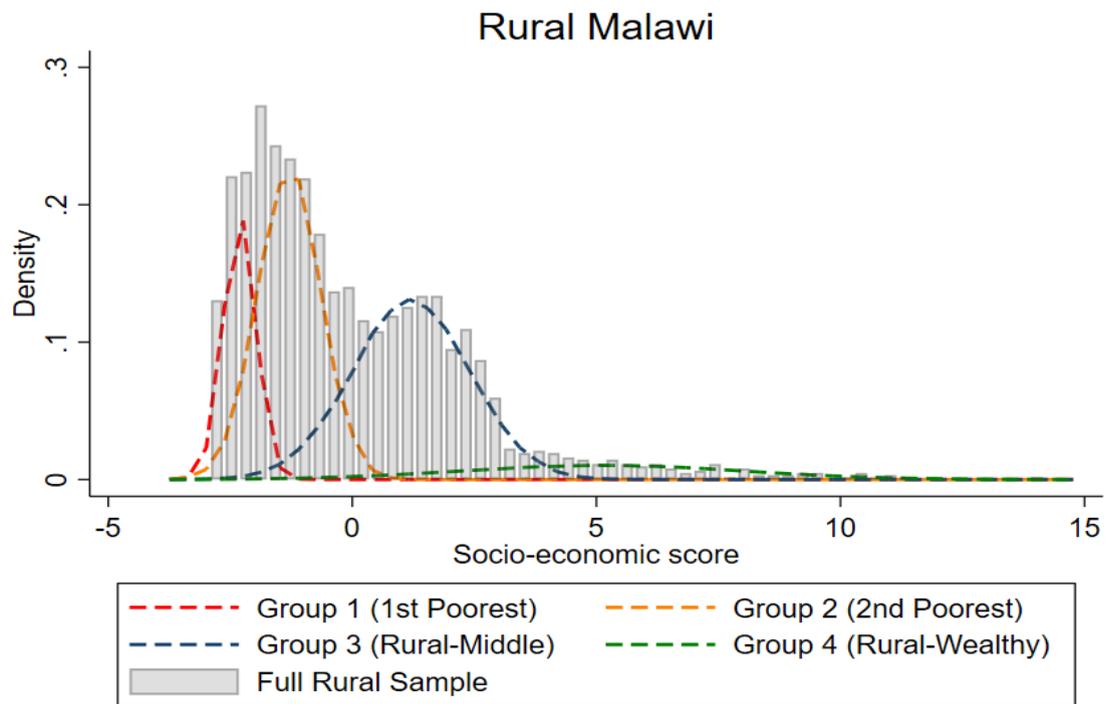
Figure B.2: Kernel densities of wealth classes and histogram of wealth index for the rural Malawi sample

B.2 Asset ownership among urban and rural classes

Table B.1: Asset ownership between urban class 1 (poorest) and urban class 2 (middle)

Variable description	Class 1	Class 2	Diff (C2-C1)	P-Value
Newspaper	0.2103	0.4339	0.2236	0
Magazine	0.1326	0.2572	0.1246	0.01
Clock	0.6231	0.8007	0.1776	0
Piped water	0.2157	0.4313	0.2156	0
Bore hole	0.3655	0.3406	-0.0249	0.616
Table	0.3957	0.7048	0.3091	0
Bed	0.422	0.7803	0.3583	0
Private study area	0.1011	0.2017	0.1006	0.004
Bicycle	0.3675	0.5811	0.2136	0
Horse cart	0.0371	0.0313	-0.0058	0.762
Car	0	0.028	0.028	0.007
Motorcycle	0	0.023	0.023	0.023
Tractor	0	0.0127	0.0127	0.056
Electricity	0	0.128	0.128	0
Refrigerator	0	0.0132	0.0132	0.03
Air conditioner	0	0.0017	0.0017	0.331
Electric fan	0	0.0126	0.0126	0.088
Washing machine	0	0	0	
Vacuum	0	0	0	
Computer	0	0	0	
Internet	0	0.0037	0.0037	0.32
Radio	0.8338	0.9378	0.104	0.001
TV	0.0266	0.2016	0.175	0
VCR	0	0.0369	0.0369	0.002
DVD player	0	0.0288	0.0288	0.006
CD player	0	0.0553	0.0553	0.002
Cassette	0.0671	0.2636	0.1965	0
Camera	0.0406	0.1314	0.0908	0
Camera (digital)	0	0.0373	0.0373	0.006
Video camera	0	0.0101	0.0101	0.095
Telephone/Mobile Phone	0.1585	0.5694	0.4109	0

Table B.2: Asset ownership between urban class 2 (middle) and urban class 3 (richest

Variable description	Class 2	Class 3	Diff (C3-C2)	P-Value
Newspaper	0.4339	0.6887	0.2548	0
Magazine	0.2572	0.3989	0.1417	0.005
Clock	0.8007	0.9283	0.1276	0
Piped water	0.4313	0.7141	0.7141	0
Bore hole	0.3406	0.2638	-0.0768	0.172
Table	0.7048	0.845	0.1402	0.007
Bed	0.7803	0.8896	0.1093	0.002
Private study area	0.2017	0.3644	0.1627	0.001
Bicycle	0.5811	0.5732	-0.0079	0.896
Horse cart	0.0313	0.0684	0.0371	0.06
Car	0.028	0.2264	0.1984	0
Motorcycle	0.023	0.117	0.094	0
Tractor	0.0127	0.0311	0.0184	0.231
Electricity	0.128	0.9483	0.8203	0
Refrigerator	0.0132	0.511	0.4978	0
Air conditioner	0.0017	0.0842	0.0825	0.016
Electric fan	0.0126	0.5722	0.5596	0
Washing machine	0	0.0366	0.0366	0.016
Vacuum	0	0.0509	0.0509	0.002
Computer	0	0.0459	0.0459	0.002
Internet	0.0037	0.0239	0.0202	0.112
Radio	0.9378	0.9791	0.0413	0.044
TV	0.2016	0.808	0.6064	0
VCR	0.0369	0.5326	0.4957	0
DVD player	0.0288	0.5948	0.566	0
CD player	0.0553	0.4836	0.4283	0
Cassette	0.2636	0.6344	0.3708	0
Camera	0.1314	0.269	0.1376	0
Camera (digital)	0.0373	0.1382	0.1009	0
Video camera	0.0101	0.1239	0.1138	0
Telephone/Mobile Phone	0.5694	0.9473	0.3779	0

Table B.3: Asset ownership between rural class 1 (poorest) and rural class 2 (middle)

Variable description	Class 1	Class 2	Diff (C2-C1)	P-Value
Newspaper	0.0524	0.1947	0.1423	0
Magazine	0.0278	0.17	0.1422	0
Clock	0.3888	0.597	0.2082	0
Piped water	0.1286	0.2017	0.0731	0.023
Bore hole	0.4747	0.5334	0.0587	0.108
Table	0.1889	0.5588	0.3699	0
Bed	0.1809	0.5188	0.3379	0
Private study area	0.0629	0.2145	0.1516	0
Bicycle	0.4378	0.5915	0.1537	0
Horse cart	0.0185	0.0478	0.0293	0.024
Car	0.0017	0.0105	0.0088	0.045
Motorcycle	0	0.0146	0.0146	0.002
Tractor	0	0.0033	0.0033	0.164
Electricity	0	0.0025	0.0025	0.316
Refrigerator	0	0	0	
Air conditioner	0	0	0	
Electric fan	0	0	0	
Washing machine	0	0	0	
Vacuum	0	0	0	
Computer	0	0	0	
Internet	0	0	0	
Radio	0.7641	0.8784	0.1143	0
TV	0	0.0217	0.0217	0.001
VCR	0	0.0045	0.0045	0.152
DVD player	0	0.006	0.006	0.114
CD player	0.0017	0.0296	0.0279	0
Cassette	0.0463	0.2486	0.2023	0
Camera	0.0111	0.0523	0.0412	0
Camera (digital)	0.0042	0.0104	0.0062	0.212
Video camera	0	0.0035	0.0035	0.15
Telephone/Mobile Phone	0.0083	0.1854	0.1771	0

Table B.4: Asset ownership between the two middle rural classes (2 and 3)

Variable description	Class 2	Class 3	Diff (C3-C2)	P-Value
Newspaper	0.1947	0.2708	0.0761	0.003
Magazine	0.17	0.2111	0.0411	0.053
Clock	0.597	0.7579	0.1609	0
Piped water	0.2017	0.2689	0.0672	0.026
Bore hole	0.5334	0.5583	0.0249	0.336
Table	0.5588	0.7156	0.1568	0
Bed	0.5188	0.7022	0.1834	0
Private study area	0.2145	0.2693	0.0548	0.048
Bicycle	0.5915	0.7094	0.1179	0
Horse cart	0.0478	0.0827	0.0349	0.013
Car	0.0105	0.0504	0.0399	0
Motorcycle	0.0146	0.0499	0.0353	0.001
Tractor	0.0033	0.0176	0.0143	0.015
Electricity	0.0025	0.0846	0.0821	0
Refrigerator	0	0.0137	0.0137	0.009
Air conditioner	0	0.0019	0.0019	0.32
Electric fan	0	0.0007	0.0007	0.32
Washing machine	0	0.0006	0.0006	0.324
Vacuum	0	0	0	
Computer	0	0.0029	0.0029	0.086
Internet	0	0.0032	0.0032	0.316
Radio	0.8784	0.8786	0.0002	0.993
TV	0.0217	0.1756	0.1539	0
VCR	0.0045	0.0315	0.027	0.002
DVD player	0.006	0.0343	0.0283	0
CD player	0.0296	0.0884	0.0588	0.001
Cassette	0.2486	0.3339	0.0853	0.001
Camera	0.0523	0.0903	0.038	0.014
Camera (digital)	0.0104	0.0324	0.022	0.029
Video camera	0.0035	0.0131	0.0096	0.05
Telephone/Mobile Phone	0.1854	0.4429	0.2575	0

Table B.5: Asset ownership between rural class 3 (upper middle) and rural class 4 (richest)

Variable description	Class 3	Class 4	Diff (C4-C3)	P-Value
Newspaper	0.2708	0.3765	0.1057	0.134
Magazine	0.2111	0.3442	0.1331	0.032
Clock	0.7579	0.8493	0.0914	0.106
Piped water	0.2689	0.4741	0.2052	0.001
Bore hole	0.5583	0.4331	-0.1252	0.017
Table	0.7156	0.8294	0.1138	0.005
Bed	0.7022	0.8392	0.137	0.005
Private study area	0.2693	0.4113	0.142	0.009
Bicycle	0.7094	0.7269	0.0175	0.787
Horse cart	0.0827	0.155	0.0723	0.031
Car	0.0504	0.2368	0.1864	0
Motorcycle	0.0499	0.2126	0.1627	0.001
Tractor	0.0176	0.0986	0.081	0.01
Electricity	0.0846	0.7066	0.622	0
Refrigerator	0.0137	0.3067	0.293	0
Air conditioner	0.0019	0.0495	0.0476	0.093
Electric fan	0.0007	0.2599	0.2592	0
Washing machine	0.0006	0.0388	0.0382	0.015
Vacuum	0	0.0306	0.0306	0.073
Computer	0.0029	0.0785	0.0756	0.001
Internet	0.0032	0.035	0.0318	0.054
Radio	0.8786	0.9344	0.0558	0.06
TV	0.1756	0.8608	0.6852	0
VCR	0.0315	0.5202	0.4887	0
DVD player	0.0343	0.5024	0.4681	0
CD player	0.0884	0.5108	0.4224	0
Cassette	0.3339	0.5476	0.2137	0.003
Camera	0.0903	0.225	0.1347	0.006
Camera (digital)	0.0324	0.1094	0.077	0.036
Video camera	0.0131	0.108	0.0949	0.002
Telephone/Mobile Phone	0.4429	0.8461	0.4032	0

Table B.6: Ownership of durable assets and housing characteristics by location

Variable description	Full Sample	Urban Malawi	Rural Malawi
Newspaper	0.2695	0.4628	0.2091
Magazine	0.1933	0.2685	0.1699
Clock	0.6754	0.8014	0.6361
Piped water	0.2904	0.4696	0.2345
Bore hole	0.4758	0.3214	0.524
Table	0.5948	0.6854	0.5664
Bed	0.5905	0.74	0.5438
Private study area	0.2196	0.2286	0.2167
Bicycle	0.5975	0.5439	0.6143
Horse cart	0.0576	0.0451	0.0616
Car	0.0484	0.0792	0.0388
Motorcycle	0.0403	0.0501	0.0373
Tractor	0.0141	0.0152	0.0137
Electricity	0.1407	0.3436	0.0773
Refrigerator	0.0544	0.1541	0.0232
Air conditioner	0.0086	0.0248	0.0036
Electric fan	0.0535	0.1743	0.0157
Washing machine	0.0043	0.0104	0.0024
Vacuum	0.0051	0.0145	0.0022
Computer	0.0077	0.013	0.006
Internet	0.0048	0.0087	0.0036
Radio	0.876	0.9286	0.8596
TV	0.1764	0.3418	0.1247
VCR	0.0742	0.1712	0.0439
DVD player	0.0778	0.1847	0.0444
CD player	0.0972	0.169	0.0748
Cassette	0.2782	0.3318	0.2614
Camera	0.0897	0.1556	0.0691
Camera (digital)	0.0313	0.0579	0.023
Video camera	0.0197	0.0418	0.0128
Telephone/Mobile Phone	0.3651	0.598	0.2923
Type of floor material			
Earth/clay	0.5434	0.2638	0.6308
Canvas	0.0124	0.0085	0.0136
Wooden planks	0.0162	0.0094	0.0184
Cement	0.4164	0.6991	0.328
Carpet/Tiles	0.0116	0.0192	0.0092
Type of wall material			
Cardboard	0.0192	0.0136	0.021
Reeds/sticks/grass	0.087	0.0396	0.1018
Stones	0.4283	0.4701	0.4153
Metal sheets/Asbestos	0.0327	0.0424	0.0297
Wood	0.0208	0.019	0.0214
Cut stone	0.4119	0.4153	0.4109
Type of roof material			
Cardboard	0.0199	0.0224	0.0191
Grass thatch/mud	0.503	0.2102	0.5945
Metal/Asbestos	0.4436	0.706	0.3616
Cement	0.0205	0.0392	0.0147
Tiles	0.013	0.0223	0.0101

APPENDIX C

Chapter 4 Appendix

C.1 Levels and changes in employment shares by race and gender, 1997-2015

Table C.1: Levels and changes in employment shares by race and gender, 1997-2015

Occupations ranked by mean earnings	ISCO-88 code	Sub-group	Percent in 1997	Perecent in 2015	Percent point change
Legislators and senior officials	11	Black	0.3	0.06	-0.3
		Coloured	0.2	0.09	-0.1
		Indian	0.5	0.05	-0.5
		White	1.3	0.53	-0.8
		Male	0.4	0.17	-0.2
		Female	0.8	0.12	-0.7
Corporate managers	12	Black	0.6	3.66	3.1
		Coloured	1	6.31	5.3
		Indian	3.7	17.7	14.0
		White	7.4	24.32	16.9
		Male	2.4	9.36	7.0
		Female	1.8	7.08	5.3
General managers	13	Black	3.6	1.78	-1.8
		Coloured	4.5	1.7	-2.8
		Indian	12.4	6.81	-5.6
		White	13.4	7.26	-6.2
		Male	6.7	3.83	-2.9
		Female	5.03	1.72	-3.3
Physical, mathematical, and engineering professionals	21	Black	0.7	0.37	-0.4
		Coloured	0.6	0.14	-0.5
		Indian	0.6	1.96	1.4
		White	5.0	3.65	-1.3
		Male	2.1	1.36	-0.7
		Female	0.4	0.49	0.1
Life science and health professionals	22	Black	0.1	0.05	-0.1
		Coloured	0.0	0.05	0.0
		Indian	0.3	0.14	-0.1
		White	0.5	0.98	0.5
		Male	0.2	0.21	0.1
		Female	0.3	0.26	0.0
Teaching professional	23	Black	0.2	0.03	-0.2
		Coloured	0.2		-0.2
		Indian	0.1		-0.1
		White	1.0	0.17	-0.8
		Male	0.32	0.04	-0.3
		Female	0.5	0.08	-0.4

Occupations ranked by mean earnings	ISCO-88 code	Sub-group	Percent in 1997	Perecent in 2015	Percent point change
Other professionals	24	Black	1.0	1.5	0.5
		Coloured	1.3	1.6	0.3
		Indian	4.5	6.7	2.2
		White	6.3	6.6	0.3
		Male	2.2	2.3	0.0
		Female	2.7	3.3	0.6
Natural and engineering science associate professionals	31	Black	1.3	2.13	0.8
		Coloured	3.1	2.73	-0.3
		Indian	3.1	6.36	3.3
		White	5.2	5.64	0.5
		Male	2.5	3.8	1.3
		Female	2.4	1.9	-0.6
Life science and health associate professionals	32	Black	0.2	0.2	0.0
		Coloured	0.2	0.3	0.1
		Indian	0.1	0.4	0.3
		White	0.3	0.4	0.2
		Male	0.1	0.2	0.1
		Female	0.5	0.5	0.0
Teaching associate professionals	33	Black	0.1	0.1	0.0
		Coloured	0.2	0.08	-0.1
		Indian	0.2		-0.2
		White	0.1	0.1	0.0
		Male	0.1	0.01	-0.1
		Female	0.22	0.2	0.0
Other associate professionals	34	Black	2.6	3.2	0.6
		Coloured	2.6	4.0	1.3
		Indian	5.4	7.0	1.6
		White	9.9	9.6	-0.3
		Male	3.5	3.7	0.2
		Female	6.2	6.1	-0.1
Office clerks	41	Black	4.1	5.7	1.6
		Coloured	7.3	9.9	2.6
		Indian	14.7	18.3	3.6
		White	13.5	14.1	0.6
		Male	4.2	4.2	0.0
		Female	14.3	14.8	0.5
Customer services clerks	42	Black	3.9	6.4	2.5
		Coloured	4.2	7.0	2.8
		Indian	7.9	6.5	-1.4
		White	5.3	3.3	-2.0
		Male	1.6	1.9	0.2
		Female	11.2	12.1	0.9
Extraction and building trades workers	71	Black	10.2	6.2	-4.0
		Coloured	6.7	4.7	-2.1
		Indian	3.3	2.5	-0.8
		White	4.1	4.8	0.7
		Male	10.8	8.7	-2.1
		Female	1.4	0.9	-0.5

Occupations ranked by mean earnings	ISCO-88 code	Sub-group	Percent in 1997	Perecent in 2015	Percent point change
Metal, machinery, and related trade work	72	Black	6.1	4.5	-1.6
		Coloured	4.4	5.1	0.7
		Indian	4.3	5.2	0.8
		White	7.3	6.8	-0.5
		Male	8.1	8.0	-0.1
		Female	0.9	0.5	-0.4
Precision, handicraft, craft printing, and related trade workers	73	Black	0.9	0.5	-0.4
		Coloured	1.1	0.8	-0.3
		Indian	1.7	0.7	-1.0
		White	0.7	0.6	-0.1
		Male	0.9	0.6	-0.3
		Female	0.8	0.4	-0.4
Other craft and related trade workers	74	Black	4.0	2.2	-1.8
		Coloured	5.4	1.6	-3.9
		Indian	6.5	0.5	-6.0
		White	0.9	0.4	-0.5
		Male	2.2	1.4	-0.8
		Female	7.3	2.1	-5.2
Stationary plant and related operators	81	Black	1.7	2.2	0.4
		Coloured	0.6	1.0	0.3
		Indian	0.7	0.3	-0.5
		White	0.8	0.6	-0.2
		Male	1.7	2.4	0.7
		Female	0.4	0.4	0.0
Machine operators and assemblers	82	Black	4.8	4.1	-0.7
		Coloured	5.8	4.8	-1.0
		Indian	4.5	3.3	-1.2
		White	0.9	0.6	-0.3
		Male	4.0	3.6	-0.4
		Female	4.3	3.3	-1.0
Drivers and mobile plant operators	83	Black	13.9	7.5	-6.3
		Coloured	6.7	5.7	-1.0
		Indian	4.2	3.4	-0.8
		White	3.2	1.0	-2.2
		Male	13.7	9.4	-4.4
		Female	1.1	0.6	-0.5
Personal and protective service workers	51	Black	7.4	13.2	5.9
		Coloured	4.1	6.9	2.8
		Indian	1.3	2.3	1.1
		White	3.8	3.0	-0.8
		Male	5.8	9.2	3.4
		Female	5.9	11.4	5.5

Occupations ranked by mean earnings	ISCO-88 code	Sub-group	Percent in 1997	Perecent in 2015	Percent point change
Models, salaespersons and demonstrators	52	Black	4.6	4.7	0.0
		Coloured	4.9	5.1	0.2
		Indian	11.5	6.7	-4.8
		White	5.2	2.9	-2.4
		Male	3.8	4.3	0.5
		Female	8.4	4.8	-3.7
Skilled agricultural and fishery workers	61	Black	2.9	0.6	-2.4
		Coloured	2.3	0.7	-1.7
		Indian	0.2	0.1	-0.1
		White	0.5	0.2	-0.3
		Male	2.3	0.5	-1.8
		Female	2.04	0.5	-1.5
Subsistence agricultural and fishery workers	62	Black	0.05	0	-0.1
		Coloured	0	0.02	0.0
		Indian	0	0	0.0
		White	0	0	0.0
		Male	0.04	0	0.0
		Female	0	0.01	0.0
Sales and service elementary occupations	91	Black	5.0	10.4	5.4
		Coloured	3.0	7.7	4.7
		Indian	1.5	1.4	-0.1
		White	1.0	0.8	-0.2
		Male	3.1	4.7	1.7
		Female	5.1	12.7	7.6
Agricultural, fishery and related labourers	92	Black	5.0	6.5	1.6
		Coloured	15.4	12.5	-2.8
		Indian	0.2	0.2	-0.1
		White	0.2	0.2	0.0
		Male	5.0	5.7	0.7
		Female	6.01	6	0.0
Labourers in mining, construction, manufacturing, and transport	93	Black	14.9	12.4	-2.5
		Coloured	14.1	9.6	-4.4
		Indian	6.92	1.9	-5.1
		White	2.3	1.5	-0.9
		Male	12.5	10.6	-1.9
		Female	9.9	7.9	-1.9

C.2 Changes in Occupation Employment Shares

Table C.2: Full sample vs private sector sample

	1997		2015		Difference	
	Private	**Full**	**Private**	**Full**	**Private**	**Full**
Low	33.9	31.6	38.2	36.2	4.4	4.6
Middle	45.6	39.3	38.1	33.9	-7.5	-5.4
High	20.6	29.1	23.7	29.9	3.1	0.8

Table C.3: Private sector changes in occupation employment Shares by gender

	1997		2015		Diference	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**
Low	32.5	37.3	35.0	43.3	2.5	6.0
Middle	47.1	41.8	40.1	35.0	-7.0	-6.8
High	20.4	21.0	24.9	21.7	4.5	0.8

Table C.4: Private sector changes in occupational employment shares by race

	1997				2015				Difference			
	Black	**Coloured**	**Indian**	**White**	**Black**	**Coloured**	**Indian**	**White**	**Black**	**Coloured**	**Indian**	**White**
Low	39.8	43.9	21.6	13.1	47.7	42.6	12.6	8.5	7.9	-1.3	-9.0	-4.6
Middle	49.4	42.3	47.7	36.7	39.2	40.4	40.5	32.3	-10.2	-1.9	-7.3	-4.4
High	10.8	13.9	30.7	50.3	13.1	17.0	47.0	59.3	2.3	3.1	16.3	9.0

Table C.5: Between and within industry decompositon of the increase in the share of workers in employment, 1997-2015

	Low		Middle		High	
	Between	Within	Between	Within	Between	Within
Agriculture, hunting, forestry and fishing	-2.32	0.62	-0.56	-0.54	-0.14	-0.08
Mining and quarrying	-0.71	0.43	-1.70	-0.08	-0.42	-0.35
Manufacturing	-1.75	-0.86	-4.23	0.73	-1.46	0.13
Utilities	-0.08	-0.15	-0.27	0.05	-0.10	0.11
Construction	0.49	0.47	1.22	-0.84	0.28	0.36
Trade	0.45	0.29	0.37	-0.43	0.20	0.14
Transport	-0.14	-0.10	-0.47	-0.42	-0.20	0.52
Finance	2.47	1.04	2.50	-1.37	3.11	0.33
Services	1.09	3.96	0.59	-0.35	1.78	-3.61