Understanding Poverty and Inequality in Mozambique: The Role of Education and Labour Market Status

by

Carlos Chadreque Penicela da Maia

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

Promoter: Professor Servaas van der Berg

Faculty of Economic and Management Sciences

Department of Economics

December 2012

Declaration

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

December 2012

_____

(Carlos da Maia)

Abstract

This thesis deals with poverty and inequality in Mozambique and with the link of education to well-being through the labour market. Earlier studies that analysed well-being in Mozambique drew counter-intuitive conclusions about the spatial distribution of poverty and inequality. They focused excessively on money-metric indicators of well-being and adjusted the poverty line so as to make it reflect taste and price differentials across regions. This thesis suggests the use of a wealth index based on asset holdings and derived by employing Multiple Correspondence Analysis to support the money-metric results. If results are not also confirmed by other indicators of well-being, one should be sceptical of simply unquestioningly applying best practice approaches. In this thesis the money-metric results drawn by earlier studies are not confirmed by this other indicator of well-being.

Since education is a policy lever that can be used to influence the existing patterns of poverty and inequality, one needs to understand how it operates through the labour market in improving well-being. Developing and poor economies such as Mozambique are characterised by a very segmented labour market and by a small wage sector. A large proportion of the working-age population is engaged in subsistence agriculture and self-employment activities. Using a multinomial logit model this thesis demonstrates that schooling has an influence on the choice of employment segment. For instance, schooling increases an individual's chances of getting a public sector job, but lowers his or her chances of falling into self-employment activities. This study also links schooling to earnings. It argues that when analysing the relationship between schooling and earnings in a poor developing economy one should account for the multiple segmentation of the labour market as well as for sample selection bias. To estimate the effects of schooling on earnings this thesis thus employs a modified version of Dubin and McFadden's model. It finds a positive association between education and earnings in the public wage sector, the private wage sector and in the self-employment segment. Convex returns to education are also found, and accounting for selectivity bias does improve the earnings functions relative to those based on ordinary least squares regressions.

Education quality has a bearing on an individual's performance in the labour market and therefore affects the role of education in alleviating poverty. Thus, this thesis identifies the correlates of education quality in Mozambique. Employing education production functions based on ordinary least squares multivariate regressions it finds that most of the correlates of educational achievement suggested by the literature are indeed associated with educational outputs. Employing Oaxaca-

Blinder decomposition techniques often used in labour studies to study earnings discrimination, the thesis attempts to explain the reasons behind the average deterioration in education quality in Mozambique. The initial hypothesis on this matter was that the average deterioration in education quality over time was associated with the increase in the proportion of pupils from low socio-economic backgrounds. This hypothesis, however, is not confirmed. Likely explanations include the decline in the efficiency of the education system and more lenient pupil promotion policies.

Opsomming

Hierdie proefskrif handel oor armoede en ongelykheid in Mosambiek en die effek van onderwys op welsyn deur die arbeidsmark. Vroeëre studies het gevolgtrekkings oor welsyn in Mosambiek getrek wat nie met intuïsie oor die ruimtelike verdeling van armoede en ongelykheid strook nie. Sulke studies het slegs geldelike maatstawwe van welsyn gebruik en die armoedelyn aangepas om pryse en smaak in verskillende streke te reflekteer. Hierdie tesis stel die gebruik van nie-geldelike maatstawwe voor om geldelike maatstawwe aan te vul, en spesifiek ’n bate-indeks van welsyn wat verkry word deur die gebruik van Veelvuldige-Korrespondensie-Analise. Daar is rede tot skeptisisme wanneer die sogenaamde ‘beste’ metode vir die berekening van geldelike maatstawwe sonder bevraagtekening gebruik word en die resultate nie deur ander indikatore van welsyn bevestig word nie.

Aangesien onderwys ’n beleidsinstrument bied om bestaande patrone van armoede en ongelykheid te beïnvloed, is dit nodig om te verstaan hoe dit deur die arbeidsmark werk om welsyn te verbeter. Ontwikkelende en arm ekonomieë soos Mosambiek word gekenmerk deur ’n baie gesegmenteerde arbeidsmark en ’n klein loonsektor. ’n Groot persentasie van die bevolking van werkende ouderdom is by onderhoudslandbou en ander self-indiensneming betrokke. ’n Veelvoudige-logit-model toon hoe opvoeding die keuse van indiensnemingsektor beïnvloed. Onderwys verhoog byvoorbeeld iemand se kanse om ’n pos in die openbare sektor te kry, maar verlaag die waarskynlikheid van self-indiensneming. Die studie koppel verdienste ook aan onderwys. Daar word aangevoer dat die groot arbeidsmarksegmentasie en seleksie-sydigheid in berekening gebring moet word wanneer die verband tussen onderwys en lone in ’n arm ontwikkelende land bestudeer word. Dus word ’n aangepaste vorm van Dubin en McFadden se model in hierdie proefskrif gebruik om die effek van onderwys op verdienste te bereken. ’n Positiewe verband bestaan tussen onderwys en lone in die openbare loonsektor, die private loonsektor en self-indiensname. Die opbrengsstruktuur op onderwys is konveks, en inagneming van seleksie-sydigheid verbeter die verdienstefunksies relatief tot gevalle wat net op gewone kleinste-kwadrate-regressies gebaseer is.

Onderwysgehalte het ’n invloed op ’n persoon se vertoning in die arbeidsmark en raak daarom die rol van onderwys in armoedeverligting. Faktore wat met onderwysgehalte in Mosambiek verband hou word dus geïdentifiseer. Die gebruik van gewone-kleinste-kwadrate-veelvoudige-regressies in onderwysproduksiefunksies toon dat die meeste van die bepalende faktore wat in die literatuur genoem word inderdaad met onderwysuitsette verband hou. Deur gebruik van Oaxaca-Blinder

dekomposisie-tegnieke – wat meer dikwels gebruik word om arbeidsmarkdiskriminasie te ontleed – word gepoog om die redes vir die agteruitgang van gemiddelde onderwysgehalte in Mosambiek te verklaar. Die aanvanklike hipotese hieroor was dat die agteruitgang in die gemiddelde vertoning deur die toename van leerlinge van laer sosio-ekonomiese agtergrond verklaar sou kon word. ’n Ontleding van die data bevestig egter nie hierdie hipotese nie. Moontlike verklarings sluit in ’n agteruitgang in die doeltreffendheid van die onderwysstelsel en minder streng beleid rakende promosie van leerlinge na hoër grade.

Acknowledgements

Working on this thesis was a challenge, and in many occasions I doubted of my capabilities of completing it. I spent many hours looking at the computer screen trying to find solutions to the hurdles that this thesis posed every now and then. Once all my ideas were exhausted, I resorted in many occasions to *Professor Van der Berg*, who always did a brilliant job, guiding me and putting me back on track. For that and also for the opportunity he gave me to pursue my Doctoral studies, I would like to thank him from the very bottom of my heart.

Many were the weekdays, the weekends and the holidays that I spent working on this thesis. Also many were the weekdays, the weekends and the holidays that I had to spend away from *Ivenilde*, my wife. She was very understanding and always supported me. Her words of encouragement and her delicious food gave me strength to keep walking, to keep going forwardly. Many thanks for that *Ivenilde*.

I also would like to acknowledge the support of my parents. From the very beginning they encouraged me with my studies. My family and relatives, *Ivo*, uncle *Jorge*, uncle *Mário Jorge* and many others, also supported me. I thank them all.

I would like to thank ReSEP, the research team led by *Professor Van der Berg*, for all the comments and suggestions made in my presentations. In particular, I would like to thank *Cobus Burger*, *Derek Yu* and *Dieter von Fintel* for their personal guidance, and *Paula Armstrong* for helping with my English insufficiencies.

The *World Bank*, the Mozambican *Instituto Nacional de Estatística* and *Finório Castigo* provided me with data without which this thesis would not be undertaken. I really appreciate that.

To all my friends that encouraged me to face this challenge, *Odília Massangaie*, *Eusébio Pequenino*, *Aurélio Bucuane*, *José Guirrugo, Sebastião Matsinhe*, *Norman Mafumo*, *Policarpo Matiquite*, the late *Zaqueo Sande* and all those that one way or another gave me strength to keep going, *Kanimambo.*

Lastly but not least, I would like to thank *God* for all the blessings given throughout this mountainous journey.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 What this Thesis is About

This thesis deals with poverty and inequality in Mozambique, a very poor country, and with the link of education to welfare through the labour market. The literature on the analysis of poverty and inequality is very limited and earlier studies that analysed these phenomena in Mozambique drew counter-intuitive conclusions about the spatial distribution of well-being. This has implications for policy targeting and is detrimental to the poor. This thesis therefore firstly addresses this issue.

Education is one of the policy levers that can be used to influence the existing patterns of poverty and inequality in Mozambique and one thus needs to understand how it operates through the labour market in improving well-being. But the global agenda of Education For All, which promoted fast increases in access to education, could have had an influence on education quality in Mozambique. The fall in education quality is detrimental to the role of education in alleviating poverty and inequality, and again, has implications for the poor. This thesis also deals with these aspects.

Before addressing the research question in more detail, it is useful to look at the background information on the Mozambican economy.

1.2 Background Information

Mozambique is situated in the South-East coast of Africa and shares borders with Tanzania, Malawi, Zambia, Zimbabwe, South Africa and Swaziland. It has eleven administrative regions called provinces, distributed across three main geographical regions. The provinces are shown in Figure 1.1 on page 15. Niassa, Cabo Delgado, and Nampula are part of the Northern region, Zambezia, Tete, Manica, and Sofala are located in the Central part of Mozambique, and Inhambane, Gaza, Maputo Province and Maputo City are part of the Southern region. All provinces except Maputo City have urban and rural locations. Maputo City is all urban and is the political, economic and cultural capital of Mozambique.

According to the latest population census (Censo 2007), Mozambique has about 20.0 million inhabitants. Nampula and Zambezia are the most populated provinces, with a total population of about 4.0 million and 3.8 million, respectively. Maputo City is the smallest province in size, is one of the least populated provinces in absolute terms (about 1.0 million inhabitants), but has the highest population density.

Figure 1.1 – Political Map of Mozambique

Mozambique has been a Portuguese colony from the 16th century until it gained political independence in 1975, following 10 years of armed conflict. After the country's independence the ruling government – FRELIMO – followed the socialist ideology of a centrally planned economy. In opposition to this, RENAMO, a military rebel group backed by the then Rhodesia and later by (apartheid) South Africa, initiated a series of sabotage attacks on Mozambican economic infrastructures. By 1976-1977 sabotage had escalated into a civil war between the two forces. By 1988 approximately 2.0 million Mozambicans had fled their homes and another 200,000 had died as a consequence of the war (MacFarquhar, 1988). The civil war only came to an end when a peace treaty between the belligerent forces was signed in 1992. By then 58% of the schools existing in 1983 had been destroyed or closed as an outcome of the war (Ministry of Education, 1996, p. 40), about 1.0 million people had died, at least 3.0 million people had relocated from their places of origin and roughly 1.7 million had migrated to neighbouring countries (Hanlon, 1997, p. 14).

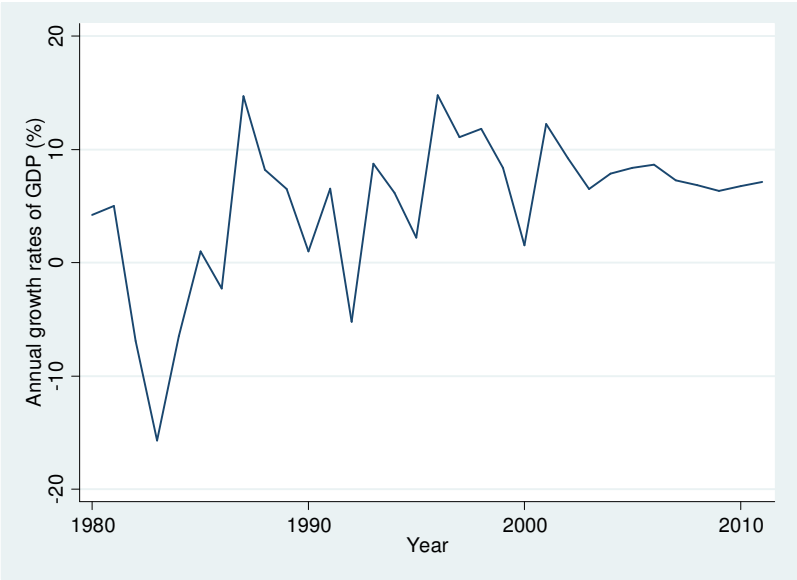
There is the possibility that the large-scale migration to neighbouring countries caused by the civil war have resulted in a 'brain drain', with highly skilled labour being lost to the labour market. Such possibility however, is not likely to have occurred. The reason is that the civil war was more intense

in the rural parts of the Central and Northern provinces of Mozambique, with most of the population from these regions being forced to migrate. As will be seen throughout this thesis these regions historically present the highest proportions of people with low or without any school education.

In addition to the civil war which affected this economy, in the beginning of the 1980s Mozambique also suffered from natural disasters and a profound economic crisis. Between 1981 and 1984 the country experienced a massive drought which resulted in about 100,000 human deaths (PreventionWeb, 2012). The drought compounded the negative effects of the war, thus crippling the rural economy. For the first time Mozambique became a net importer of food and a recipient of food donations. The period was characterised by growing imports which depleted its international reserves. The strong subsidies to the social sectors of education and health which characterised the centrally planned economy resulted in a huge budget deficit. To finance this deficit the Mozambican authorities resorted to domestic and external loans. Around 1984 Mozambique faced difficulties in repaying its external debts and consequently lost credibility in the international financial markets (PNUD, 1998, p. 51). Around the same time the Mozambican authorities started questioning the feasibility of continuing with the centrally planned economy, and thus initiated contacts with the Bretton Woods institutions, the World Bank and the International Monetary Fund.

Most of the 1980s were characterised by very small and in some cases negative economic growth rates. For instance, between 1982 and 1983 the Mozambican economy, measured by its gross domestic product (GDP), shrank by roughly 16%. Since the 1990s the performance of the economy improved significantly. Between 1993 and 2011 the economy grew by an average of 8% per year, which constitutes very strong growth (see Figure 1.2 on page 17).

Figure 1.2 – Growth rates of the Mozambican economy



Source: adapted using data from International Monetary Fund, World Economic Outlook Database, April 2012

In 1992, the year the peace treaty was signed, Mozambique's rank on the Human Development Index (HDI) was very low. It ranked 146th out of 160 countries. While Mozambique has experienced strong growth rates of GDP per capita in the last two decades, its HDI did not improve over time. As indicated by the latest human development report (Human Development Report, 2011), out of 187 countries Mozambique ranked in the 184th place. Due to methodological differences the HDI values and ranks published in the latest report are not comparable to those published in earlier reports. Despite this fact, irrespective of the method applied, Mozambique always ranks very low on human development and is therefore one of the world's poorest countries. Being one of the poorest countries in the world, Mozambique lacks the fiscal resources for asset accumulation and service delivery. Due to limited capacity in terms of internal resources the Mozambican authorities, as mentioned, had to resort to external financing, particularly in terms of loans and donations. By 1980 Mozambique's external debt amounted to 49.0 million US$. By 1997, however, it had risen to 5,877.6 million US$, more than 120 times the 1980 figure (Francisco, 2002, p. 33). To counteract this situation of extreme poverty and huge external debts, in 1998 Mozambique was declared eligible to benefit from the Heavily Indebted Poor Countries (HIPC) initiative. This initiative constituted the international community's response to alleviate budget pressures for the world's poorest and most indebted countries. Mozambique thus became eligible to receive comprehensive debt relief from its multilateral and bilateral partners.

A number of programmes have been put in place to help alleviate the high poverty levels in Mozambique. In 1987 the government introduced the PRE (*Programa de Reabilitação Económica*,

Economic Rehabilitation Programme). Its ultimate objective was to improve the living standards of the Mozambican population. It was a structural plan which basically marked the change from a centrally planned economy to a free market economy. It involved the deregulation of the economy, the devaluation of the national currency (the *Metical*), the privatisation of state-owned enterprises, the liberalisation of prices, the reduction of the budget for the social sectors, and the reduction of public service employment and real wages, among other things. Mole (1994, pp. 158-159) argued that coupled with the effects of the war and the natural disasters the programme did not succeed in improving the living standards of the Mozambican population, but exacerbated poverty and inequality.

Acknowledging these deficiencies, the donor community and the Mozambican authorities decided to add a social component to the programme in 1990. The programme was re-baptised with the name PRES (*Programa de Reabilitação Económica e Social*, Economic and Social Rehabilitation Programme). It basically involved the introduction of a set of safety nets aimed to improve the situation of the poorest. A few elements of the safety nets included the distribution of free food, medicines, seeds, fertilisers and agricultural tools, among other things (Bata, 2006, p. 17).

In 2001 the government approved the implementation of PARPA 2001-2005 (*Plano de Acção para a Redução da Pobreza Absoluta*, the first Poverty Reduction Strategy Paper – PRSP). The central objective of PARPA was the substantial reduction of absolute poverty by improving the capabilities and opportunities of all Mozambicans, in particular the poor. The specific objective was to cut down the poverty incidence of roughly 70% in 1996-1997 to a figure below 60% by 2005, and to below 50% by 2010 (República de Moçambique, 2001, p. 1). PARPA II was the version of PARPA for the period 2006-2009, and was introduced in 2006. It aimed to reduce the incidence of poverty from roughly 54% in 2003 to 45% by 2009 (República de Moçambique, 2006, p. 1). It will be seen throughout the next chapter that the official government reports claimed that poverty declined significantly between 1996-1997 and 2002-2003. It will also be seen that such claims, however, are very sensitive to the methodologies employed to define the poor.

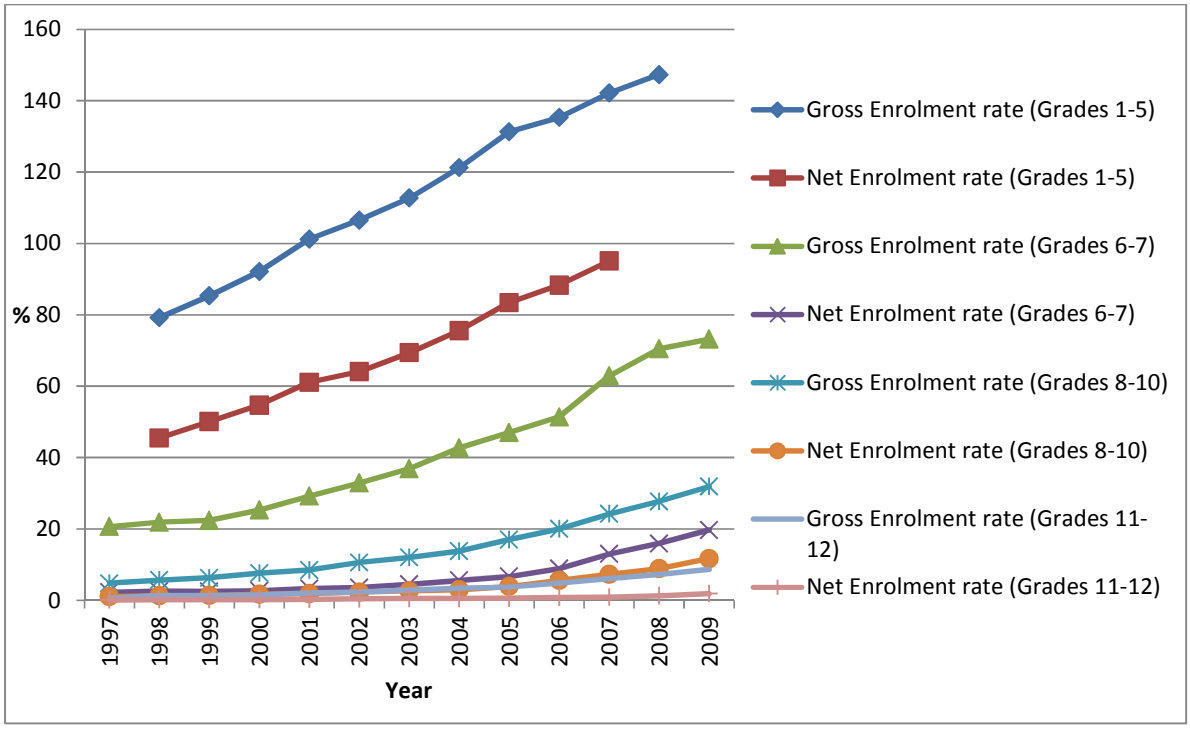
In both the first PARPA and PARPA II education is one of the main pillars of the poverty reduction strategy. Education is seen as a means for improving human development, by increasing the capabilities and opportunities of the poor, and by promoting social, regional and gender equity. However, it is important to understand the link between education and the mentioned outcomes. How does education relate to human development? How does education increase the capabilities and opportunities of the poor? How does education increase equity? What are the prospects for

poverty reduction if some socio-economic groups are left behind in access? This thesis also deals with these questions in the chapters that follow.

The main education objectives of PARPA were the achievement of universal primary education, the expansion of secondary education, non-formal education, and technical-professional education, with most of the resources employed in primary education. An improvement in quality and efficiency as well as the reduction of costs also constituted education objectives of PARPA (República de Moçambique, 2001, pp. 3-4, 44). In PARPA II the government put more emphasis on education quality. The main education objective of PARPA II was to ensure quality education for all, in particular at the primary level of education (República de Moçambique, 2006, p. 100).

In 1990 governments of a number of countries met in Thailand at the World Conference on Education for All. They recognised the existence of extreme disparities in access to education in their respective countries. Yet, it is known that people lacking numeracy and literacy skills face heightened risks of falling into poverty. Therefore, the governments declared the adoption of consistent measures to overcome the inequalities in access to education (UNESCO, 2010, p. 136), and set out six important goals to be met within specified time periods. One of these important goals was to ensure that "by 2015 all children, particularly girls, children in difficult circumstances and those belonging to ethnic minorities, have access to and complete free and compulsory primary education of good quality" (UNESCO, 2000, p. 15). This has had an effect on the Mozambican education system. The global agenda of Education For All (EFA) promoted fast increases in enrolments in the Mozambican education system, in particular at the primary level. As is illustrated in Figure 1.3 on page 20, gross and net enrolment rates at the lower primary (grades 1 to 5) and upper primary (grades 6 and 7) levels of education increased sharply in the first decade of the 21st century.

Figure 1.3 – Gross and Net Enrolment Rates between 1997 and 2009



Source: adapted using data from Mozambique Education Statistics 1998 – 2009

These fast increases in enrolments could have had an influence on education quality. It seems that efforts aimed at rapidly increasing enrolments, in particular at the primary level of education, had detrimental effects on quality. This thesis also investigates the quantity versus quality of education trade-off in Mozambique, in Chapter 4.

1.3 Problem Statement and Research Questions

In order to inform policy analysis and design programmes aimed to curb poverty and inequality the state requires a good understanding of these phenomena. A deficiency of the existing literature, however, is that it focuses too much attention on the money-metric analysis of poverty and inequality. The poor is defined as everyone who within a certain period of time fails to achieve a minimum income or consumption threshold called the poverty line. Inequality then measures the distribution of the money-metric indicator (i.e., income or consumption expenditures) across the population.

The design and implementation of PARPA and PARPA II were informed by earlier studies based on the existing household surveys at the time, IAF[1] 1996-1997 and IAF 2002-2003, respectively. Such studies are the official government reports on poverty and well-being and include Ministério

[1] IAF stands for *Inquérito aos Agregados Familiares* in Portuguese, household survey in English.

do Plano e Finanças (1998) and another study done by a group of researchers from Mozambique's National Directorate of Planning and Budget, Ministry of Planning and Finance, Economic Research Bureau, International Food Policy Research Institute and Purdue University in 2004 (hereafter DNPO et al., 2004). Both studies employed consumption expenditures per capita as the main living standards indicator of the population.

Poverty, however, is a multidimensional concept. Poverty is also defined as the absence of the freedom to choose arising from a lack of capability to function effectively in society (Sen, 2001). As argued by Van der Berg (2008, p. 1), "this multidimensional interpretation moves far beyond the notion of poverty as being solely related to a lack of financial resources". Poverty is also defined as lack of education, deficient health, and weak access to basic goods and services such as water, electricity, shelter, sanitation and communications, among other things. These alternative definitions of poverty were recognised in PARPA and PARPA II, but the main objectives of these programmes were defined in terms of the money-metric approach, as mentioned above.

Another deficiency existing in the literature relates to the estimation of poverty across regions. In using survey data for money-metric analysis of poverty and well-being across regions, it is customary to adjust the poverty line to reflect local tastes and prices. This approach is known as the cost of basic needs method. In developing and poor countries such as Mozambique where recorded price differentials between regions or provinces are large, using the remedy of adjusting for price differentials may sometimes lead to very wrong conclusions about the spatial distribution of poverty. This may have severe consequences for policy and may be detrimental to the poor.

In creating the poverty profile for Mozambique for the periods 1996-1997 and 2002-2003 both government reports (i.e., Ministério do Plano e Finanças, 1998 and DNPO et al., 2004, respectively) followed the cost of basic needs approach to derive region-specific poverty lines, which throughout this thesis are referred to as variable poverty lines. These poverty lines were intended to reflect as closely as possible the local tastes and prices. But in DNPO et al. (2004), for instance, the poverty line derived for Maputo City was very high, more than twice the poverty lines of most Central and Northern parts of the country and more than three times the poverty lines of the rural parts of Nampula, Sofala and Zambezia. It will be argued that this extremely high poverty line overstates poverty in Maputo City, with consequences for the targeting of policy.

The central research problem addressed in this thesis is that poverty and inequality in the Mozambican context are not well understood. The analysis is overly focused on money-metric indicators and the adjustments made to the poverty line as to make it reflect local tastes and prices

may bring more harm than good. Not addressing this problem might have consequences for the targeting of policy with detrimental effects to the poor. There are so far no studies that addressed this research problem in the particular case of Mozambique.

The research problem addressed in this thesis leads to other interesting questions that this work tries to answer. Having access to appropriate poverty and inequality profiles is only an initial step in the design of programmes to curb these phenomena. Policy levers are needed that can be used to influence the existing patterns of poverty and inequality. Education is one of these policy levers and we thus need to understand how it operates in improving well-being.

The association of education with well-being occurs through the labour market. Education is associated with access to better jobs, and conditional on this it is associated with better earnings. In developing and poor countries such as Mozambique, however, the labour market is very segmented. The Mozambican labour market, for instance, could be split into six distinct employment segments. These include wage employment in the private sector, wage employment in the public sector, self-employment, unpaid family work, unemployment and non-participation. Most people work outside the wage sectors, particularly in self-employment activities such as subsistence agriculture and informal activities. On this matter, this thesis tries to answer two research questions:

- What is the effect that schooling and other individual characteristics have on a person's chances of falling into each one of the employment states that characterise the Mozambican labour market?
- How does individual schooling relate to earnings in each of the employments states for which reliable earnings data are available?

The global agenda in education has contributed to fast increases in access to education in many developing and poor countries. In Mozambique, the number of pupils enrolled in grade 6 has increased from about 120,000 to roughly 320,000 in the first seven years of the 21st century. This very fast increase in enrolments is likely to have put pressure on the education system, with detrimental effects on education quality. One example that is explored in this thesis is the fall in the average performance of Mozambican pupils in standardised cross-country tests of cognitive achievement, signalling deterioration in average education quality of those pupils tested.

The fall in education quality is detrimental to the role of education in alleviating poverty and inequality. Most parents will invest in the education of their offspring only if it brings adequate returns. In turn education will bring adequate returns if it is education of quality. Employers will

employ and reward better those individuals coming from better quality schools, and this group of individuals will have better chances of escaping poverty compared to those coming from poor quality schools. If the deterioration in education quality affects some groups more than others, like was the case in Mozambique, inequality might increase. In Mozambique the fall in the quality of education seems to have affected those coming from lower socio-economic backgrounds more, in particular those in the Central and Northern provinces. Therefore, this thesis also addresses the following research questions:

- What are the correlates of pupil cognitive achievement in Mozambique?
- How can the average deterioration in education quality over time be explained?

1.4 Research Objectives and Theses Statements

The empirical work was largely divided into three chapters. In order to deal with this they are discussed separately in chapters 2, 3 and 4.

In **Chapter 2** the overall objective is to improve the understanding of poverty and inequality in the Mozambican context. This will be done by demonstrating that focusing only on the money-metric analysis of poverty and inequality gives a narrow understanding of the well-being of the Mozambican population. It will also be demonstrated that the practice of adjusting poverty lines to reflect local tastes and prices might lead to very wrong conclusions in terms of the spatial distribution of poverty and that this deficiency is detrimental for policy targeting. Thus, the main thesis defended in this chapter is that the provincial poverty profile estimated by DNPO et al. (2004) for the period 2002-2003 is inappropriate. By investigating this matter the spatial distribution of poverty and inequality in Mozambique will be better understood.

This thesis also deals with the relationship between education, the labour market and poverty alleviation, in **Chapter 3**. In this chapter the aim is to demonstrate that in developing and poor countries such as Mozambique, schooling has a crucial influence on the choice of employment sectors. For instance, it is shown that schooling increases an individual's chances of getting a public sector job, but lowers his or her chances of falling into self-employment activities. The thesis defended in this chapter is that in economies characterised by a very segmented labour market, as is the case in Mozambique, the effects of schooling on employment prospects should be analysed in each labour market segment, because they are likely to be different.

Another objective of this chapter is to demonstrate that schooling is associated with greater individual earnings and therefore has a potential to alleviate poverty and inequality. Empirical

estimates of returns to education in Mozambique are rare, despite their advantage for policy. A discussion paper on the topic was recently published in Mozambique's Ministry of Planning and Development. In the paper, Simione (2011) measures labour market returns to education in Mozambique. The paper, however, does not take into account the segmentation existing in the Mozambican labour market. Also, there is no account for the sample selection bias that originates from the omission of some groups from the analysis, which means that results cannot be generalised to the entire working-age population. Thus, this study aims to present the first empirical estimation of returns to education in Mozambique accounting for the existing multiple employment segments in the labour market as well as for sample selection bias.

The last group of research objectives relate to education quality and are discussed in **Chapter 4**. The main objectives are to identify the correlates of pupil educational achievement in Mozambique, and lastly to explain the reasons behind the fall in education quality over time. The main thesis associated with this chapter is that the average deterioration in education quality over time was caused by a massive influx of pupils from disadvantaged socio-economic backgrounds as well as the consequent pressure put on existing school resources. Poor pupils are likely to fare worse than their rich counterparts since they often have to walk long distances to school, have poor health and nutrition, lack learning materials and get limited help with school work at home. The increased number of pupils causes pressure on existing resources, given that the supply of school factors such as school infrastructures and teachers grows slower than school demand. The new school resources such as teachers are likely to be of lower quality themselves.

1.5 Significance and Limitations

The practical significance of this work is its direct contribution to a finer understanding of the well-being of the Mozambican population, by combining money-metric and other dimensions of poverty and inequality. By better understanding who the poor are and how unequal the various socio-economic groups are, policymakers, researchers and the like can improve the targeting of poverty and inequality alleviation programmes.

The general limitation of this thesis relates to the research design employed as dictated by data availability. The type of study that best addresses the formulated research problem is secondary analysis and statistical modelling of existing survey data. This type of study provides the required evidence to better understand poverty and inequality in Mozambique, as well as the role education and labour markets have. This type of research design allows savings in time and costs because of the use of existing data or the possibility of reanalysing previous findings. However, the analyst of secondary data is not able to control for data collection errors and is constrained by the fact that

government (i.e., the statistical office) collected information on some variables but not on others (Mouton, 2001, p. 165).

1.6 Chapter Overviews

This thesis is made of five chapters. **Chapter 2** makes the case that money-metric analysis of poverty and inequality, as well as adjusting poverty lines to reflect local tastes and prices in the manner an earlier work did, are unsuitable procedures to study living standards and lead to counter-intuitive conclusions regarding the spatial distribution of well-being in Mozambique. The first section presents the background to the chapter, including, for instance, the huge decline in poverty between the periods 1996-1997 and 2002-2003 reported by previous studies. It is shown how some scholars dispute the extent of the mentioned decline in the incidence of poverty.

Section 2.2 reviews the literature on poverty and inequality, and is divided into four subsections. First, it discusses the theoretical considerations behind the estimation of money-metric poverty, that is, the choice between using income or consumption as the living standards indicator, the choice between using absolute versus relative poverty lines and the choice of a poverty measure to report the results, in particular the Foster, Greer & Thorbecke (1984) class of decomposable poverty measures. This subsection also discusses the most commonly used inequality measures in the literature, the Gini and the Theil indices. Then, the next subsection considers studies that employed the money-metric approach to analyse poverty and inequality in Mozambique. One of the reviewed studies is the government report produced by DNPO et al. (2004). In particular it is shown how the cost of basic needs approach was used to derive a different poverty line for each geographical region and the implications of that. The third subsection reviews studies that employed an asset index to analyse poverty in Africa, an approach which is followed later in the chapter to understand well-being better in Mozambique. The last subsection of Section 2.2 summarises the literature review, pointing out the main deficiencies of the existing literature on the matter discussed in this chapter.

Section 2.3 describes the data as `well as the methods used to analyse poverty and inequality. The data are sourced from the second Mozambican income and expenditure household survey. The main method applied is the derivation of an asset index employing a technique named Multiple Correspondence Analysis (MCA). The fourth section reports the poverty estimates across various groups. It also includes a subsection which estimates the correlates of poverty using multivariate ordinary least squares (OLS) regressions. Section 2.4 also reports the Gini and Theil inequality estimates. Section 2.5 summarises and concludes the chapter.

Chapter 3 investigates the relationship between education, labour markets, and poverty alleviation. It studies the effects of schooling on the choice of employment segment as well as on earnings within each sector for which earnings data are available. It makes the case that the study of the relationship between education and earnings in countries with very segmented labour markets such as Mozambique should account for that segmentation. After a brief introduction and background to the chapter, Section 3.2 reviews studies that analysed the relationship between education and labour market outcomes in African countries with similar characteristics to Mozambique. Section 3.3 describes the data used in the chapter. The main data are sourced from *Inquérito Integrado à Força de Trabalho*, the only labour force survey in Mozambique. Section 3.4 analyses the relationship between education and employment in the presence of multiple employment segments. The main technique applied is a regression of employment segments on education (and other controls) based on a multinomial logit model. Section 3.5 of this chapter models the relationship between education and earnings in the public sector, the private sector and self-employment segment. First, simple OLS regressions are employed, thus not accounting for the existing multiple employment statuses nor for potential sample selection bias. Then, a version of Dubin & McFadden's (1984) model is used to correct for the deficiencies inherent in the OLS approach. The last section summarises and concludes the chapter.

As mentioned, this thesis is also concerned with the quality of education, as it has implications for poverty and inequality. Low quality education is less likely to contribute to poverty alleviation and if poorer socio-economic groups are more affected than their richer counterparts, a deterioration in the quality of education might lead to greater inequality. Thus, **Chapter 4** uses data sourced from the Southern and Eastern Africa Consortium for Monitoring Education Quality (SACMEQ), first to identify the correlates of education quality in Mozambique and then to understand the reasons behind the average deterioration in education quality over time.

The first section of this chapter situates the topic, in particular describing the huge expansion in enrolments in the Mozambican education system between the years 2000 and 2007, as well as the detrimental effects that that expansion seems to have had on education quality. Section 4.2 reviews the literature on school effectiveness, and lists the most important correlates of educational achievement found in the literature. To identify the correlates of educational achievement, OLS education production functions are employed. To separate the effects of increased enrolments from those of a change in education quality, this chapter employs the Oaxaca-Blinder decomposition technique (Oaxaca, 1973; Blinder, 1973). Section 4.3, thus, reviews the Oaxaca-Blinder decomposition technique and Section 4.4 discusses how this technique is adapted to the data used in

this chapter. Section 4.5 describes the SACMEQ data, sections 4.6 and 4.7 report the results, and Section 4.8 concludes the chapter.

Chapter 5 concludes this thesis. It discusses what was discovered throughout the thesis and its value. It is divided into four sections. The first section is a summary of findings. Then, Section 5.2 answers the thesis statements, i.e., it shows what can be deduced from the thesis as a whole. The third section presents the summary of contributions made throughout the thesis. It shows what this research has added in terms of new knowledge as well as discusses the implications of that. Section 5.4 presents suggestions for further research.

Chapter 2

Are Money-Metric Indicators of Poverty and Inequality Enough?

2.1 Chapter Introduction

Mozambique is one of the poorest countries in the world. The evidence on this is overwhelming. For instance, in 1992, the year in which a political settlement ended 16 years of a civil war, Mozambique had the 14th lowest Human Development Index (HDI) in the world, indicating that it was a very poor country. Between 1992 and 2011 this economy has experienced strong growth rates of Gross Domestic Product (GDP), averaging 8% per year (International Monetary Fund, 2012). Its HDI, however, continues to be one of the lowest in the world. As indicated by the latest Human Development report (Human Development Report, 2011), out of 187 countries Mozambique ranked in the 184th place.

This situation of extreme poverty has implications for the collection of fiscal resources. The tax base is very small and therefore the state has to resort to external finances in order to cover its budget deficit. In the current fiscal year, for instance, the budget deficit amounts to approximately 67.0 billion Meticais[2], which corresponds to 2.4 billion US$[3]. From this figure, 41.3 billion Meticais (1.5 billion US$) are supported by external finances, in the form of loans and donations. In other words, about 61.6% of the state budget deficit is financed by external resources.

The civil war, the extreme poverty, the lack of fiscal resources plus the resulting dependence on external financing limited asset accumulation and service delivery. As mentioned in the background section of Chapter 1, a number of programmes have been put in place in order to increase well-being in Mozambique. This requires a good understanding of poverty and inequality in order to inform policy analysis. But the literature on the analysis of poverty and inequality in Mozambique is limited and presents two main deficiencies: the analysis is only focused on money-metric indicators of living standards and the methods employed to derive poverty lines make them vary widely across regions.

It is often said, however, that focusing only on money-metric indicators does not give a complete picture of what is really happening to the population, i.e., it does not show accurately who the poor are (Kingdon & Knight, 2003, p. 3). The inclusion of non-monetary indicators in the analysis may shed some light on this aspect (Ravallion, 1996, pp. 1331-1332). Earlier studies using a money-

[2] Data collected from the Mozambican state budget for 2012.
[3] The latest Central Bank Economic Bulletin indicates that in June 2012 the exchange rate averaged 27.9 Meticais per US$ (Banco de Moçambique, 2012, p. 17).

metric indicator of well-being and variable poverty lines drew counter-intuitive conclusions about the extent of poverty reduction between the periods 1996-1997 and 2002-2003 as well as about the spatial distribution of poverty in the period 2002-2003. The latter implication will be dealt with later.

With respect to the change in well-being over time, according to a group of researchers from Mozambique's National Directorate of Planning and Budget, Ministry of Planning and Finance, Economic Research Bureau, International Food Policy Research Institute and Purdue University who produced a poverty report in 2004 (hereafter DNPO et al., 2004), the Mozambican government has been very successful in reducing the incidence of poverty between the periods 1996-1997 and 2002-2003, from 69.4% to 54.1%. During the same period inequality levels increased only slightly from 0.40 to 0.42 (Fox, Bardasi & Van den Broeck, 2005; World Bank, 2007).

Day-to-day reality, however, suggests that the decline in poverty was neither widespread nor fast. For instance, on the one hand, malnourishment, access to formal quality education, access to basic services such as clean water and electricity, access to food markets and access to financial services, among other things, are still major problems in this economy, mainly in the rural parts, where the biggest share of the population lives.

These perceptions are supported by a number of studies. The World Bank (2007) argued that despite the apparent economic success, Mozambique paradoxically suffers from rising chronic child malnutrition. A Poverty and Vulnerability Survey conducted in four provinces in 2006 indicated that around three-quarters of the Mozambican population felt that their socio-economic situation did not improve or even that it became worse (World Bank, 2007, p. 24). Hanlon (2007) argued that the headline suggesting that poverty declined dramatically to 54.1% is exaggerated. The headline is used by ministers and donors alike to give the impression that though Mozambique continues to be very poor, development is taking place (Hanlon, 2007, pp. 8-9). In another article, the same author indicated that while this economy benefited from substantial economic growth, "[it] does not trickle down to the poor segments of the population [and this fact is therefore] increasing tensions" (Hanlon, 2010, p. 79).

While studying the change in poverty and inequality over time is important to evaluate the performance of programmes aimed to increase well-being, this is not the primary focus of this work. The focus is instead on the poverty profile estimated for the period 2002-2003. In the study done by DNPO et al. (2004), the poverty line derived for Maputo City was very high, more than twice the poverty lines of most Central and Northern parts of the country and more than three times

the poverty lines of the rural parts of Nampula, Sofala and Zambezia. This extremely high poverty line overstates poverty in Maputo City, with consequences for the targeting of policy.

The research problem this chapter addresses is that since the existing studies are exclusively based on money-metric indicators and poverty lines are adjusted to reflect local tastes and prices, poverty and inequality in Mozambique are not well understood. Not addressing this research problem might result in a flawed poverty mapping. Owing to this, policymakers will be improperly informed, and therefore will take erroneous decisions regarding poverty and inequality reduction policies, while the poorest quintiles of the population continue enduring the negative effects of these phenomena.
So far for the case of Mozambique no study yet tried to address this research problem using the approach followed here[4]. It is thus the overall objective of this study to understand poverty and inequality in Mozambique better by using micro-level survey data sets and by exploring other dimensions of these phenomena. More specifically, the aims of the study are the following:

(i) to understand how the existing money-metric poverty and inequality indicators are compiled. This objective is worth achieving so the components of the living standards indicator are identified and the variable itself better evaluated. Further, in earlier studies it is not clear how such a measure is compiled.

(ii) to evaluate the money-metric results. This indicates whether other dimensions of the studied phenomena should also be analysed.

(iii) to propose an asset index computed using Multiple Correspondence Analysis (MCA) as a technique for complementing (but not substituting) the money-metric indicator of poverty and inequality. The use of an asset index is also a way of trying to show whether variable poverty lines present answers which are not convincing. This objective addresses directly the central research problem this study tackles.

In the light of the above mentioned research problem and aims, money-metric analysis should be complemented by analysis of other dimensions of poverty and inequality. Rather than considering an individual poor just by the amount he or she earns or consumes in a certain period of time, it is critical to also analyse the type of housing he or she lives in, the type of assets owned by the household he or she lives in, and the access to public services such as electricity and water, among other things. This therefore will result in a more thorough understanding of the individual's well-being.

[4] See the asset index approach in Section 2.3.2 for details.

One potential limitation when analysing money-metric poverty, is the variable poverty lines developed in DNPO et al. (2004, p. 37), the Mozambican official poverty lines. These poverty lines vary by region, taking into account local tastes and prices of the bundle of goods the poor consume. The adoption of the official poverty lines is justified since the intention is to evaluate the money-metric results produced by DNPO et al. (2004). Nevertheless, this study also makes use of other poverty lines, which permits analysing the sensitivity of the official poverty estimates to the assumptions regarding the design of the poverty lines.

The practical implication of this work is its direct contribution to a finer understanding of the well-being of the Mozambican population, by combining money-metric and other dimensions of poverty and inequality. By better understanding who the poor are and how unequal the various socio-economic groups are, policymakers, researchers and the like will be able to target their poverty and inequality reduction policies with greater precision.

After this brief introduction to the chapter, Section 2.2 reviews the literature on poverty and inequality analysis. This includes both studies that have dealt with money-metric indicators and those that analysed other dimensions of poverty and inequality. Next, Section 2.3 expounds the methods used to accomplish the objectives of this study. It sets out the necessary steps for the construction of a comprehensive measure of consumption for welfare analysis. It also puts forward the methods followed for the computation of the asset index that will also be used for welfare analysis. Section 2.4 reports the results, their interpretation and discussions. Lastly, Section 2.5 concludes the chapter.

2.2 Understanding Poverty and Inequality

The first part of this section reviews works that have to do with poverty and inequality measurement. It is the theory base of this chapter in order to give the reader 'the big picture' (Hofstee, 2006, p. 95). The second part considers studies that followed a money-metric approach to analyse poverty and inequality in Mozambique and discusses the techniques used by other scholars for poverty and inequality analyses. The third part deals with studies that used an asset index to analyse welfare. This group of studies are reviewed last since they lead directly into what this work proposes to do. The fourth and last part summarises and concludes the section.

2.2.1 Theoretical Considerations

As stated by the World Bank (2004, p. 30), computing a poverty measure entails choosing the relevant dimensions and indicators of well-being, selecting a poverty line, and then selecting the poverty measure to be used for reporting the results. With respect to the first requirement, the

analyst has to choose between monetary and non-monetary indicators. When a monetary dimension is chosen then he or she has to decide between using income or consumption expenditures as the living standards indicator. The World Bank (2004) argued that, for the case of developing countries, data on income are poorly available. Therefore, consumption might be a better indicator since, among other advantages, it reflects better the actual standard of living of the household and its ability to meet basic needs. Other benefits of using consumption over income include the fact that consumption has fewer fluctuations than income, consumption has lower shares of non-responses in household surveys and the measurement of income from self-employment activities is rather difficult.

As is also argued in this chapter, the World Bank (2004) stated that poverty is also associated with inadequate outcomes in other dimensions. Inadequate health, deficient nutrition and insufficient literacy, deficient social relations, insecurity and a low self-esteem and powerlessness are also synonymous with poverty. This view is shared by a number of scholars, who argued that the money-metric poverty indicators' coverage is too limited (Sen, 1985; Sen, 1987; Ravallion, 1996; Kingdon & Knight, 2003).

With respect to the second requirement for estimating poverty, poverty lines are cut-off points separating poor individuals or households from the non-poor (World Bank, 2004, p. 33). The researcher should opt between using an absolute or a relative poverty line. In the first case the line is set for a particular person or household without reference to other persons or households, and it is anchored in some absolute standard of what a person or household is able to count on to meet its basic needs. In the second case the line is defined in relation to the overall distribution of the selected indicator in the population. Common examples include, for instance, setting the poverty line at 50% of the population's mean or median consumption expenditures, as is the case for the OECD countries.

The third requirement for computing a poverty measure is the creation of a statistical function translating the selected living standards indicator and the poverty line into an aggregate number for the population (World Bank, 2004, p. 34). A number of statistical functions were proposed in the literature[5], but most of them failed to satisfy basic axioms required to consider a poverty measure reliable. Sen (1976) put forward the monotonicity and the transfer axioms. Kakwani (1980) proposed the transfer sensitivity axiom, and Foster, Greer & Thorbecke (1984) suggested the subgroup monotonicity axiom.

[5] For examples see Watts (1968), Sen (1976), Anand (1977), Kakwani (1977), Van Ginneken (1980), Clark et al. (1981), and Foster, Greer & Thorbecke (1984).

Foster et al., (1984), in their seminal work "*A Class of Decomposable Poverty Measures"*, made a case for a statistical function that satisfies all the required axioms. They argued that their measure was the only one that, under certain conditions, also satisfied the axiom they proposed, therefore permitting poverty analyses by population subgroups.

The basic equation for their statistical function is the following:

$$P_{\propto}(y;z) = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^{\propto}, \qquad (1)$$

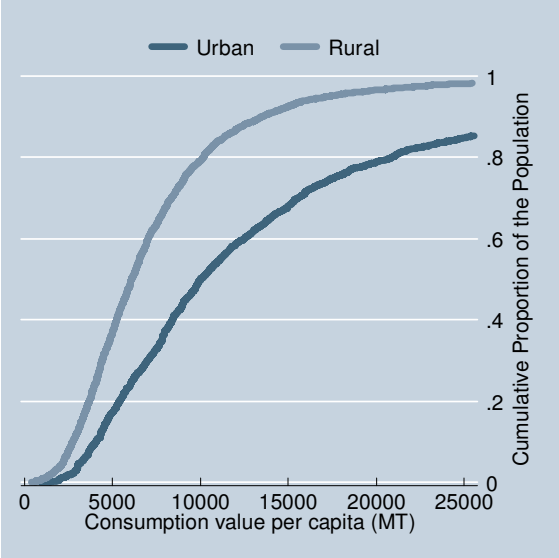
where P_α stands for the poverty index; q stands for the number of poor persons in the population; n is the total population size; y_i stands for income, consumption expenditures or any other variable proxying for welfare of the *i*th individual; z is the poverty line; and α is a measure of poverty aversion. According to the authors, a "[l]arger α gives greater emphasis to the poorest poor" Foster et al. (1984, p. 763).

Making use of the above statistical function Foster et al. (1984) proposed three poverty measures (hereafter FGT), namely the poverty headcount index (P_0), the poverty gap ratio (P_1) and the squared poverty gap ratio (P_2). The poverty headcount index measures the incidence of poverty, i.e., it gives the percentage of the population below the poverty line. The poverty gap ratio measures the depth of poverty. It is the mean income shortfall with respect to the poverty line across the whole population, expressed as a percentage of the poverty line. In other words, "it estimates the total resources needed to bring all the poor to the level of the poverty line (divided by the number of individuals in the population)" (World Bank, 2004, p. 35). Lastly, the squared poverty gap ratio has no easy interpretation. It is commonly known to give the severity of poverty, as it considers the position of the poorest individuals in the population.

The measures described above, however, suffer from two deficiencies. First, the rankings of two or more income distributions in terms of poverty are sensitive to where in the income distribution the poverty line, z, is set. Second, the ranking is sensitive to the particular measure chosen to report poverty, i.e., P_0, P_1 and P_2. Poverty dominance addresses both of these deficiencies (Madden & Smith, 2000, p. 189). If one considers two income distributions y_1 and y_2 with cumulative distributions $F(y_1)$ and $F(y_2)$, if $F(y_1)$ lies nowhere above and at least somewhere below $F(y_2)$, then distribution y_1 displays first-order stochastic dominance over distribution y_2 (Haughton & Khandker, 2009, p. 97). The poverty ranking remains the same irrespective of using P_0, P_1 or P_2.

Poverty dominance may be better understood by an illustrative example using cumulative density functions (CDFs). CDFs measure the cumulative percentage of the population on the vertical axis and the living standards indicator on the horizontal axis. Figure 2.1 shows two CDFs which describe the distribution of consumption value per capita across urban and rural locations. Assuming that the range of consumption value per capita displayed in the figure is the relevant range, the urban CDF lies below the rural CDF for all levels of the consumption value per capita. Thus the urban distribution displays first-order stochastic dominance over the rural distribution, i.e., irrespective of the poverty line chosen the conclusion is that poverty is higher in rural locations. Since the CDFs do not cross, the poverty ranking is unambiguous. If the CDFs cross, however, it is not clear whether poverty is greater in rural or in urban locations. For some poverty lines poverty will be greater in rural locations, while for other poverty lines poverty will be greater in urban locations.

Figure 2.1 – First-order poverty dominance



To resolve the issue when CDFs cross, one could either restrict the range of the poverty line and search for dominance within that range, or analyse second-order stochastic dominance. In the latter case, if one, for example, is comparing poverty between two years, "a fall in poverty then requires that the poverty deficit curve, given by the area under the cumulative distribution, be nowhere lower for year 1 at all points up to the maximum poverty line, and at least somewhere higher" (Haughton & Khandker, 2009, pp. 95-96). One should keep in mind however, that because the CDFs cross, poverty rankings given by the headcount poverty index are not unambiguous, and therefore P_0 is excluded from the analysis. Second-order stochastic dominance requires that the poverty analysis focuses on P_1 and P_2 only, additive measures that reflect the depth of poverty.

With respect to inequality, as stated in Litchfield (1999, p. 1), it is a concept broader than poverty since it is defined over the whole population. It considers individuals or households both below and above the poverty line. Inequality means that if for instance the population is ranked on income from poorest to richest and then divided into fifths (quintiles), those in the top fifth (quintile 5) get a greater share of total income than those say in the bottom fifth (quintile 1). The greater the share of income allocated to richer quintiles relative to poorer quintiles the greater the level of inequality. The reliability of the inequality measures also depends upon whether they satisfy certain basic axioms. Cowell (1985) puts forward a number of these axioms, namely the Pigou-Dalton transfer principle, the income scale independence, the principle of population, the anonymity (symmetry), and the decomposability axioms.

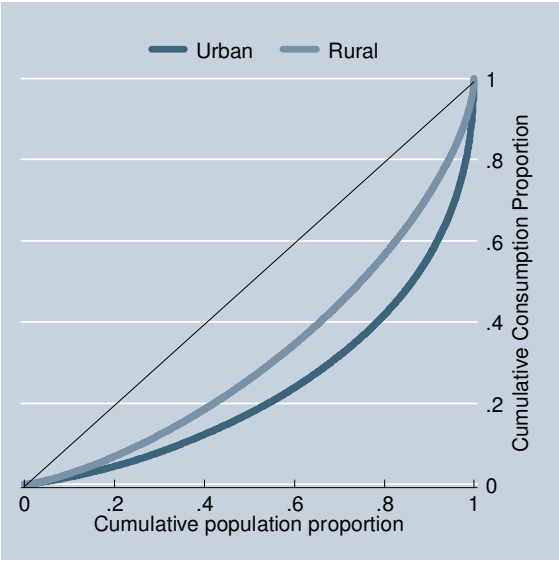
According to the World Bank (2004, p. 78), the most commonly used inequality measure is the Gini index. It varies from zero, which means complete equality among the members of the analysed population, to one, which means complete inequality. It represents the expected difference in income, consumption expenditures or any other living standards indicator of two individuals or households randomly selected from the population as a whole (World Bank 2004, p. 79). It has the advantage of being easy to interpret and of handling negative values. Nevertheless, it is not decomposable[6] and although it is sensitive to changes in the distribution, it does not discriminate as to where in the distribution these changes occur. The Gini index can be presented in terms of a curve named Lorenz curve. A Lorenz curve is constructed by graphing the cumulative percentage of the population on the horizontal axis and the cumulative percentage of the living standards indicator (e.g. consumption) on the vertical axis. "If the Lorenz curve of distribution y_1 lies nowhere below and at least somewhere above of the Lorenz curve of distribution y_2, then y_1 Lorenz dominates y_2" (Litchfield, 1999, p. 6). In other words, distribution y_2 is more unequal than distribution y_1.

Figure 2.2 on page 36 shows two Lorenz curves which describe the distribution of consumption expenditure across locations. The Lorenz curve of the rural distribution lies above the Lorenz curve of the urban distribution. Thus rural locations Lorenz dominate the urban, i.e., urban areas are more unequal than rural areas.

Another inequality measure commonly used in research is the Theil index. It is basically a measure of discrepancy between the income share and the population share of the analysed groups. It is interpreted as "the expected information content of the indirect message which transforms the population shares as prior probabilities into the income shares as posterior probabilities" (Theil, 1967, pp. 125-126).

[6] This means that the total Gini of society is not equal to the sum of the Gini coefficients of its subgroups.

Figure 2.2 – Lorenz-dominance



In contrast to the Gini index, the Theil index is additive across different groups and largely sensitive to transfers from poor to rich people or households. It is always positive, but the contribution of each group can be negative. If each group has the same share of income as the share of the population, this index is equal to zero (complete equality) and groups whose population share is the same as the income share contribute neutrally to the index, i.e., these groups' contribution is zero (Conceição & Ferreira, 2000, p. 13). This index is decomposable into between-groups and within-group inequality. However, its interpretation is not as straightforward and as appealing as is the case for the Gini index (World Bank, 2004, p. 48).

2.2.2 Money-metric Poverty and Inequality

After this 'big picture' on poverty and inequality measurement, the second part of this section reviews studies that followed a money-metric approach to analyse poverty and inequality in Mozambique. Maximiano, Arndt & Simler (2005) analysed the correlates of poverty using the second Mozambican national income and expenditure household survey, performed in the period 2002-2003. The authors used a multivariate ordinary least squares (OLS) model. Their objective was first to identify the correlates of poverty in 2002-2003 and then to analyse their dynamics between the periods 1996-1997 and 2002-2003. Their model was specified as follows:

$$\ln c_j = \beta' x_j + \eta_j, \tag{2}$$

where $\ln c_j$ stands for the natural logarithm of real consumption per capita for household j, x_j is a group of characteristics of the household plus other determinants, and η_j is a random error term. The real consumption was normalised using price indices available in the survey and spatially

deflated by the poverty line (Maximiano et al., 2005, p. 3). They ran a model by location and by region.

In sum, Maximiano et al. (2005) found that bigger households and female headed households tended to be poorer. Second, they found that older people tended to be poorer. Third, they found that educational attainment was positively related to consumption per capita. In other words, schooling was negatively associated with poverty. Lastly, the authors concluded that the correlates of poverty in the period 1996-1997 were still roughly the same as in 2002-2003.

A work more related to the research problem addressed in this study is DNPO et al. (2004). It is a government report that aimed to present the methods used and the results of the poverty analysis using the second Mozambican national income and expenditure household survey (IAF2002-03). The report also compared these results to the poverty rates estimated using the preceding national income and expenditure household survey (IAF1996-97).

To analyse poverty DNPO et al. (2004) used a comprehensive measure of consumption, i.e., they analysed the money-metric dimension of poverty. Their indicator included both food and non-food items, and imputed use values for household durable goods and owner-occupied housing, drawing from several modules of the survey. They argued that their approach closely followed Deaton & Zaidi's (1999) and Deaton and Grosh's (2000) recommendations for constructing consumption aggregates for welfare analysis.

To construct poverty lines they applied the popular cost of basic needs (CBN) approach. This approach entails selecting a bundle of goods (containing food and non-food components) consumed by the poor and thought to be appropriate for satisfying a person's basic consumption needs. Once the bundle is selected and assuming it is socially accepted, it is costed, accounting for price differentials across spatial regions (Ravallion & Bidani, 1994). For the case of Mozambique they first looked at the typical diet of the poor in each of thirteen spatial regions which they defined. Then they constructed food bundles for each spatial domain and calibrated them so as to attain an average of 2,150 calories per person per day. By costing the food bundles with the prices prevailing in each spatial domain, they generated thirteen food poverty lines.

Concerning the non-food poverty lines they simply used the non-food shares derived in the previous report, which used data from IAF1996-97. In the previous report the approach was to look at how much households that were in the neighbourhood of satisfaction of their food basic needs spent on non-food items. In other words, the non-food poverty lines were set by examining the non-food consumption among those households whose total expenditure per capita was roughly equal to the

food poverty line in each spatial domain[7]. This is based on the argument that, since the non-food consumption of those who could just barely satisfy their food consumption needs displaces essential food consumption, such non-food consumption could itself be considered essential (DNPO et al., 2004, pp. 14-15).

The total poverty line was simply derived as the sum of the food and non-food poverty lines in each spatial region. Table 2.1 on page 39 shows these poverty lines across spatial domains. The total poverty line derived for Maputo City was very high, more than twice the poverty lines of most Central and Northern parts of the country and more than three times the poverty lines of the rural parts of Nampula, Sofala and Zambezia. It is interesting to note, however, that mean consumption expenditures per capita were significantly greater in Maputo Province and Maputo City. This suggests that standards of living were highest in these two regions, assuming only small price differentials across regions.

In order to analyse the change in the poverty estimates between the periods 1996-1997 and 2002-2003, they first used fixed bundles of goods to construct thirteen region-specific food poverty lines. In other words, the bundle of goods derived in each spatial domain for the period 1996-1997 was exactly the same in 2002-2003. The food poverty lines for the latter period were simply updated for inflation. This approach was justified on the grounds that "it [was] the only way to be sure that the food poverty lines represent[ed] equal levels of welfare" (DNPO et al., 2004, p. 7).

However, when the relative prices of food are subject to significant changes over time, the estimates of poverty will not be consistent if the initial consumption bundle is used in the analysis. Using a fixed bundle would overstate the poverty levels in the next period. For the particular case of Mozambique, between the periods 1996-1997 and 2002-2003 the price of maize flour increased relative to the price of cassava flour (DNPO et al., 2004, p. 9). As a result, consumers substituted away from consuming maize flour to cassava flour. This required the creation of bundles of goods that reflected that substitution in the diet of the poor, i.e., it required the use of flexible bundle of goods instead. The poverty lines resulting from the flexible approach are those reported in Table 2.1.

DNPO et al. (2004) found that between the periods 1996-1997 and 2002-2003, the FGT poverty measures declined. Using the fixed bundle of goods to construct the food poverty lines, the poverty headcount index declined from 69.4% to 63.2%, the poverty gap ratio declined from 29.3% to 25.8%, and the squared poverty gap ratio declined from 15.6% to 13.5%. Employing the flexible

[7] The spatial domains used in the period 1996-1997 were the same in the period 2002-2003.

bundle approach, however, the FGT poverty measures declined to 54.1%, 20.5%, and to 10.3%, respectively. Thus using flexible poverty lines resulted in a stronger decline in poverty, giving the impression that the decline in poverty was substantial.

Table 2.1 – Official money-metric poverty lines in the period 2002-2003

Spatial Regions	Food poverty lines	Non-food poverty lines	Total poverty lines	Mean consumption per capita
North				
Niassa and Cabo Delgado-rural	5,434	1,665	7,099	8,125
Niassa and Cabo Delgado-urban	7,540	2,690	10,230	14,341
Nampula-rural	4,471	1,501	5,972	6,460
Nampula-urban	4,853	1,807	6,660	9,711
Centre				
Sofala and Zambezia-rural	4,155	1,318	5,473	8,672
Sofala and Zambezia-urban	6,591	2,183	8,774	16,373
Manica and Tete-rural	5,629	1,304	6,933	8,401
Manica and Tete-urban	7,145	2,545	9,690	14,472
South				
Gaza and Inhambane-rural	6,614	2,394	9,008	7,732
Gaza and Inhambane-urban	7,264	3,457	10,721	13,726
Maputo Province-rural	11,801	4,963	16,764	9,852
Maputo Province-urban	11,898	6,398	18,296	21,519
Maputo City	12,224	7,291	19,515	36,367

Note: figures in Meticais (MT) per person per day; Source: DNPO et al. (2004, p. 37); The consumption data are sourced from the data set

Hanlon (2007) disagreed with the use of the flexible bundle approach. He argued that the correct figures for the FGT poverty estimates in the period 2002-2003 are those that used fixed food bundles to derive the region-specific poverty lines, i.e., 63.2%, 25.8% and 13.5%. According to him when DNPO et al. (2004) used the flexible bundle approach they only looked at calories, therefore not taking into account other nutrients. The difficulty is the fact that cassava flour is much less nutritive than maize flour. Moreover, "it is a poverty food and the poorest eat proportionately more cassava and less other grains" (Hanlon, 2007, p. 10). The flexible bundle approach did not take into account that the flexible food poverty lines were in fact of lower quality, in the sense that the proportion of inferior goods in the bundle increased. This therefore biased the poverty estimates downwards. He also argued that the flexible approach led to huge differences between the spatial poverty lines. For instance, the poverty line in Maputo City was more than three times the line in some rural areas (Hanlon, 2007, p. 10).

DNPO et al. (2004) tested the sensitivity of their results to the choice of the poverty line following the welfare dominance approach. They found that between the two income and expenditure household surveys (i.e., IAF1996-97 and IAF2002-03), for most poverty lines their results were

robust. However, by comparing rural to urban areas they could not find stochastic poverty dominance, meaning that the poverty estimates were not robust to the choice of the poverty line.

Another work that fits in this section is Fox, Bardasi & Van den Broeck (2005). In their study the authors used IAF1996-97 and IAF2002-03 to analyse growth in household consumption, the change in the distribution of that growth and the interaction of these two factors in reducing poverty. They also analysed the non-money-metric dimension of poverty such as the change in access to services and the change in asset ownership by the households. In brief, they found that due to a strong increase in agricultural production in particular, poverty declined sharply over the analysed period, mainly in the rural areas. They argued that that was mostly due to the fact that inequality did not change much. The increase in inequality was small, thus allowing aggregate consumption growth to reach the lowest income quintiles of the population (Fox et al., 2005, p. 1).

As was the case with DNPO et al. (2004), Fox et al. (2005) also used consumption as the living standards indicator. However, in adjusting their indicator for the size of the household for comparisons of welfare among the population, instead of doing just a per capita basis adjustment, they adjusted the per capita consumption to reflect the needs of each type of household member (i.e., they used adult equivalent scales). Nonetheless, the aggregate poverty rate was not sensitive to this different approach to poverty measurement. In their words, "[the] most important point to note about poverty trends in Mozambique is that regardless of which method is used to adjust for household composition, poverty in Mozambique fell dramatically between [1996-1997] and [2002-2003]" (Fox et al., 2005, p. 2).

Fox et al. (2005) also found that, on the one hand, the expenditure share on durable goods such as radios, bicycles, television sets and housing, among other things, increased. On the other hand, they found that, between the two surveys, the food expenditure share declined in all income quintiles. This is consistent with the fact that as households got richer they spent proportionately less on food items, and proportionately more on durable goods, following Engel's Law[8]. In other words, as households' income rise their marginal propensity to consume food declines, while their marginal propensity to consume durable goods increases. Simply put, food is an inferior good.

Concerning access to public services, the authors found that water and sanitation became more available in the period 2002-2003, for all income quintiles. Similar results were found for education and health facilities. However, with respect to electricity they found that the lowest income quintile did not experience any increased benefit from this service.

[8] Engel's Law states that as income rises the proportion of expenditure on food decreases (Zimmerman, 1932, p. 78).

In addition to poverty analysis, Fox et al. (2005) studied the inequality changes between the two household surveys. Using both the Gini coefficient and the Theil index they found that for Mozambique as a whole inequality increased from 0.38 to 0.40 and from 0.29 to 0.34, respectively (Fox et al., 2005, p. 12). Though qualitatively similar results, the World Bank (2007, p. 91) found that for the case of the Gini coefficient the change between IAF1996-97 and IAF2002-03 was from 0.40 to 0.42. In terms of locations, Fox et al. (2005) found that inequality within urban locations was higher than within rural locations. The Theil decomposition also showed that the existing gap between urban and rural areas was not the main factor explaining national inequality, but instead within location inequality. The authors also found that inequality within provinces was higher than inequality between provinces. Lastly, Fox et al. (2005) also made use of IAF2002-03 to identify the poverty correlates in Mozambique, employing multivariate regressions. They found that the most important correlate of household consumption per capita was the education of the head of the household, particularly in urban areas (Fox et al., 2005, p. 45). People living in households whose heads were more educated tended to be less poor.

2.2.3 Multidimensional Analysis of Welfare

The next part of this literature review deals with studies that used an asset index to analyse living standards. Two important articles on poverty analysis which applied the method proposed in this chapter are Sahn & Stifel (2000) and Booysen, Van der Berg, Burger, Von Maltiz & Du Rand (2008). As also suggested in this study, they applied a multidimensional approach to the analysis of poverty. The first article used Demographic and Health Surveys to compute the asset index. This index was then used to analyse poverty in nine sub-Saharan countries (Mozambique not included). In this work the asset index was computed using the factor analysis approach (FA).

Booysen et al. (2008) extended the work of Sahn & Stifel (2000), firstly by making use of more recent surveys, thus allowing them to compare poverty at three rather than two points in time for each country. This made their conclusions regarding trends in poverty more stable given the longer time span covered. In contrast to Sahn & Stifel (2000), however, to compute the asset index Booysen et al. (2008) employed an approach named Multiple Correspondence Analysis (MCA), rather than FA. Another novelty is that they reported their results using both an asset index based on pooling across countries and individual indices compiled for each country separately. This permitted an investigation of the robustness of their conclusions with respect to the poverty trends. They tested the sensitivity of their results to measurement and sampling errors, and to the choice of the poverty line (Booysen et al., 2008, p. 1114). In sum, both studies found that poverty declined in most of the analysed sub-Saharan African countries.

2.2.4 Summary of Literature

The first part of this section introduced 'the big picture' on the matter. It expounded the requirements for the compilation of a reliable poverty measure. Further, it described the most commonly used inequality measures in the literature. The second part reviewed works that followed a money-metric approach to analyse poverty and inequality in Mozambique. While Maximiano et al. (2005) analysed the poverty correlates in Mozambique, they did not tackle the research problem this study addresses. The authors did not attempt to measure poverty and inequality and did not look at other dimensions of welfare analysis. Nevertheless, their study is important since knowing what 'causes' or reduces poverty is essential to inform policy.

DNPO et al. (2004) provided the official Mozambican poverty estimates and is a work more directly related to this chapter. The authors analysed poverty in Mozambique using the second national income and expenditure household survey. Though they gave a few explanations on their methods, it is not transparent how exactly their consumption aggregate was compiled. An accurate replication of their results is thus not possible. The derivation of the consumption aggregate is no easy matter and how it is done has implications for policy targeting. The poverty estimates depend on which elements of consumption are included or excluded from its derivation.

Another deficiency with this work is that, as Hanlon noted, when using the flexible bundle approach to estimate the change in poverty over time they did not take into account the fact that households substituting their initial basket of goods for a lower quality bundle of goods were in fact becoming worse off, even if they spent the same amount of money as before. Further, their variable poverty lines lead to counter-intuitive results regarding the spatial distribution of poverty. Maputo City comes as one of the poorest provinces in Mozambique, an aspect that this chapter aims to disprove. Additionally, the authors did not analyse inequality. Lastly, though they tested for the robustness of their results, the welfare dominance approach was applied, but there was no testing for the sensitivity of the poverty measures to the poverty line in each of the thirteen spatial regions.

Fox et al. (2005) is a more complete work in the sense that it addressed the research problem this chapter tackles. The authors analysed poverty and inequality, and incorporated a non-money-metric dimension in their analysis of poverty. However, their non-money-metric dimension did not consider inequality analysis, and the non-money-metric poverty analysis was not as comprehensive as is the case for the asset index method suggested in this chapter.

The third part of this section reviewed articles that used an asset index to analyse welfare. Both the studies by Sahn & Stifel (2000) and Booysen et al. (2008) are important to the problem addressed in

this study. Though they applied the asset index approach to sub-Saharan countries, they did not include Mozambique. For the case of Booysen et al. (2008), the most likely reason for excluding Mozambique would be that only two sets of Demographic and Health Surveys were available for this country, while the authors wanted to analyse poverty across three points in time. This chapter however focuses on poverty and inequality analysis at one point in time and the author is therefore able to apply the asset index approach to analyse well-being in Mozambique.

In the light of the above arguments, this chapter addresses the identified lacunas in the works reviewed. Doing so will permit a greater understanding of poverty and inequality in Mozambique.

2.3 Methods and Data

By using the methods proposed in this section, that is, the compilation of a consumption aggregate and of an asset index for welfare analysis, this study aims to better understand the concepts of poverty and inequality in the Mozambican context. Using consumption as the living standards indicator is very common in the poverty and inequality literature, but the computation of an asset index using Multiple Correspondence Analysis (MCA) is a rather new technique and it has never been applied to any Mozambican data set so far. The latter method will help in coming to a solid conclusion about the thesis defended in this study, since it introduces a non-money-metric dimension to the poverty and inequality analyses.

The data used in this chapter are sourced from the second Mozambican income and expenditure household survey (IAF2002-03). This survey was produced by the National Statistics Institute (*Instituto Nacional de Estatística* – INE), and was performed during the period July 2002 to June 2003. It was based on a stratified random sample and was designed to be representative of the nation, provinces, and area of residence (rural and urban). It covered 44,083 individuals living in 8,700 households. The main purpose of this survey was to measure the progress in the fight against poverty, and to provide information for future policies (DNPO et al., 2004, p. 1). The survey collected, among other things, detailed information on household consumption expenditures, asset ownership, housing characteristics and access to public services such as water and electricity.

The extrapolations of the data using the sampling weights provided in the data set suggest that in the analysed period the Mozambican population was around 18.3 million inhabitants[9]. This is consistent with the results from the population census performed in 2007 that suggested that there

[9] According to Deaton (1997, p. 48) the total of individual weights, or inflation factors, is the total number of people in the population. Individual weight is simply the product of household weight and household size, both given in the household survey.

were 20.0 million inhabitants in that year[10]. The number of households was 3,808,434, and the percentage of females (52%) in the population was higher than the percentage of males (48%). In terms of age, 50% of the population was aged 17 years old or younger and only about 1% was aged 75 years or older. If non-respondents on education status are dropped from the analysis, 50% of household heads had no formal education in the period 2002-2003, and only 25% of them completed the first level of primary education, i.e., only a quarter of household heads were functionally literate. These figures become more drastic if it is assumed that non-respondents never attended school; the figures change to 64% and 18%, respectively.

In terms of methods, this study applied descriptive and inferential statistics, in particular tabulations, correlations, regression techniques, multiple correspondence analysis, maps and statistical graphs for presentations, as well as reviews of the existing literature on the topic. From these techniques, two deserve a subsection of their own for a detailed explanation. Therefore, Subsection 2.3.1 deals with the compilation of the consumption aggregate and Subsection 2.3.2 describes the computation of the asset index.

2.3.1 The Consumption Aggregate

This subsection describes the approach followed for the compilation of the comprehensive measure of household consumption per capita. It closely follows the recommendations of Deaton & Grosh (1998) and Deaton & Zaidi (2002) on the matter, the current best practice. The first few paragraphs describe the questionnaires contained in IAF2002-03, the second income and expenditure household survey. Then, a brief explanation of the way the consumption aggregate was assembled follows.

Theoretically compiling an aggregate measure of consumption is a straightforward exercise. One just needs information on consumption value per household and household size. In practice, however, the exercise was not clear-cut. The following paragraph summarises the issues:

> *"Even for consumer expenditure surveys in industrialised countries, which have been extensively documented and where there has been a good deal of experimentation, the literature has not produced a satisfactory synthesis between theory and practice. The literature itself is difficult to find and is scattered across various disciplines, including economics, marketing, psychology, sociology, and statistics. Much of it is contained in poorly catalogued government reports and conference proceedings rather than in academic journals" (Deaton & Grosh, 1998, pp. 5-6).*

[10] From 2002-2003 to 2007 the population grew by about 1.7 million.

Thus compiling the consumption aggregate involved many practical complexities. The aggregate consumption indicator had to be built from several components spread among various modules of IAF2002-03. This was necessary since the reliability of this indicator depended on whether it was indeed comprehensive, i.e., it depended on whether all the important goods and services that contributed to the population's living standards were included[11]. Nevertheless, "there is not a clear right or wrong way of doing things for many issues about how to measure consumption. Rather, there is a range of good practice techniques and not enough empirical evidence about which is best" (Deaton & Grosh, 1998, p. 8). The poverty estimates will be sensitive to which method is used to compute the consumption aggregate.

IAF2002-03 contains four questionnaires. The first questionnaire has questions on the general characteristics of the household. It collected, among other things, demographic information of the household members, information on their education, health status, employment status and housing characteristics. Information on household size was obtained from the first module of this questionnaire. The second questionnaire, in its first module, gathered data on household daily expenditures on food. These included expenditures on food consumed both inside and outside the dwelling. Its second module collected data on *autoconsomation*, i.e., food that was home-produced for own consumption purposes. The last module dealt with revenues in-kind. The third questionnaire collected data with respect to general expenditures and revenues of the household. The first part asked questions on ownership of durable goods. The second part collected data on annual expenditures on education. The remainder of the modules gathered data on monthly expenditures on clothing, rents, furniture, household appliances, health, transportation and leisure, among other things. The fourth one was a community questionnaire. It was administered only in rural areas. It gathered data on the village's economy and infrastructures, villagers' access to education and health facilities, agricultural production as well as prices in the local village markets.
To compute the aggregate measure of consumption first the household food component was compiled. It was simply the row sum of food purchases, *autoconsomation* and food in-kind received by each of the uniquely identified households. The total of this sum was the total food consumption value, in Meticais[12] (MT) per household per day[13].

[11] Not all goods and services contained in the data set were included. For more information on the components that should not be included in the aggregate consumption indicator, please see Deaton & Grosh (1998) and Deaton & Zaidi (2002).
[12] Metical is the Mozambican official currency. According to the Penn World Table 6.2, in 2003 the Metical per US$ exchange rate averaged 23,782.27.
[13]The questionnaires contained in the World Bank's Living Standards Measurement Study surveys normally contain three questions which make compiling the food purchases aggregate an easy process: (a) the number of months the specific good was bought in the past 12 months; (b) the quantity bought in a typical month; and (c) the amount normally

The second step dealt with the non-food component. The yearly expenditures on education and the other monthly non-food expenditures had first to be converted into the same unit as above (i.e., MT per household per day). Then, the row sum of the education expenditures, imputed rents for owner-occupied housing, imputed use-value of services provided by the ownership of durable goods (assets)[14] and the other non-food expenditures resulted in the total non-food consumption value in MT per household per day.

After merging the different files together[15] the total consumption value in MT per household per day was simply the row sum of the total food consumption value and the total non-food consumption value, by household. To obtain a per capita consumption value just entailed dividing this variable by the household size. The variable created out of this was then used for the money-metric poverty and inequality analyses. Adult equivalent scales could have been used to reflect lower consumption needs of children relative to adults. But that was not the main focus of this chapter and, as mentioned before, the aggregate incidence of poverty in Mozambique was not sensitive to adult equivalent scales (Fox et al., 2005, p. 2).

When computing this indicator, there was the risk of double-counting products. In other words, there were products such as wild animal meat, sausages, wood, matches and candles, among other things, that were present in both the second and the third questionnaires (See Appendix A for a list of these products and respective codes). The data set included data from both sources. The approach followed to deal with this issue was to delete the data for the product which could be double-counted from the questionnaire that reported the lowest frequency for that product. For example, if new matches were reported to be consumed twice a week in the second questionnaire, but four times a week in the third questionnaire, the consumption data on new matches were deleted from the second questionnaire. Deleting the product from the questionnaire which presented the lowest frequency rather than deleting from the questionnaire which presented the highest frequency avoided biasing the consumption aggregate downwards.

---

spent in total in a typical month. The food purchases aggregate is just the product of (a) and (c) (Deaton & Zaidi, 2002, p. 69). However, the Mozambican questionnaire was organised in a different and more complicated manner. The only important question asked was on "the amount spent to buy the good during the week the diary was left in the household". "While it may not be strictly the case that all food acquired in a given week is consumed that week, the general assumption is that at the monthly or annual level, total food expenditures indicate the value of total food consumed by the household" (Lanjouw, 2009, p. 7). Therefore, for the case of Mozambique, to attain the daily purchases aggregate the author simply divided the relevant amount by seven.

[14] The imputations were done by the National Statistics Institute (INE).

[15] Each of the consumption components came in its own data file.

2.3.2 The Asset Index

A wealth index, also known as an asset index, is a dimension-reducing indicator that ranks households by their long-run socio-economic status. Households with a high wealth index score are expected to be better off than households with lower figures for this index. The computation of the asset index is made difficult since the weights for each of the assets included in the index are not grounded theoretically (Filmer & Pritchett, 2001, p. 116). A simple but not flawless technique to get around this issue is to give equal weights to each of the assets in the household. But this method implies that, say, having a television set carries the same weight in socio-economic status as having a radio, a car or a bicycle, and therefore that any of these assets can correctly and equally predict the household's socio-economic status, something which is not true (Taylor & Yu, 2009, p. 14).

The solution applied in this study was to use a technique named Multiple Correspondence Analysis (MCA)[16]. Greenacre (2007, p. 137) suggested that MCA should be thought of as the analysis of an indicator matrix – the analysis of the whole data set coded in the form of dummy variables. He also suggested that the analyst could think of MCA as the analysis of a Burt matrix – the analysis of all two-way cross-tabulations amongst the variables. Basically, the technique uses the correlations of household ownership of a variety of assets to compute an index that proxies for household wealth.
Principal Components Analysis (PCA) is the most widely used technique for computing asset indices. However, it works better for continuous variables. If the data set in use contains instead categorical variables, the most appropriate technique to compute an asset index is MCA, since it imposes fewer constraints on the data (Blasius & Greenacre, 2006; Booysen et al., 2007). Since in this chapter the variables used to compute the index are categorical socio-economic variables, the asset index was computed using MCA.

Booysen et al. (2007, p. 1115) quoted Asselin (2002, p. 14) for the description of the computation of an asset index using MCA. According to them, the calculation of this index using MCA involves four steps: The first step is the construction of the indicator matrix. This matrix shows the asset ownership by each of the households. Simply put, it tells whether or not a particular household owns a radio, a bicycle, and so forth, or whether it has a house with certain predefined characteristics. The second step is the calculation of the household profile with respect to the categories of asset ownership. Third, MCA is applied to the indicator matrix, therefore providing the first dimension's set of column weights. Lastly, the weights are applied to the households' profiles, therefore generating the following MCA asset index:

$$MCA_i = R_{i1}W_1 + R_{i2}W_2 + R_{i3}W_3 + \cdots + R_{iz}W_z, \quad (3)$$

[16] See Appendix B for further details on the construction of the wealth index for this study.

where MCA_i is the ith household's wealth score, R_{iz} is the response of household i to category z, and W_z is the first dimension MCA weight applied to category z. The wealth index is then used for the non-money-metric poverty and inequality analyses.

For a comparison with the money-metric results, in this study the poverty line was set at the percentile which corresponds to the national estimate of the headcount poverty rate. A potential limitation this study faced was in setting the non-money-metric poverty line. The poverty line chosen might be considered arbitrary or subjective. However, the essential feature was to use a line that permitted comparisons of the poverty analysis using a money-metric indicator to the analysis using the wealth index.

Another issue with this index was that 21.03% of the households interviewed did not respond on their ownership of assets. Since most of these households had no missing information on the remainder of the variables and to avoid having a much smaller data set, these households were included in the asset index. In spite of this drawback, Appendix C makes the case that the data at hand are good enough for the purposes of this study.

A final potential limitation was the fact that the assets included in the wealth index were mostly urban assets. Urban households when compared to rural households of similar economic means are more likely to own a house of cement block walls, concrete roof, piped water, with running electricity and with a flush toilet with sewerage. The same could be argued for the case of ownership of each durable private asset. Consequently, the wealth index might be slightly biased downwards against rural households. But since the data set does not contain data on rural assets, the wealth index was computed using the available data.

2.4 Poverty and Inequality in Mozambique

This section reports the results of the poverty and inequality analyses. The first part deals with poverty, and the second deals with inequality. The third and last subsection discusses the results.

2.4.1 Poverty Analysis

The current subsection introduces the poverty estimates and their analysis. Then, the next subsection runs regressions aimed at identifying the poverty markers, i.e., the correlates of consumption per capita and of the wealth index.

2.4.1.1 Poverty estimates

To carry out the poverty analyses this study made use of the Foster, Greer & Thorbecke (1984) FGT poverty measures. Stata11 (Statacorp, 2009) has a routine of estimating poverty rates. The *povdeco*

command gives the three FGT poverty measures, namely the poverty headcount index (P_0), the poverty gap ratio (P_1) and the squared poverty gap ratio (P_2).

Table 2.2 on page 51 reports the poverty estimates. According to that table using the DNPO et al.'s (2004, p. 37) variable (and official) poverty lines, by the time of the survey (i.e., in the period 2002-2003) 54.36% of the Mozambican population were below the poverty line. The depth (P_1) and severity (P_2) of poverty were approximately 21.02% and 10.82%, respectively. This set of poverty estimates is a very close match to the set of official poverty estimates produced by DNPO et al. (2004).[17] Further, overlooking the issue of potential risk for double-counting products, in this study the FGT poverty estimates were, respectively, 53.50%, 20.34% and 10.37%. This suggests that not correcting this issue did not affect the results much, and it could therefore have been ignored.

As explained in Appendix B, negative values complicate the analysis of poverty. The wealth index was therefore transformed in order to include only positive values. A number slightly higher than the greatest absolute negative value was added to each household's wealth index score. Booysen et al. (2008, p. 1117) argued that the transformation of the asset index into positive values does not preserve the mean, and therefore P_1 and P_2 based on such indices only have meaning in the context of a specific study. Therefore, for the purposes of asset poverty analysis, this study made use of the poverty headcount index only. The poverty line was set at 2.31, the 54.36th percentile of the wealth index, so as to have poverty estimates at the same level as in the money-metric analysis, therefore permitting meaningful comparisons of the two approaches.

In terms of location the table reports that rural regions were poorer (55.28%) than the urban (52.43%). The analysis done using the wealth index leads qualitatively to the same conclusion. However, the gap between rural and urban areas was found to be wider for the latter case. While only 15.07% of the urban population was deemed poor by this measure, 72.39% of the rural population was below the poverty line. Though the latter might be true, it should be taken cautiously since the wealth index tends to be biased against rural households, as mentioned.

Figure 2.3 and Figure 2.4 on page 52 test the robustness of the poverty estimates by location, using cumulative density functions (CDFs). According to these figures, irrespective of the indicator chosen (i.e., consumption aggregate or asset index), the urban CDF does not cross the rural one and it is to the right of the latter. This indicates that the urban distribution first-order dominates the rural distribution, i.e., rural regions are poorer than the urban.

[17] DNPO et al. (2004) found, respectively, 54.1%, 20.5%, and 10.3%.

The CDFs for the wealth index have a bump at the lower end of the distribution. This reflects the fact that there were about 21.03% of households with missing data on household ownership of durable private assets. As indicated, instead of dropping these observations they were included in the computation of the wealth index. In Appendix C Kernel density curves are employed to make the case that the data at hand are good enough for the purposes of this study. As long as the people with missing observations were already at the bottom of the distribution, the wealth index's poverty estimates will not be much affected. Apart from that, the poverty line was set at 2.31, above the missing values, clearly not altering the poverty estimates much. While the latter two points might be considered rather strong assumptions, they reflect some of the difficulties encountered when working with developing country data sets.

Another issue with the wealth index's CDFs is that they become very steep after the bump. This makes it clear that there is a big divide between the various socio-economic groups concerning the possession of assets and household housing characteristics. There is a cluster of people who has nothing and others who have plenty of assets and the best type of household housing characteristics. According to the wealth index's CDFs this aspect is more prevalent amongst the poorest groups of the population.

Table 2.2 also reports poverty estimates across regions[18]. Using the consumption aggregate, the Southern region of Mozambique was the poorest, with 71.14% of the population below the poverty line. It was followed by the Northern and Central regions, with 52.25% and 45.72% of incidence of poverty, respectively. In contrast, the non-money-metric analysis suggests a different ranking. On the one extreme, the South had only 16.82% of the population below the poverty line. On the other extreme, the Central and the Northern regions, in that order, had 69.19% and 63.82% of incidence of poverty.

With respect to the robustness of the poverty estimates to the choice of the poverty line, Figure 2.5 on page 52 suggests that in 2002-2003 there was no poverty dominance by region, unless a poverty line of about 5,000 Meticais (MT) or higher was chosen. In that case, the North would have been the poorest region, followed by the Central and Southern regions, respectively. The analysis done using the wealth index (see Figure 2.6) suggests no poverty dominance between the Northern and the Central regions, but the figure illustrates that the South was clearly the least poor region.

[18] The Southern region of Mozambique includes Maputo City, Maputo Province, Gaza and Inhambane. The Central region comprises Manica, Sofala, Tete and Zambezia provinces. Lastly, the Northern consists of Niassa, Cabo Delgado and Nampula provinces.

Concerning the composition of the population in poverty, both the consumption aggregate and the wealth index suggest that the incidence of poverty among females was higher than that of males. Poverty analysis across genders of the household head, however, leads to different conclusions depending on the indicator used. Using the consumption aggregate, on the one hand, 63.27% of the population living in households headed by a female were deemed poor. That percentage drops to 52.07% for male headed households. On the other hand, using the wealth index, male headed households were poorer than female headed households, with 54.75% and 51.08% of incidence of poverty, respectively. Nevertheless, and as indicated by Figure 2.7 and Figure 2.8 these estimates are not robust to the choice of the poverty line since the CDFs cross.

Table 2.2 – Poverty Estimates

	Money-metric Analysis (%)					Non-money-metric Analysis (%)
	Variable Poverty Line			Other Poverty Lines		
				7,598.83 MT per person per day[19]	1 US$ per person per day	
	P_0	P_1	P_2	P_0	P_0	P_0
National	54.36	21.02	10.82	*54.36*	93.36	*54.36*
By Location						
Urban	52.43	20.67	10.39	33.46	83.64	15.07
Rural	55.28	21.19	11.03	64.24	97.95	72.39
By Regions						
North	52.25	17.65	7.87	59.02	96.77	63.82
Centre	45.72	15.93	7.83	55.80	95.27	69.19
South	71.14	33.58	19.4	46.17	85.92	16.82
By Gender						
Female	55.32	21.54	11.16	55.30	93.62	54.37
Male	53.34	20.46	10.46	53.37	93.09	53.62
By Gender of Head						
Female	63.27	26.31	14.17	57.31	94.15	51.08
Male	52.07	19.66	9.96	53.61	93.15	54.75
By Education of Head						
At least Primary Education Completed	33.01	11.00	5.00	20.47	73.49	12.33
Less than primary education completed	57.45	22.47	11.67	59.27	90.06	60.02
By Provinces						
Niassa	48.60	14.10	6.14	47.57	95.62	55.11
Cabo Delgado	62.22	20.96	9.28	60.39	95.28	71.93
Nampula	48.78	17.13	7.71	61.50	97.75	62.54
Zambezia	46.90	14.72	6.44	64.54	97.18	81.06
Tete	57.36	25.09	14.16	60.98	97.55	76.25
Manica	41.58	16.01	8.66	38.73	93.05	50.29
Sofala	35.66	10.25	4.17	44.62	90.59	50.58
Inhambane	83.54	45.60	29.06	75.10	97.75	35.92
Gaza	66.51	26.92	13.47	55.55	93.34	20.56
Maputo Province	74.62	35.83	21.19	30.96	83.57	4.26
Maputo City	57.55	24.13	12.68	12.51	64.02	0.03

[19] In order to check whether the poverty estimates were sensitive to the assumptions regarding the design of the DNPO et al.'s (2004) variable poverty lines, a poverty line was set at 7,598.83 MT per person per day for all regions of the country. This is the consumption value that leads to a headcount ratio of 54.36%.

Figure 2.3 – CDFs by location (consumption aggregate)

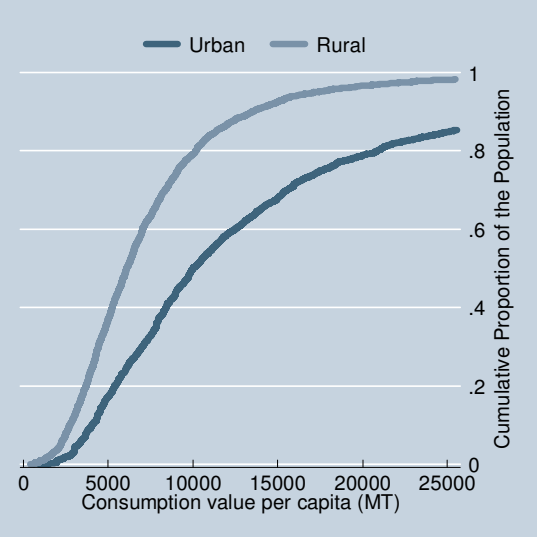


Figure 2.4 – CDFs by location (wealth index)

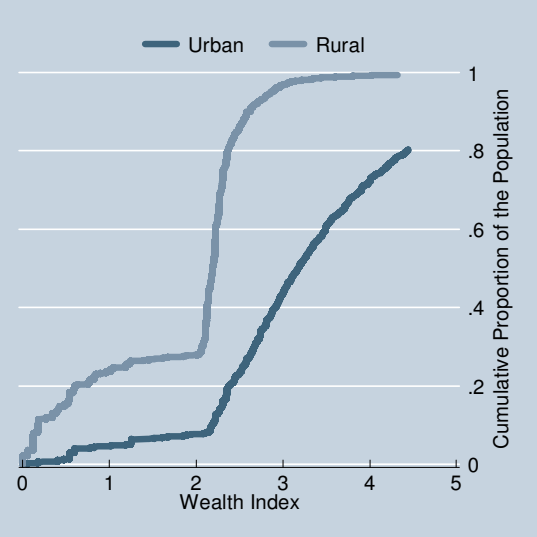


Figure 2.5 – CDFs by region (consumption aggregate)

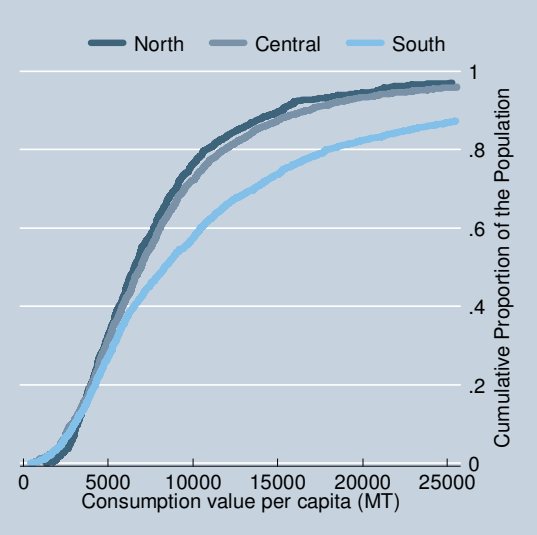


Figure 2.6 – CDFs by region (wealth index)

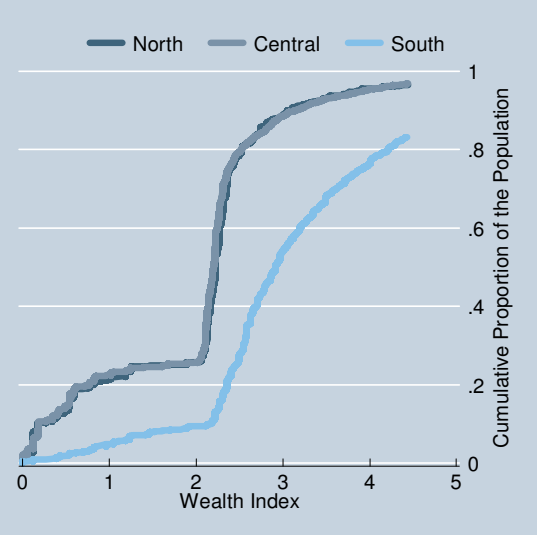


Figure 2.7– CDFs by gender of head (consumption aggregate)

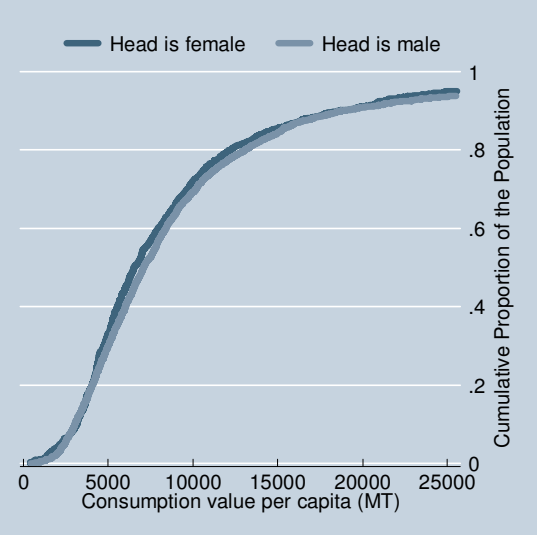


Figure 2.8 – CDFs by gender of head (wealth index)

Table 2.2 also suggests that individuals living in households headed by somebody who at least completed the primary level of education were less poor (33.01%) than their less educated counterparts (57.45%). A similar analysis employing the wealth index supports this fact. Actually, the wealth index shows a wider gap between these two categories of household heads. Nonetheless, part of this gap might be due to the fact that most of the less educated heads lived in rural areas and the wealth index tends to be biased against rural households, as mentioned. Concerning the welfare dominance approach, both the CDFs using consumption aggregate and the wealth index indicate that these estimates are robust to the choice of the poverty line, since they do not cross (observe Figure 2.9 and Figure 2.10).

With respect to the provincial poverty profile, the money-metric estimates suggest that in the period 2002-2003, on the one extreme, Inhambane was the poorest province with 83.54% of the population below the poverty line. On the other extreme, Sofala was the least poor province, with only 35.66% of incidence of poverty. The provincial ranking suggested by this measure, however, is not the same as the one suggested by the asset index. Making use of the wealth index it was found that Zambezia was the poorest province, with 81.06% of the population below the poverty line. Maputo Province and Maputo City were the least poor regions, with only 4.26% and 0.03% of the population below the asset poverty line, respectively.

Figure 2.9 – CDFs by education of head (consumption aggregate)

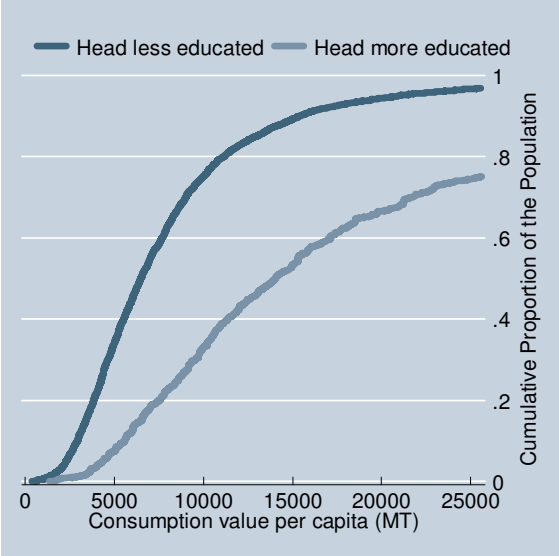


Figure 2.10 – CDFs by education of head (wealth index)

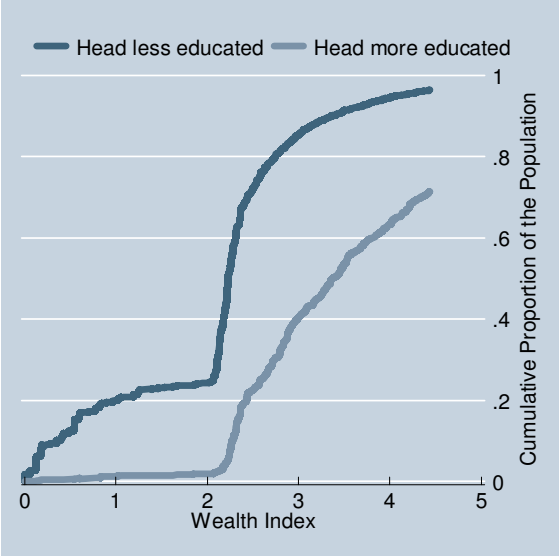


Figure 2.11 – CDFs by provinces (consumption aggregate)

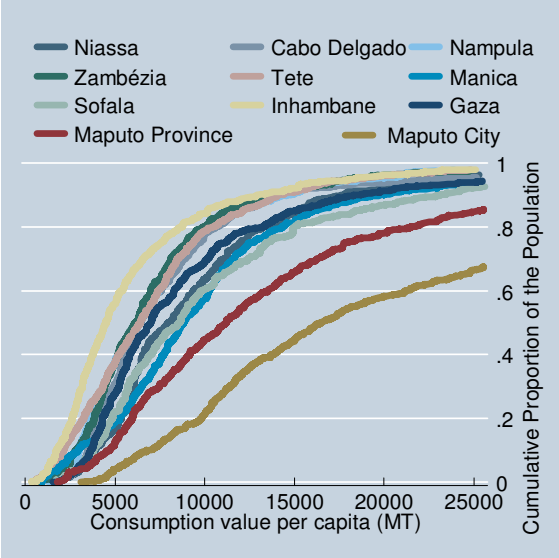
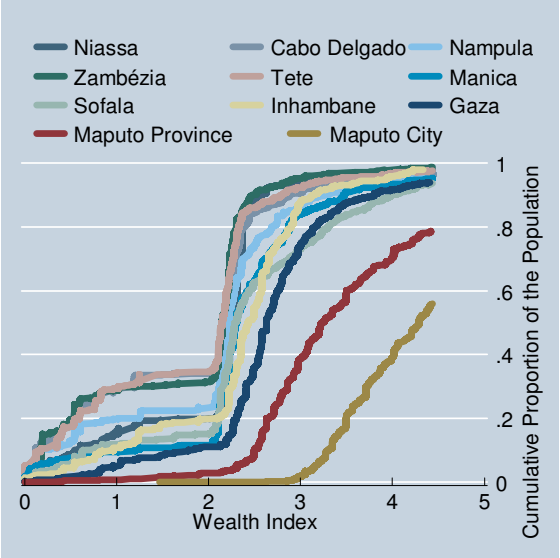


Figure 2.12 – CDFs by provinces (wealth index)



The welfare dominance approach suggests that Maputo Province and Maputo City were by far the least poor provinces of Mozambique, irrespective of the poverty line chosen (see Figure 2.11). The CDFs corresponding to these two provinces are below the CDFs corresponding to the remainder of the provinces. The dominance approach also suggests that Inhambane was one of the poorest provinces, since its CDF is above the other ones except Zambezia's and Tete's CDFs (see also Figure 2.13 and Figure 2.14 on page 56). With respect to the remainder of the provinces' poverty

estimates, this approach suggests that there is no stochastic poverty dominance since the CDFs cross. But if Zambezia is excluded, Tete is the poorest province (see Figure 2.15 to Figure 2.18 on page 57).

The wealth index's poverty estimates had already suggested that the Southern region of Mozambique, in particular Maputo Province and Maputo City, were the least poor in the period 2002-2003. These results are robust to the choice of the poverty line (see Figure 2.12). Concerning the remainder of the provinces Figure 2.19 and Figure 2.20 on page 57 indicate that there is no poverty dominance. But if Sofala is left aside, Gaza displays first-order stochastic dominance over the remaining provinces since their CDFs are above Gaza's CDF.

Lastly, Table 2.2 also reports money-metric poverty estimates using poverty lines which do not vary across spatial domains, in contrast to DNPO et al.'s (2004) variable poverty lines. One poverty line is set to 7,598.83 MT per person per day, for all regions of the country. This is the consumption value that leads to an estimate of poverty incidence equal to 54.36%, the headcount poverty rate this study found and which was very similar to DNPO et al.'s (2004) result. The other poverty line was set to 1US$ per person per day, again for all regions. The results of the estimates suggest that DNPO et al.'s (2004) poverty estimates are extremely sensitive to the assumptions regarding the design of their variable poverty lines. For instance, both the new poverty lines suggested that, in 2002-2003, the gap between rural and urban areas was indeed wider, as suggested by the wealth index. They also suggested that the South was the least poor region. Lastly, just as the wealth index indicated that Zambezia[20], on the one extreme, was the poorest province, and on the other, that Maputo Province and Maputo City were the least poor, these new poverty lines led to very similar conclusions.

[20] Making use of the 7,598.83 MT and the 1US$ poverty lines Zambezia was the second and fourth poorest province, respectively.

Figure 2.13 – CDFs for selected provinces (i) (consumption aggregate)

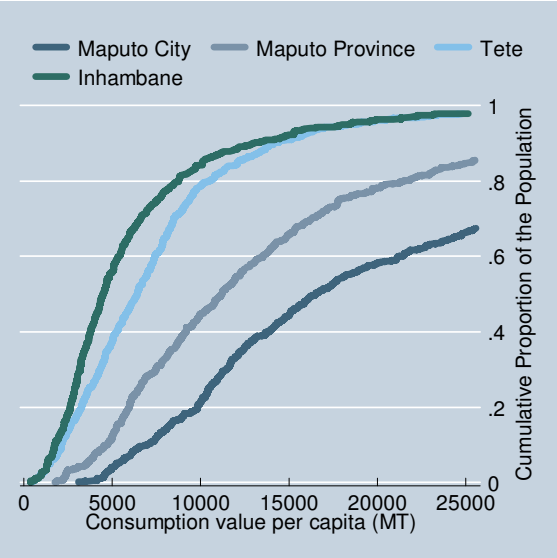


Figure 2.14 – CDFs for selected provinces (ii) (consumption aggregate)

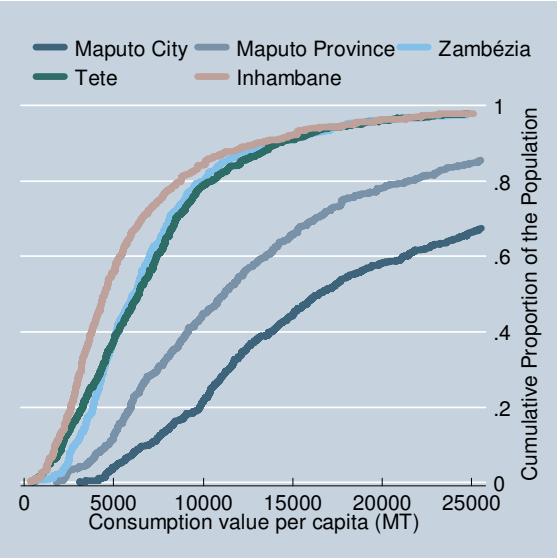


Figure 2.15 – CDFs for selected provinces (iii) (consumption aggregate)

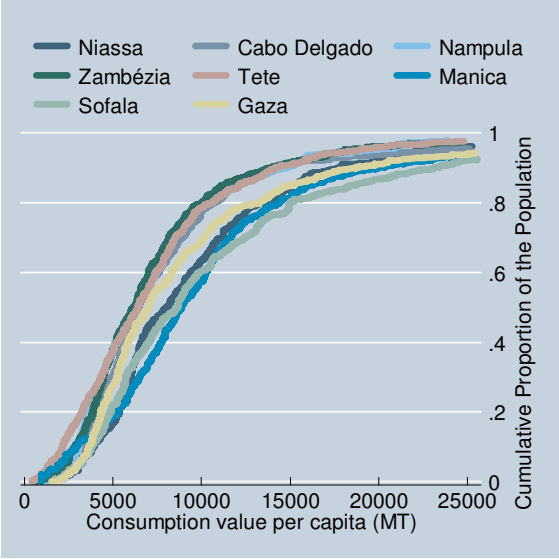


Figure 2.16 – CDFs for selected provinces (iv) (consumption aggregate)

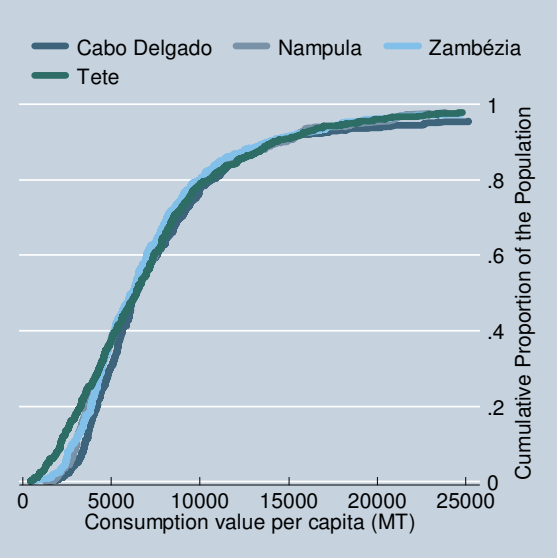


Figure 2.17 – CDFs for selected provinces (v) (consumption aggregate)

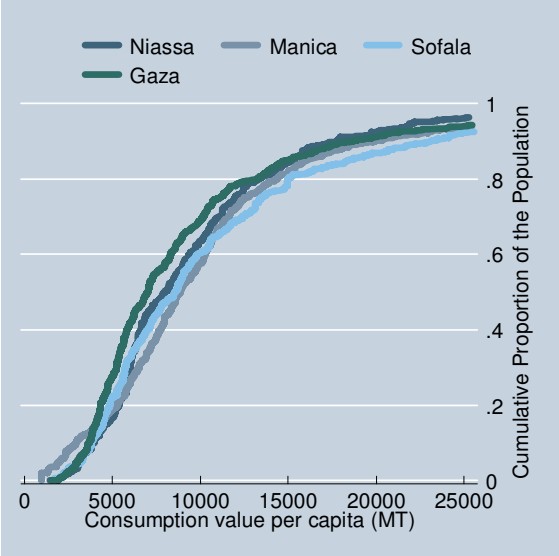


Figure 2.18 – CDFs for selected provinces (vi) (consumption aggregate)

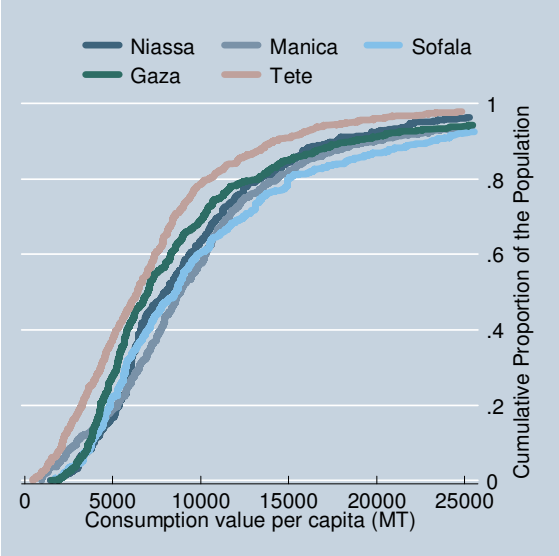


Figure 2.19 – CDFs for selected provinces (i) (wealth index)

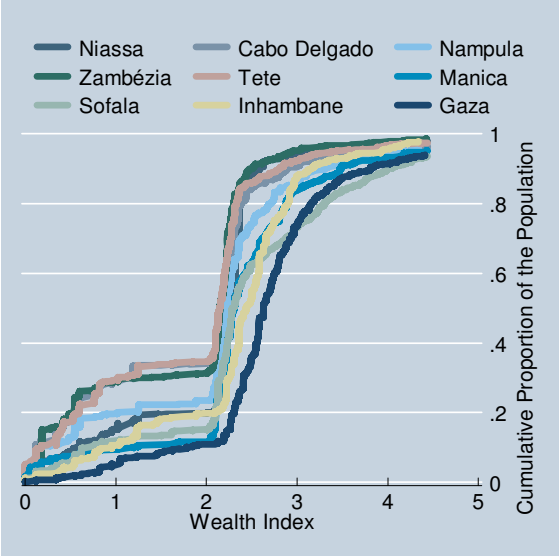
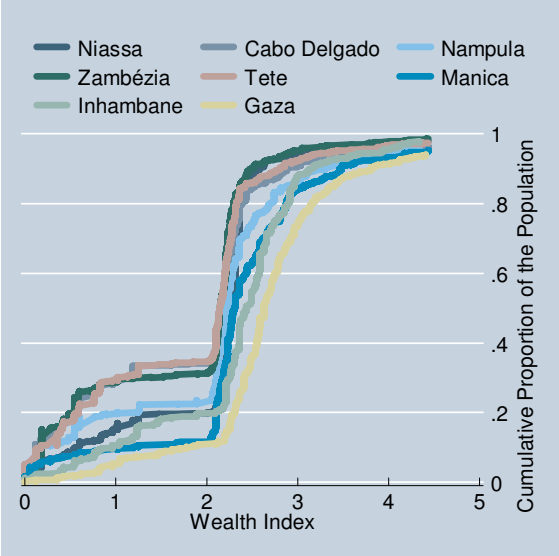


Figure 2.20 – CDFs for selected provinces (ii) (wealth index)



2.4.1.2 Poverty markers

The analysis of poverty markers is of great importance, particularly for poverty reduction policy purposes, as it permits to identify the proximate correlates of poverty. This study followed a specification similar to the one presented in Maximiano, Arndt & Simler (2005). Formally, the model is described as:

$$c_j = \beta' x_j + \mu_j, \tag{4}$$

where c_j stands for logarithm of real consumption per capita of household j, or for wealth index score of household j, depending on the focus of the analysis; x_j is a group of household and individual characteristics; and μ_j is a random error term. The wealth index was standardised to have mean zero and standard deviation of one.

Table 2.3 on page 59 reports the results of two ordinary least squares (OLS) regressions modelling the effects household and individual characteristics have on poverty in Mozambique. The potential poverty markers include age, household size, age-dependency ratio, and health status of the household head within two weeks prior to the interview. With respect to expected results, first, assuming that older individuals accumulated more skills and on-the-job experience, and that they need more calories than the younger ones, one expects that age will have a positive impact on consumption per capita. By the same reasons one expects that it will affect wealth positively.

Second, concerning the consumption aggregate, in bigger households 'the pie is shared among many individuals'. Therefore, bigger households tend to be poorer on average. In contrast, it can be argued that bigger households are more likely to own an asset than smaller households, if at least a household member has a source of income. It is then expected that household size will impact negatively on consumption per capita but positively on the asset index.

Third, households with a higher number of dependents, on average, are expected to be poorer. According to the Mozambican National Statistics Institute (*Instituto Nacional de Estatística* – INE), the economically active population include those aged 15 years and above, with no upper limit. The approach followed to create the age-dependency ratio was to divide the number of individuals aged below 15 years by the household size, within each household.

Fourth, on average sick household heads are expected to have difficulties in providing for their respective families. It is thus expected that this variable will enter the regressions negatively.

Fifth, educated household heads are likely to be in a better position to provide for their families than uneducated ones. One therefore expects that people living in households whose head has at least primary education completed will be less poor than those living in households headed by less educated people.

According to Table 2.3, firstly, people living in households with more dependents tended to be poorer on average. Controlling for other household and individual characteristics that have a

bearing on living standards, an increase by one unit in the age-dependency ratio leads to a decline in consumption per capita by 0.88% and to a decline in the wealth index by 0.38 standard deviations. Both coefficients are statistically significant at the 1% level of significance.

Secondly, employing the consumption aggregate, age has no effect on consumption per capita, *ceteris paribus*. The age coefficient is statistically insignificant. Nevertheless, contrary to expected, the wealth index suggests a negative relationship between age and wealth.

Table 2.3 – Poverty Markers

Dependent Variable / Correlates	Log of Consumption per capita	Wealth Index
Constant	19.547***	0.997***
Age-dependency ratio	-0.888***	-0.382***
Age	0.000	-0.001***
Household size	-0.037***	0.033***
Rural	-0.243***	-0.619***
Female	-0.009	-0.012*
Female head	-0.044***	-0.117***
Head was sick in the previous 2 weeks	-0.017**	-0.022**
Head completed at least the primary education level	0.571***	0.566***
Niassa	-0.631***	-1.148***
Cabo Delgado	-0.841***	-1.192***
Nampula	-0.866***	-1.105***
Zambezia	-0.737***	-1.061***
Tete	-0.844***	-1.069***
Manica	-0.500***	-0.951***
Sofala	-0.465***	-0.846***
Inhambane	-1.047***	-0.801***
Gaza	-0.545***	-0.607***
Maputo Province	-0.405***	-0.338***
R-Squared	0.392	0.520
Number of observations	44,047	44,047

Notes: Regressions based on robust standard errors; * Significant at 10% level; ** Significant at 5% level; * Significant at 1% level**

Thirdly, making use of the consumption aggregate, bigger households tended to be poorer, *ceteris paribus*. The coefficient on household size is negative and statistically significant at 1% level of significance. But employing the wealth index it is concluded that bigger households tended to be less poor, *ceteris paribus*, as expected. Its coefficient is positive and statistically significant at 1% level of significance.

Being female rather than male (for the asset index only), living in a female headed household rather than living in a male headed household, living in rural areas rather than in urban, living in provinces other than Maputo City and living in households whose head had less than the primary level of

education are features associated with poverty. In other words, individuals with these characteristics tend to be poorer, on average. The coefficients on these dummies are negative and statistically significant at 1% level of significance. Lastly, making use of the consumption aggregate and of the wealth index, the R-squared suggests that the models explain 39% and 52% of the variation in poverty, respectively.

It is pertinent to point out that two poverty markers have a relatively large bearing on observed differences in poverty. The first is living in a household whose head is educated (which lowers the chances of poverty) and the second is living in rural locations (which increases the chances of falling into poverty). As mentioned before, Fox et al. (2005) had also found that the education of the head (in particular in urban areas) was the most important correlate of poverty.

2.4.2 Inequality Analysis

This subsection reports the inequality estimates making use of the Gini and Theil indices. It also makes use of the Lorenz dominance approach. The dominance approach is important to ensure that the rankings suggested by the inequality indices are robust and not just a result of an artefact of the particular inequality index used. As mentioned, if a Lorenz curve of a certain distribution lies above, and never below the Lorenz curve of another distribution, it means that the former distribution Lorenz dominates the latter, and therefore it is less unequal than the latter distribution (Fields, 2000, p. 19). In contrast, if two or more Lorenz curves cross, inequality rankings can no longer be made, and therefore nothing can be said about the inequality rankings between those distributions.

The first two columns of Table 2.4 on page 61 report the inequality estimates for the consumption aggregate and the last two report the estimates for the wealth index. Firstly, at the national level, making use of the money-metric indicator the Gini and Theil indices suggest that in the period 2002-2003 inequality was approximately 0.48 and 0.39, respectively. Employing the wealth index the figures change to 0.30 and 0.42, respectively. Secondly, applying the consumption aggregate the inequality indices suggest that urban regions were more unequal than the rural. This conclusion is corroborated by the Lorenz dominance approach which suggests that rural regions Lorenz-dominated the urban (see Figure 2.21 on page 62). In contrast, employing the wealth index, the inequality indices indicate that rural regions were more unequal than the urban. Figure 2.22, however, suggests that there is no Lorenz dominance since the curves cross.

Table 2.4 also reports the decomposition of the Theil index into ‘within-group’ and ‘between-group’ inequality, by location (rural and urban). Both dimensions suggest that inequality within the

locations was higher than inequality between them. 79% and 88% of inequality in the distribution of the consumption aggregate and of the wealth index, in that order, was within-locations inequality. The other 21% and 12%, respectively, corresponded to between-locations inequality.

Table 2.4 – Inequality Estimates

	Consumption Aggregate		Wealth Index	
	Gini Index	Theil Index	Gini Index	Theil Index
Within-group by location	-	0.31	-	0.37
Between-groups by location	-	0.08	-	0.04
Within-group by province	-	0.30	-	0.38
Between-groups by province	-	0.09	-	0.04
Within-group by region	-	0.36	-	0.40
Between-groups by region	-	0.03	-	0.02
Within-group by education of head	-	0.30	-	0.40
Between-groups by education of head	-	0.08	-	0.02
National	0.48	0.39	0.30	0.42
By location				
Urban	0.52	0.47	0.23	0.12
Rural	0.36	0.23	0.26	0.49
By education of head				
Head has less than primary level of education completed	0.40	0.27	0.29	0.44
Head has at least primary level of education completed	0.55	0.52	0.22	0.10
By region				
North	0.41	0.29	0.29	0.45
Centre	0.42	0.30	0.29	0.50
South	0.56	0.55	0.23	0.16
By province				
Niassa	0.38	0.24	0.21	0.28
Cabo Delgado	0.50	0.43	0.34	0.55
Nampula	0.36	0.21	0.28	0.45
Zambezia	0.37	0.24	0.30	0.46
Tete	0.41	0.30	0.33	0.83
Manica	0.39	0.28	0.21	0.36
Sofala	0.47	0.37	0.26	0.31
Inhambane	0.45	0.36	0.20	0.24
Gaza	0.43	0.31	0.19	0.12
Maputo Province	0.46	0.36	0.18	0.06
Maputo City	0.54	0.51	0.16	0.04

Also from looking at Table 2.4 one concludes that, for the consumption aggregate, inequality was higher in the South. It was followed by the Centre and then by the North. Nevertheless, Figure 2.23 exposes the fact that inequality rankings between the North and the Centre cannot be made given that the Lorenz curves cross. But the South was beyond doubt the most unequal region.

Table 2.4 also reports inequality figures by regions for the wealth index. But contrary to what was found for the consumption aggregate, it suggests that the South was the least unequal region. The North and the Centre had roughly the same inequality levels. Figure 2.24 confirms this result. The Lorenz curve for the Southern region lies above the other two. The same table indicates that 92% and 95% of the inequality in the distribution of the consumption aggregate and of the wealth index, respectively, was within-regions inequality. The remaining 8% and 5%, in the order mentioned, was between-regions inequality.

Figure 2.21 – Lorenz curves by location (consumption aggregation)

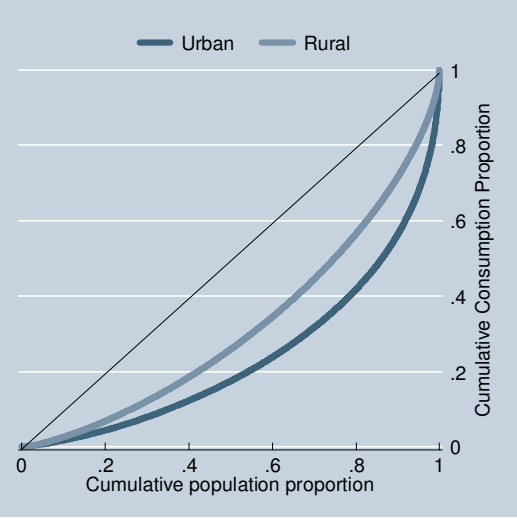


Figure 2.22 – Lorenz curves by location (wealth index)

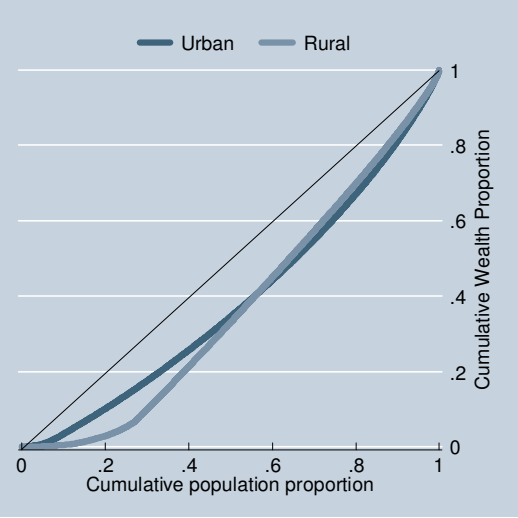


Figure 2.23 – Lorenz curves by region (consumption aggregation)

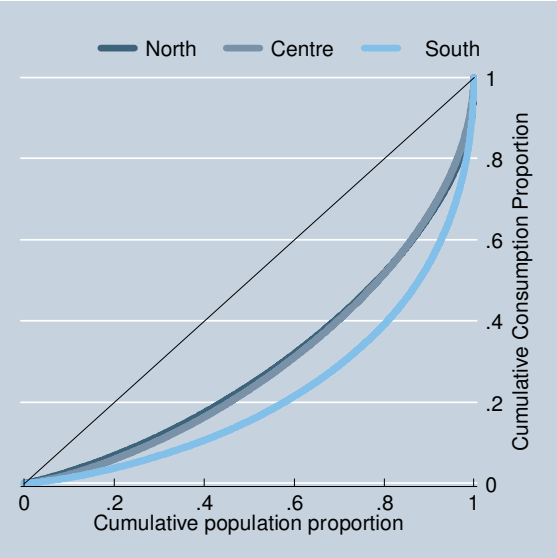
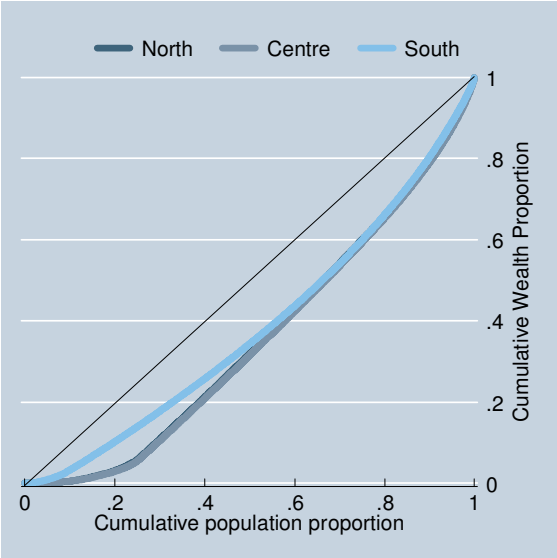


Figure 2.24 – Lorenz curves by region (wealth index)



With respect to inequality by education of head, the money-metric indicator suggests that individuals living in households with less educated heads were less unequal than those living with more educated heads. Figure 2.25 confirms that the former group Lorenz-dominated the latter. Again, the wealth index estimates contrast with the money-metric ones. Making use of the wealth index individuals living in households with heads who completed at least the primary level of education were less unequal than their less educated counterparts. As clearly stated in Figure 2.26, the 'educated' distribution Lorenz dominates the 'non-educated' one. Once again, making use of both indicators, the inequality decomposition by education of head suggests that inequality within groups explained most of the existing inequality.

Lastly, Table 2.4 reports inequality estimates across provinces. Making use of the consumption aggregate, on the one extreme, Maputo City was the most unequal province in the period 2002-2003. On the other extreme, Nampula was the least unequal province. Though not very clear in Figure 2.27, Figure 2.29 shows that Nampula was indeed less unequal than Maputo Province and Maputo City, since their Lorenz curves do not cross. Actually, the figures confirm that these provinces were the lower and upper extremes, in that order. But nothing can be said about the inequality rankings between Maputo City and Cabo Delgado (see Figure 2.30), because the Lorenz curves cross.

Figure 2.25 – Lorenz curves by education of head (consumption aggregate)

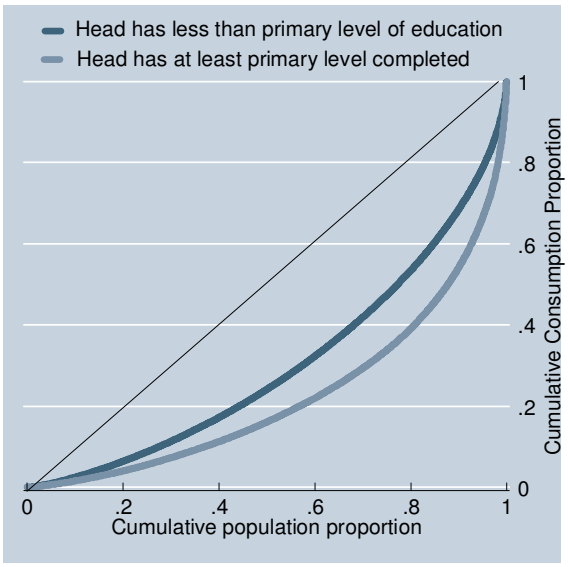
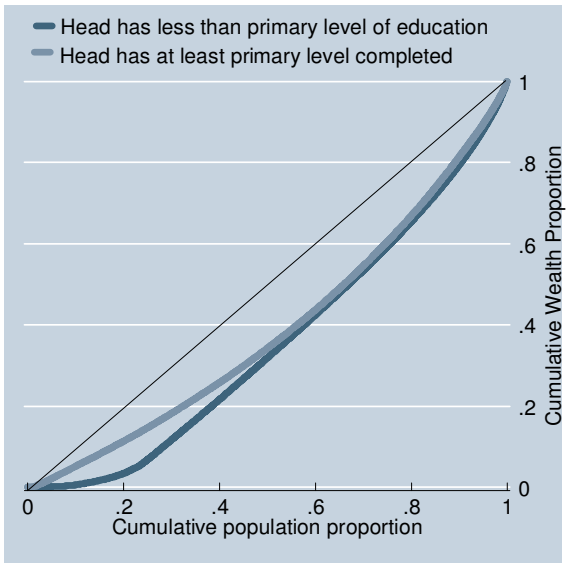


Figure 2.26 – Lorenz curves by education of head (wealth index)



Making use of the wealth index, though, Maputo City was the least unequal province and Tete and Cabo Delgado the most unequal provinces[21]. The Lorenz dominance approach suggests that the lower bound was Maputo City and the upper bounds were both Cabo Delgado and Tete, since the Lorenz curves for the latter two provinces trace one another and are below Maputo City's Lorenz curve, without crossing it (see Figure 2.28 and Figure 2.31).

Concerning the inequality decomposition by province, Table 2.4 indicates that within-group inequality was much higher than between-groups inequality. Both dimensions suggest that inequality within provinces was higher than inequality between provinces. 77% and 90% of inequality in the distribution of the consumption aggregate and of the wealth index, in that order, was within-province inequality. The other 23% and 10%, respectively, was between-provinces inequality, meaning that Lorenz dominance across provinces was not the problem but income differences within provinces.

[21] Applying the Gini coefficient Cabo Delgado was the most unequal province, while employing the Theil index Tete was the most unequal.

Figure 2.27 – Lorenz curves by province (consumption aggregate)

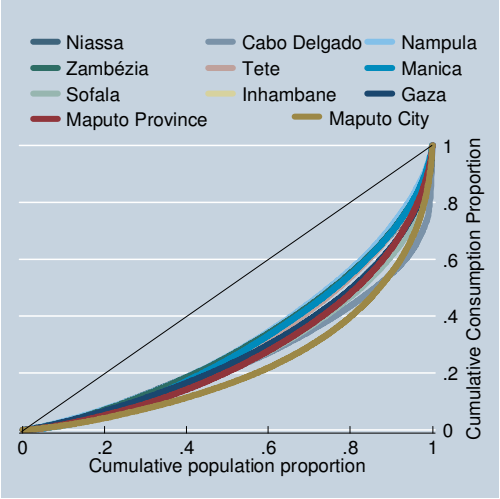


Figure 2.28 – Lorenz curves by province (wealth index)

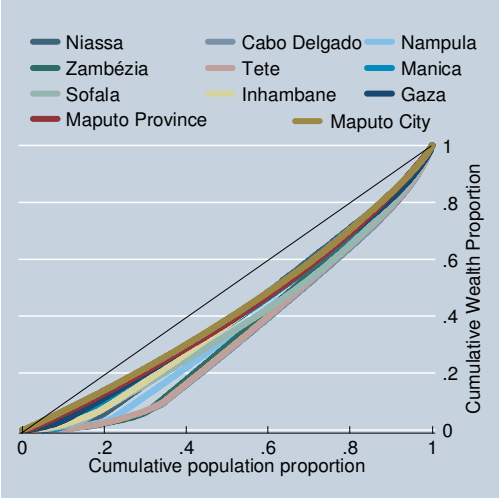


Figure 2.29 – Lorenz curves for selected provinces (i) (consumption aggregate)

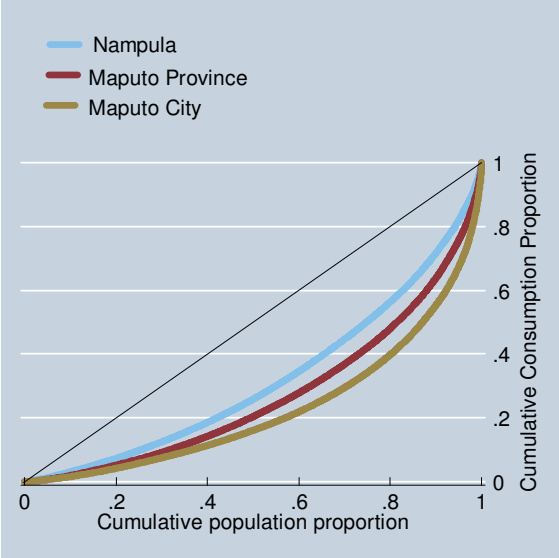


Figure 2.30 - Lorenz curves for selected provinces (ii) (consumption aggregate)

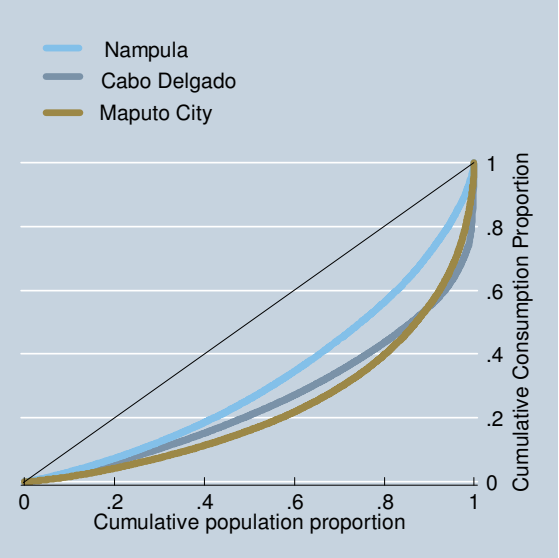
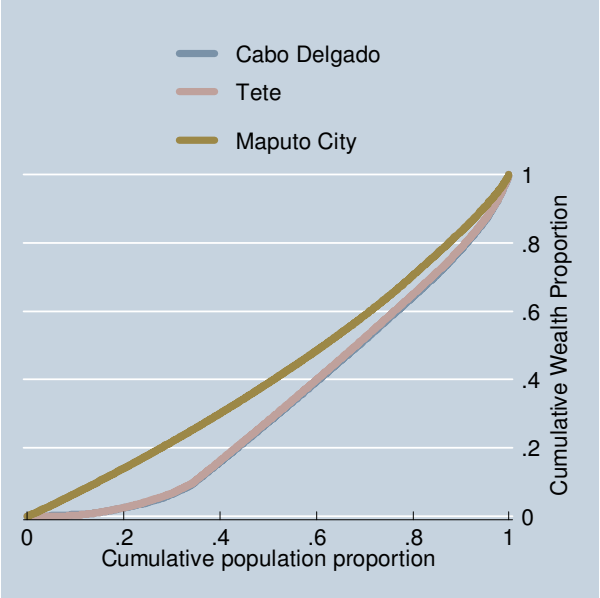


Figure 2.31 – Lorenz curves for selected provinces (wealth index)



2.4.3 Discussion

This subsection gives the building blocks that will allow coming to a final conclusion with regards to the thesis statement presented in the introductory section of this work. Figure 2.32 to Figure 2.37 on page 68 map the headcount poverty ratios and the Gini coefficients (across Mozambican provinces), and contain information similar to that presented in Table 2.2 and Table 2.4. These figures give a more illustrative picture of the welfare status of the Mozambican population in the period 2002-2003. Darker colours imply higher levels of poverty or inequality, while lighter colours mean the opposite. The elaboration of these maps required the use of a software for interactive data visualisation and mapping named StatPlanet.

The first point to note is the fact that basing the analysis on the consumption aggregate and the DNPO et al.'s (2004) variable poverty lines this study got to results very similar to the Mozambican official poverty estimates. The second point concerns poverty by location. Following the DNPO et al.'s (2004) approach this study did reach similar estimates. But the wealth index's results suggest that rural regions were a great deal poorer than the urban ones. This contrasts to DNPO et al.'s (2004) work which concluded that rural and urban poverty levels were roughly similar. They did not find dominance across locations, while in this chapter, employing both indicators, it was found that urban locations displayed first-order stochastic dominance over the rural. However, since the wealth index tended to be biased against rural households, the rural-urban gap should be viewed cautiously. Despite this issue, the multivariate regression analysis presented in Table 2.3 indicated that controlling for other factors that have a bearing on living standards, people living in rural

regions were on average poorer than their urban counterparts. An immediate policy implication is that anti-poverty programmes should explicitly target rural areas.

Third, the money-metric approach suggests that the Southern region of Mozambique was the poorest one. Making use of the wealth index, Figure 2.33 suggests that this is inaccurate. The wealth index suggests that the South (Gaza, Inhambane, Maputo Province, and Maputo City) was the least poor region. The welfare dominance approach corroborated this conclusion applying both the consumption aggregate and the wealth index. Similar results are also found when the poverty lines are set at 7,598.83 MT and 1US$ per person per day – the South was the least poor region in the period 2002-2003. This suggests that DNPO et al.'s (2004) poverty estimates were very sensitive to the assumptions regarding the design of the poverty lines. Programmes aimed at poverty alleviation should therefore be less targeted on the Southern region of Mozambique and concentrate more on the Central and Northern parts of the country.

Fourth, the welfare dominance approach suggests that the poverty estimates by gender of household head are not robust to the choice of the poverty line. But the incidence of poverty across genders is greater for women than for men, using both the consumption aggregate and the wealth index. Previous works did not analyse the incidence of poverty across genders using this survey, an aspect of great importance, taking into account that females form the major part of the Mozambican population.

Fifth, both indicators suggest that education matters for poverty alleviation. People living in households whose heads were less educated tended to be poorer. If education policies do not ensure that the poor get access to quality education they will perpetuate the existing vicious cycle of poverty and inequality in Mozambique, rather than transforming them. The expansion of adult education programmes to target poorly educated household heads is likely to contribute to the reduction of the incidence of poverty.

Figure 2.32 – Poverty mapping (Consumption aggregate)

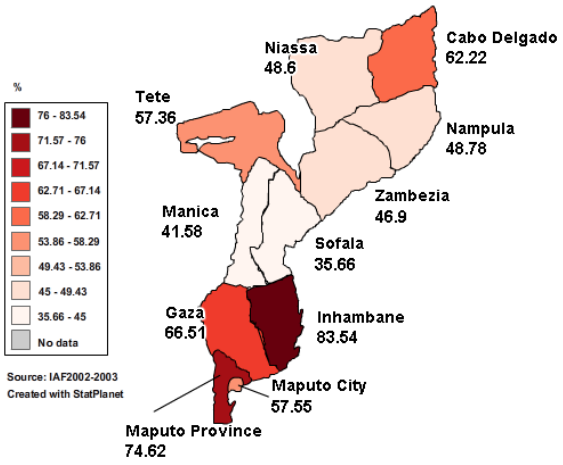


Figure 2.33 – Poverty mapping (Wealth index)

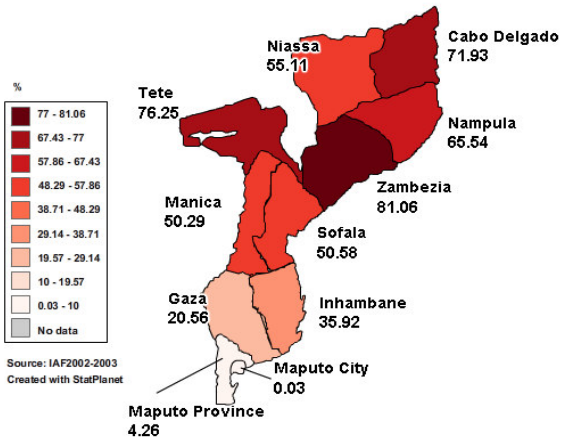


Figure 2.34 – Poverty mapping for the consumption aggregate with the poverty line set at 7,598.83 MT per capita per day

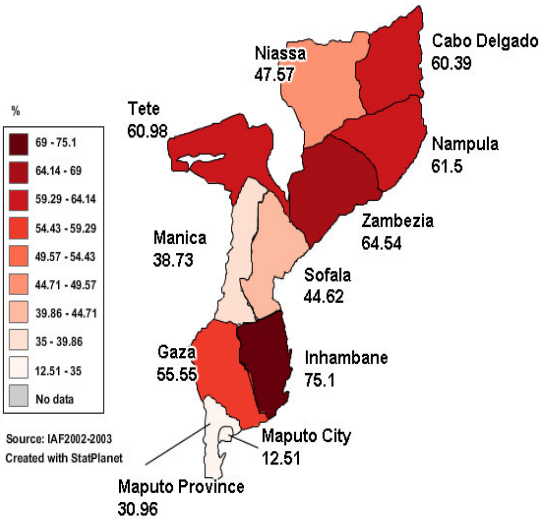


Figure 2.35 – Poverty mapping for the consumption aggregate with the poverty line set at 1US$ per capita per day

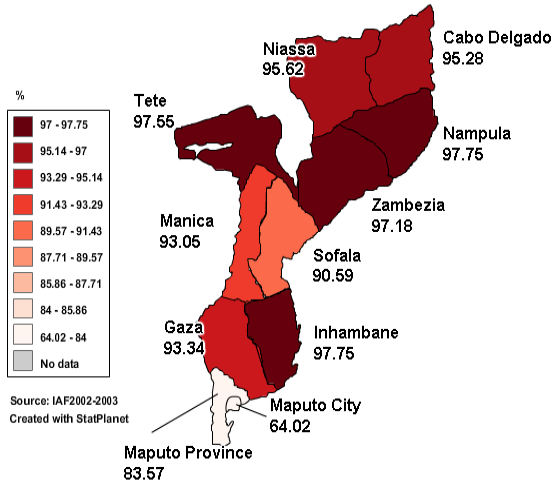


Figure 2.36 – Gini inequality mapping (Consumption aggregate)

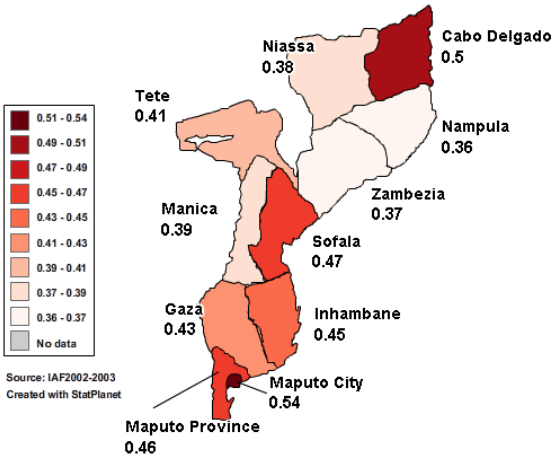
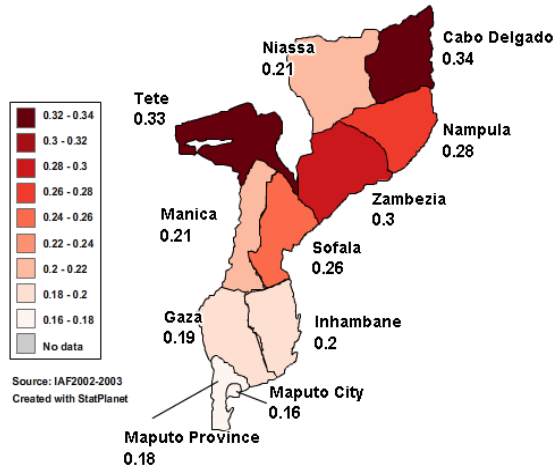


Figure 2.37 – Gini inequality mapping (Wealth index)



Sixth, a fact that is clear among all the approaches to measure poverty except the DNPO et al.'s (2004) report is that Maputo City and Maputo Province were the least poor provinces in the studied period (see Figure 2.33 to Figure 2.35). DNPO et al. (2004) did not reach the same conclusion. With respect to the other extreme, the wealth index suggests that Zambezia was the poorest province. Making use of the consumption aggregate and the poverty lines derived in this chapter it was found that Zambezia was one of the poorest provinces in the period 2002-2003. Again, DNPO et al. (2004) failed to reach this conclusion. As discussed in Da Maia & Van der Berg (2010), the problem lies in the 13 different poverty lines derived by DNPO et al. (2004) that were adjusted to reflect local diet and prices. Such adjustments ended up overstating poverty in Maputo City, at least as far as one can determine from other measures of well-being. Another way to get to similar conclusions is by looking at Table 2.3 on page 59, which reports the results of the multivariate regressions. The table shows that for both indicators an average individual living in Zambezia, compared to one living in Maputo City, was on average the fourth poorest after Inhambane, Cabo Delgado and Tete, *ceteris paribus*.

The view that Zambezia is one of the poorest provinces of Mozambique (if not the poorest one) is also shared by other analysts. For instance, Chichava (2007, pp. 15-17) argued that since most of the population from Zambezia supports *Resistência Nacional Moçambicana* (RENAMO), the biggest opposition party in Mozambique, the government does not give appropriate support to households living in this province. The transportation and road systems are so debilitated that most durable private assets reach the final consumers at very high prices, therefore making them unaffordable to most households. Most of the population lacks access to clean water or a house with a proper sanitation system. Indeed, the wealth index suggests that the government was not at the time doing much for this province in terms of making available basic socio-economic infrastructures. On the other hand, the province is considered 'the barn of the nation' since it is a major supplier of agricultural products to the rest of the country.

As mentioned in the introductory chapter of this thesis, Zambezia is the second most populated province of Mozambique, with a total population of 3.8 million inhabitants. This suggests that a large share of the national poverty is likely to be attributable to Zambezia. Concentrating anti-poverty programmes in Zambezia (in particular in its rural parts) is the most obvious policy implication suggested by this part of the empirical work. Reducing poverty in Zambezia first has the potential of quickly pulling down the national poverty levels, given the province's high poverty share.

Seventh, employing the consumption aggregate it was found that inequality at the national level was much higher than Fox et al. (2005) and the World Bank (2007) suggested. But using the wealth index the figures were lower.

Eighth, this study reached the same conclusion as Fox et al. (2005) did with respect to inequality across locations: urban areas were more unequal than the rural ones. The Lorenz dominance approach supports this finding, at least for the money-metric indicator. But making use of the wealth index there is no Lorenz dominance and therefore, nothing can be said about the inequality rankings between rural and urban locations.

Ninth, looking at Figure 2.23 on page 63 and Figure 2.36 on page 68, the impression is that the South was the most unequal region in the studied period. Applying the wealth index, this impression changes dramatically. One instead concludes that the South was the least unequal region, in particular Maputo City and Maputo Province (see Figure 2.24 and Figure 2.37).

Lastly, 'within-group' inequality across locations, provinces, regions and education of head was always much higher for the case of the wealth index. The non-money-metric approach suggests that most of the inequality comes from within the different groups, as the groups were on average similar. That could be the reason why making use of the wealth index in most cases there was no Lorenz dominance.

2.5 Conclusion

This study argued that poverty and inequality were not very well understood in the Mozambican context. The reason for this is the convention followed by most researchers of focusing their analyses on the money-metric dimension only. Researchers estimate poverty levels by calculating the proportion of the population that fails to reach a poverty line, very often set in terms of income or consumption expenditures only. This work argued that this approach was not enough to understand welfare, and therefore suggested the construction of an asset index that could be used to proxy for the living standards of the population. This would introduce other dimensions to the analysis of poverty and inequality. Welfare analysis using such an index complements (but does not substitute) the money-metric approach and improves the profile of poverty and inequality.

By using the proposed asset index this study found a poverty profile of Mozambique different from the one found when the conventional money-metric approach with variable poverty lines is applied, as had been done in earlier works. For the cases where the rankings were qualitatively similar the wealth index showed that the gap between the groups was higher. Such gaps, however, should be

viewed with caution since the wealth index tended to be biased against rural households. Where surveys are used to inform government policy, as they should, targeting of state interventions often responds to poverty profiles. The poverty profile derived from money-metric provincially-differentiated poverty lines would lead to perverse targeting that would strengthen rather than reduce the existing urban bias. It is in that sense extremely bad for the poor.

This chapter offered a detailed method for compiling a consumption aggregate for welfare analysis. This is important since there is a gap between theory and practice on this matter, notably for the case of Mozambique. There is plenty of material explaining, theoretically and generally, how to compile such a measure, but there is no material explaining how to measure consumption in surveys such as the Mozambican Income and Expenditure Household Survey, which differs a bit from the well-known World Bank Living Standards Measurement Surveys. Future income and expenditure household surveys should dedicate special attention in providing metadata detailing the steps followed to construct the living standard indicators. Even for already existing surveys the National Statistics Office (INE) could make an extra effort and make available to the wide public the relevant metadata.

Another contribution of this chapter was the use of the wealth index approach on a Mozambican data set. So far no previous study did this. By analysing other dimensions of welfare, poverty and inequality reduction policies will target the intended groups with greater precision. This study therefore suggests that every money-metric analysis of poverty and inequality should be thoroughly complemented by an analysis using a socio-economic wealth index. In this way, since other dimensions of poverty and inequality are also captured, these phenomena will be better understood and therefore better dealt with.

Identifying the most vulnerable groups in a society by creating a national poverty (and inequality) profile is only the initial step in designing policy interventions to curb poverty and inequality. Identifying the correlates of poverty is also a step in the right direction. More than these we need policy levers that could be used to help eradicate poverty and inequality. Many people associate schooling with the eradication of poverty and inequality. In this chapter, for instance, it was found that people living in households whose head completed at least the primary level of education were less poor on average. It is thus important to understand the mechanisms through which schooling is associated with better living standards, and the labour market seems to be a promising transmission mechanism.

As an increasing proportion of the population gets more and more access to education, however, schooling measured by educational attainment only (i.e., the number of years of education completed) becomes insufficient to curb poverty and inequality. Those in better schools have better chances of getting the best jobs and thus better chances to escape poverty, while those in low quality schools are less likely to get good jobs. If schooling is associated with poverty alleviation through the labour market, and the labour market rewards those coming from good schools better, it is also important to understand what makes a school good, i.e., what the correlates of pupil cognitive achievement are. These topics offer space for further research, and will be dealt with in the subsequent chapters.

Chapter 3

Education and Labour Market Outcomes in Mozambique

> *"Prospects for reducing poverty, narrowing extreme inequalities and improving public health are heavily influenced by what happens in education. Progress towards the equalisation of opportunity in education is one of the most important conditions for overcoming social injustice and reducing social disparities in any country"* (UNESCO, 2009, p. 24).

3.1 Chapter Introduction

An important linkage of education and poverty alleviation is through the labour market. Schooling is intrinsically linked to labour force participation. It shares a positive association with the probability of employment. More educated people are more likely to participate and to get the best paid jobs available (Bhorat & McCord, 2003, p. 135). Education is also associated with better command of earnings. Human capital theorists advocate schooling and post-schooling investments as sources of improvement of lifetime earnings (Mincer, 1974). In summary, in particular in sub-Saharan countries, "education is often seen as the main policy instrument in the fight against poverty because it may help individuals access better jobs and thus raise their labour incomes" (Kuepie, Nordman & Roubaud, 2007, p. 2). Further and as stated in the UNESCO quote above, if we want to deal with inequality in developing and poor countries such as Mozambique, we should start with education, by making access and quality more equal.

In developing and poor countries such as Mozambique many people work outside the wage sector. A large part of the working-age population is engaged in self-employment activities and unpaid family work, particularly subsistence agriculture and informal activities. A direct implication of this feature is that different labour market segments might also differ with respect to the determinants of entry and earnings. For instance, Glick & Sahn (1997), employing multinomial logit models and using Guinean data, found that the determinants of entry into the various sectors of the labour market were quite different. For both men and women, more education reduced the likelihood of being self-employed while it strongly increased the likelihood of working in the public sector. In Ghana, Glewwe (1991) found that education was positively associated with wage employment, and among the wage employed, the better educated were more likely to be in the public sector than in the private sector.

The main data set used in this chapter contains information on the different segments of the labour market the Mozambican working-age population is engaged in. A plausible view of the labour market in Mozambique is the existence of six distinct employment segments. These are wage

employment in the private sector, wage employment in the public sector, self-employment, unpaid family work, unemployment and non-participation. It is therefore not sufficient to state that education is associated with better chances of getting a job or with better earnings. The labour market is too segmented and analysis should be done for each of the segments.

The research problem this chapter tackles is the following: if we consider the whole range of different labour market segments that exist in developing economies such as this, there is the potential for education to impact differently on a person's probability of getting a job and on his or her earnings. Therefore, employing, say, a probit or logit model that links schooling to the probability of getting a job and an ordinary least squares (OLS) earnings function that links schooling to earnings, for the whole labour force and thus without accounting for the segmentation existing in the labour market, is likely to be misleading, thus leading to very crude and wrong conclusions.

Therefore, the purposes of this chapter are twofold. First, it aims to determine the impact schooling and other individual characteristics have on the individual probability of falling into each one of the six employment states mentioned above. Given the multitude of "options[22]" open to the working-age population, this part of the study employs a multinomial logit model. The second purpose of this study is to link education and earnings. It uses Mincerian type OLS regressions to estimate the relationship between investments in schooling and post-schooling, on the one hand, and the distribution of earnings, on the other hand, in three employment sectors for which reliable earnings data are available: the wage public sector, the wage private sector, and self-employment. But since the workforce might not have been randomly assigned to each sector, OLS earnings functions face the potential for carrying sample selection bias. To get around this deficiency, this study uses a modified version of Dubin & McFadden's (1984) model. This procedure, as is described in Section 3.5.2, is based on a multinomial logit model in its first-stage (selection) equation.

In the related literature no published studies are found employing similar methodologies to analyse such issues using Mozambican data. In addition, this analysis, in particular the first issue this chapter addresses, gives an indication of the degree of heterogeneity found in the Mozambican labour market (Demery & Grootaert, 1993). For instance, if schooling influences the chances of entry into the public sector in a different way than it influences entry into private and/or self-

[22] As pointed out by Kuepie, Nordman & Roubaud (2007, p. 10) in their Footnote 7, when the labour market is segmented, as is the case in economies such as Mozambique, despite the fact that some people may be able to choose from different alternatives, the bulk of the working-age population is faced with the impossibility of accessing certain highly (educationally) rationed jobs such as those in the public sector. This study nevertheless will assume everybody can "choose" which labour market sector to enter.

employment, then the questions raised here are important from a policy perspective. Many developing countries, including Mozambique, under the influence of the Bretton Woods institutions – International Monetary Fund and the World Bank – had undergone large scale economic liberalisation, expansion of the private sector[23], and reduction of the civil service (Glick & Sahn, 1997, p. 794). For the case of Mozambique, as the evidence below shows, this might have resulted in a greater proportion of people in self-employment/informal jobs, unpaid family work, unemployment and even non-participation, as the private sector does not seem to have grown in an adequate (quantitative and qualitative) way to absorb those released from public sector jobs. In turn, this might have exacerbated the poverty rate[24]. Further, as argued by Wambugu (2002, p. 2), the level of returns to schooling in each labour market sector will indicate the education potential for improving economic welfare and will serve as an incentive for poor individuals and households to invest in education. In other words, households' demand for schooling depends strongly on education's ability to provide them with attractive jobs (Kuepie, Nordman & Roubaud, 2007, p. 2). If households/individuals realise that education is profitable on the labour market, then they are likely to invest in it.

Section 3.2 of this chapter presents a brief review of studies employing similar techniques as in this study, but in other sub-Saharan African economies with a background similar to that of Mozambique. Section 3.3 describes the data sets and reports descriptive statistics of the variables of interest. Section 3.4 links education to employment outcomes. Earnings functions analyses without and with sample selectivity correction are presented in Section 3.5. The latter two sections study the sorting mechanisms in the Mozambican labour market as well as the determinants of earnings, with particular interest in educational investments. The chapter conclusions are presented in Section 3.6.

3.2 A Survey of Comparable African Studies

Several studies on Africa applying similar methodologies as in this study are found in the literature. This section surveys a few of these. A common feature of these studies is that they all focus on sub-Saharan economies, which are characterised by very segmented labour markets, i.e., most of the working-age population is engaged in informal/self-employment activities, including subsistence agriculture, with weak chances of working in the (relatively well paid) formal private and public

[23] The data set gives evidences of a small and lower level formal private sector in Mozambique. Education does not seem to affect significantly the chances of getting jobs in this sector. Most formal private sector workers are blue collars, as will be discussed below. To use Krishnan, Sellasie & Dercon´s words in the Ethiopian case which is reviewed below, "[t]he slow pace of privatisation and the contradictory signals to investors have also meant that the private sector remains marginal, particular in terms of employment" (Krishnan, Sellasie & Dercon, 1998, p. 24).

[24] In fact, analysis of the two latest income and expenditure surveys (IAF2002-03 and IOF2008-09) (not shown here, but available upon request) suggests that poverty levels remained the same in Mozambique, while the economy´s growth rates between 2002-2003 and 2008-2009 were rather high, as described by the online database of *Instituto Nacional de Estatística* (INE).

wage sectors. Nevertheless, so far none of these studies used Mozambican data in the analysis. This fact grants this study a pioneering position in the Mozambican context, thus initiating the debate among researchers, politicians, and ordinary citizens interested in the operation of the Mozambican labour market.

Given the similarities of the focus of these studies with the topic and research questions addressed in this chapter, it reviews them in a chronological order, from the oldest to the most recent relevant study. In this way the reader can learn cumulatively about the contributions made by each study, hence understanding the temporal evolution of what we know on the topic and research questions here discussed.

Paul Glewwe analysed private rates of returns to education in Ghana using data from the second year of the Ghana Living Standards Survey. The survey covered 3,200 households from all regions of Ghana, between 1988 and 1989. For a subsample of 1,586 households, tests on abstract reasoning, Mathematics, and reading comprehension were administered to household respondents aged nine to 55 years. Glewwe's aim with this study was to critically examine the extent to which rates of returns to education could be estimated using Becker's (1975) and Mincer's (1974) human capital theory, paying special attention to the case in which the quality of education was unevenly distributed (Glewwe, 1991, p. 11). Given that in developing countries a relatively high proportion of working-age individuals do not work in the wage sectors, and hence to avoid the selectivity bias that would stem from ordinary least squares (OLS) wage equations, he suggested a model of three possible activities: wage employment in the public sector, wage employment in the private sector, and a residual category which included both unemployed and self-employed people (Glewwe, 1991, p. 18). But unlike this chapter and the remainder of the studies surveyed below, Glewwe did not control for selection in a multiple employment segments fashion. He basically applied the Heckman two-step approach which is briefly described in Section 3.5.2.

Among other things, the author concluded that, first, accounting for the variability of school quality across time and space, years of education completed were a misleading indicator of human capital attained, as the private rate of returns to schooling was found to be substantially biased. Second, cognitive skills (Mathematics and reading comprehension tests) rather than innate ability (a test on abstract reasoning) or schooling credentials were the most important determinants of private sector wages. Nevertheless, in the public sector credentials were found to be important. Lastly, returns to schooling were found to be convex, thus reflecting the poor quality of primary schooling in that economy (Glewwe, 1991, p. 55-56). But for women in the public sector he found insignificant returns to education.

Vijverberg (1993) used a three waves overlapping panel data set from Ivory Coast (the Côte d'Ivoire Living Standard Survey) held in the period 1985-1987, to compare labour market returns to human capital between Ivorian men and women aged 20 to 60 years. His main objective was to ascertain whether women's labour market discrimination could explain the lower levels of education they usually get in such type of countries.

Unlike the case of industrialised economies which are characterised by competitive labour markets, in sub-Saharan African economies such as Mozambique, the premium that an employer pays to a worker within the wage sector with an additional year of education is very often greater than the earnings gain that such a worker could obtain in a self-employment/informal job. Reasons may include, among other things, barriers to entry into wage employment jobs and large public sectors with wage scales unreflective of worker productivity. As a result, returns to education in these alternative sectors will tend to differ substantial, and very often in favour of the wage sectors. Vijverberg therefore argued that returns to education should be estimated separately in each sector as well as account for the selection process involved (Vijverberg, 1993, p. 937-938). Hence, in his paper returns to education were analysed separately within the agricultural self-employment sector, the non-agricultural self-employment sector, and the wage sector, treating unemployment as the base category.

Instead of using the problematic method of Lee (1983) described in Section 3.5.2, which used to be considered best practice at the time, Vijverberg employed his own dual selection criteria method (Vijverberg, 1992), thus accounting for the selection process potentially present in the Ivorian labour market. This method[25] yielded information on the nature of selection both by labour force status choice as well as by migration (Vijverberg, 1993, p. 941). In other words, in his model the selectivity bias arose from the observation that wage workers selected themselves first by choosing to offer their time in the wage labour market. The offer of their time in the wage sector depended on the market wage offer. If it exceeded the value of time in the next best activity, an individual would choose wage employment. Second, if labour markets were not regionally segregated workers would have the incentive to move to higher-wage regions. Estimating wage equations only for individuals living in a particular region would yield results only applicable to persons living there. One should thus correct for selectivity by region, therefore obtaining appropriate estimates of the earnings that a randomly selected person from a different region would have earned (Vijverberg, 1995, pp. 159-160).

[25] Interested readers are referred to Vijverberg (1992) and Vijverberg (1995) for further details on the dual-λ method.

In sum, he found that returns to education were greater for women, on average. At all educational levels men tended to earn better than women, but the gap tended to narrow and even vanish the older and the more educated the women were, therefore increasing women's chances of obtaining employment in the wage sectors. Owing to this, combined with the fact that women were still less likely to get wage jobs than men, he concluded that the reason women attained, on average, significantly lower levels of education could not be found in lower returns to human capital, but rather in gender discrimination (Vijverberg, 1993, p. 962).

In a Guinean study, Glick & Sahn (1997) split the labour market into three employment segments, apart from non-participation: Self-employment, private wage employment and public wage employment. Their data set was a survey of 1,725 households performed in 1990 in Conakry (the capital of Guinea), which included 3,566 men and 3,306 women aged 15 years and above. The authors aimed to answer three research questions. Firstly, were labour market sectors different regarding the determinants of entry and earnings? Secondly, did investments in human capital yield similar returns in different labour market sectors? And lastly, how did women differ to men with respect to access to employment in different labour market sectors and earnings in the respective sectors?

To address the first question they estimated a model of labour market participation and determination of sector of employment. Given the multitude of choices involved, i.e., the four employment sectors, they employed a multinomial logit model to estimate the probabilities that an individual would have been found in each sector. They addressed the last two questions by estimating earnings functions for men and women in each sector for which earnings data were available. Since the determination of participation in a given sector was unlikely to be random they used the two-stage Lee (1983) method to correct their earnings functions for potential sample selection bias. They concluded that African urban labour markets were heterogeneous since determinants of entry into the various labour market sectors (education in particular) were quite different. For both men and women more education lowered the chances of self-employment jobs while it strongly increased the likelihood of public sector employment. Only for women did education impact positively on the likelihood of private sector employment. Nevertheless, women were found to be far less likely than men to get wage jobs. The authors also found that returns to education were positive in all sectors, both for men and women and similar across sexes. Lastly, and as pointed out by Vijverberg (1993), the fact that women fared worse than comparable men in all sectors, this reflected "either gender discrimination in pay or differences in the occupations that men and women tend[ed] to enter" (Glick & Sahn, 1993, p. 815).

Krishnan, Sellasie & Dercon (1998) used a combination of cross-section and panel data based on three household surveys conducted in 1990, 1994 and 1997, to study employment allocation in the Ethiopian urban labour market pre and post reform. Like the case of Mozambique immediately after the independence, since the 1980s the Ethiopian government implemented a highly regulated economy of the Soviet sort. By the beginning of the 1990s with the fall of the Soviet Union, the collapse in aid flows and the consequent defeat of the then government by the rebels (which then became the new government), a programme on economic recovery and reconstruction was put forward to rehabilitate (reform) the economy (Krishnan, Sellasie & Dercon, 1998, p. 2). The authors used their data on relatively new entrants into the labour market, i.e., individuals aged 15 to 29 years, in order to address the issues of changes in the labour market in that period of adjustment. They wanted to know how the labour market was changing in response to the reforms, particularly the effects of exogenous variables on allocation into employment segments as well as the determinants of earnings in the wage sectors.

In close similarity to this chapter, they split the labour market into five employment segments, namely, the public sector, the private sector, self-employment, unemployment and non-participation, and in contrast, they used the Lee approach to correct their wage equations for sample selection bias. In sum, for both men and women they found that schooling increased the likelihood of employment in the public sector. The same was found to be true for men in the private sector, but not for women. Women with completed primary education were less likely to work in the private sector relative to uneducated women, thus suggesting that the kind of activities open to poorly educated women within the private sector did not require any education (e.g. domestic workers). But women with at least secondary education completed (relative to no formal schooling) were more likely to work in the private sector. Lastly, for both men and women, schooling in comparison to no formal schooling was found to lower the chances of self-employment and not always guarded against unemployment.

With respect to their main research question, they found that while the proportion of unemployed (but educated)[26] individuals tended to increase, the Ethiopian labour market had been generally irresponsive to the pressures of the reform since 1990. In the analysed period the public sector shrank and real wages returned (declined) to the pre-reform period, formal private sector real wages

[26] According to the authors the fact that the proportion of unemployed individuals who had relatively higher levels of education completed increased, suggested that they were queuing for work, notably in the public sector (Krishnan, Sellasie & Dercon, 1998, p. 24).

increased[27], and private returns to education remained unchanged (Krishnan, Sellasie & Dercon, 1998, p. 24).

As part of his study on the Kenyan labour market, Wambugu (2002) examined separately for men and women, firstly, the impact of education and household characteristics on the chances of falling into public, private, informal, agricultural, and unpaid family work segments. Then, he compared the private returns to education across three employment sectors for which income data were available: the public sector, the private sector, and the informal sector. Like Glick & Sahn (1997) and other authors surveyed in this section, Wambugu employed a multinomial logit model to analyse the determination of sector of employment and the Lee (1983) method to control his earnings functions for selectivity bias.

The data set Wambugu (2002) used was a household survey of 59,183 individuals distributed across 10,857 households, both in the rural and urban parts of Kenya. The survey was conducted by the Keynan Central Bureau of Statistics between August and September 1994. The labour market analyses focused on non-students aged 15 to 65 years. After considering these criteria and after dropping workers lacking full information on crucial variables his sample consisted of 22,579 workers.

Wambugu concluded that additional schooling lowered the chances of agricultural and informal sector employment, but increased the chances of private (for females only) and public sector jobs (both for males and females). He argued that this might have been the case because it gave people access to better opportunities within the wage sectors, jobs which were relatively more secure and provided stable incomes. He also found that education had a stronger correlation with public sector jobs than with private sectors jobs, and the reason for this was the fact that "the public sector hiring criteria puts emphasis on formal education much more than the private sector does" (Wambugu, 2002, p. 19). On his second objective, he found positive returns to education for both men and women in all of the three sectors studied. Returns to secondary school were greater in the private sector, and women, on average, had greater returns than men. But "unless women [could] access wage employment the returns to education [might never] be realised" (Wambugu, 2002, p. 20). Lastly, returns to primary education in informal sector jobs were substantial, in particular for women.

Also employing, among other methods, a multinomial logit model and then the Lee (1983) method, Kuepie, Nordman & Roubaud (2007) studied the effects of schooling on labour market participation

[27] Thus narrowing the gap in relation to the public sector.

and earnings, respectively, in seven[28] French speaking major cities of the West African Economic and Monetary Union. They acknowledged the deficiencies of the Lee method, but argued that for their particular case, considering the small size of their labour market sector sub-samples, it performed better than alternative methods proposed elsewhere[29] (Kuepie, Nordman & Roubaud, 2007, p. 12). They were also worried about the potential endogeneity of the education variable in the earnings functions. To correct for this they used techniques such as proxies for child ability, a control function approach, and household fixed effects regressions.

Their data were taken from the cross-sectional, unique, and comparable *1-2-3 Surveys* conducted in these capital cities between 2001 and 2002. Their analysis included 69,595 individuals aged ten years and over, distributed across 17,841 households. They split the working-age population into those in non-paid work participation, informal sector jobs, formal private sector jobs, and public sectors jobs.

Among other things, they concluded that education did not always guard against unemployment, a result also found by Krishnan, Sellasie & Dercon (1998). They blamed this finding on the deterioration of African urban labour markets, caused in turn by the failure or absence of urbanisation policies and the structural adjustment policies advocated by the Bretton Woods institutions (structural adjustment policies that fostered the reduction of staff in the civil service) (Kuepie, Nordman & Roubaud, 2007, p. 19). Nevertheless, more schooling opened the doors to the best jobs, which were found in the wage sectors. They also found convex returns to education. This suggests that the potential for education to alleviate an individual or household's poverty increases with greater levels of educational attainment (Kuepie, Nordman & Roubaud, 2007, p. 32-33).

In most of the cities involved in the study education was found to be endogenous, thus leading to biases in the estimates of returns to schooling. Unlike the case when correction for the bias involved using father's characteristics as proxies for own child's ability, the control function approach, which used father's education and work status as instruments, as well as the household fixed effects regressions led to better estimations of returns to education. The latter method, in particular, suggested the existence of a strong influence of social capital on the relationship between schooling and labour market outcomes in these seven West African cities. Lastly, their study revealed that schooling, even at higher levels, was found to be strongly related to greater earnings also in the informal sector.

[28] The cities were Abidjan, Bamako, Cotonou, Dakar, Lome, Niamey and Ouagadougou.
[29] These alternative methods are scrutinised, among other sources, in Bourguignon, Fournier & Gurgand (2004; 2007) and described below in Section 3.5.2.

The studies summarised above give a general overview of the techniques one needs to bear in mind when considering tackling the relationship between education and labour market outcomes in a sub-Saharan African context, where labour markets are very segmented. Nevertheless, despite being informative in many aspects and thus advancing our knowledge on the topic, they suffer from a few insufficiencies. The next few paragraphs elaborate on this.

First, Glewwe (1991) did not consider the choice of self-employment versus wage sector jobs. Yet, the self-employment segment made up a large part of the labour force in Ghana. Second, also pointed out by Glick & Sahn (1997), Vijverberg's (1993) study did not distinguish between public and private wage employment. To assume that the determinants of earnings are identical within the public and private wage sectors is doubtful since government pay scales may reward schooling of workers in a way that does not reflect their productivity (Glewwe, 1991, p. 24).

Third, while the first two studies summarised here did not go very far with respect to accounting for sample selection bias associated with the multiple employment statuses that feature in these economies, the latter four did it. Nevertheless, all but one of the analysed studies focused only on the Lee method, despite its deficiencies, which will be discussed later. Krishnan, Sellasie & Dercon (1998) also employed alternative methods and answered informative questions, but their earnings functions aggregated men and women together. The impact of schooling on employment sector determination as well as on earnings is likely to affect men and women differently, as is shown in this chapter and was found in other studies.

Lastly, while child work is very common in African societies, including ten-year-olds in earnings functions like Glewwe (1991) and Kuepie, Nordman & Roubaud (2007) did might be misleading, since it might capture a high proportion of individuals that are still in school and do part-time work. The driving force behind their selection of labour market sectors to work in as well as the earnings they happen to get from their jobs might be different from those of adult full-time workers. While in this chapter school-age individuals are also included in the analysis, its negative impact on the results are minimised by raising the minimum age to 15 years, as per the International Labour Organisation (ILO) and Mozambican conventional definitions.

This chapter complements earlier research on the relationship between schooling and labour market outcomes in the African context, and covers some lacunas left open by past studies. More importantly, it is pioneering for tackling such questions in the Mozambican context.

3.3 Data

The main data used in this chapter are sourced from *Inquérito Integrado à Força de Trabalho*, the only labour force survey in Mozambique. This survey was conducted by the National Statistics Institute (INE) in conjunction with the Ministry of Labour over the years 2004 and 2005. It surveyed a stratified random sample of 17,151 households designed to be representative at the national and provincial levels and by area of residence (urban and rural). Among other things, the survey collected information on individual background and household characteristics, employment segment, industry, occupation, hours worked as well as labour earnings both in the wage and self-employment segments of the labour market.

Despite being relatively rich for a developing country such as Mozambique, the main data set this chapter uses is rather outdated. In this regard, INE also recently released the latest Income and Expenditure Household Survey *(Inquérito sobre o Orçamento Familiar – IOF2008-09)*, which contains comparable variables and followed similar sampling procedures. In order to gauge reliability of the analysis done based on the Labour Force Survey 2004-2005, this section therefore also makes use of the new data set. IOF2008-09 was executed between September 2008 and August 2009, inclusive, and surveyed a stratified random sample of 51,177 people distributed over 10,832 households nationwide. It was designed to be representative at the national and provincial levels, and by area of residence (urban and rural). Like the labour force survey, this survey collected information on individual background characteristics, education, employment status, sector of economic activity and occupation, as well as labour earnings, among other things. But the earnings data were collected only for those employed in the public and private wage sectors, thus making it relatively inferior relative to the labour force survey, particularly for the kind of education and labour market analyses intended here.

In this chapter conventional practice is followed and the sample is restricted to individuals aged 15 to 64 years. Table 3.1 on page 87 presents summary statistics of the variables used in the analysis. After deleting observations with missing data, the sample consisted of 15,915 men and 20,150 women, distributed between six employment states: wage employment in the public sector, wage employment in the private sector, self-employment, unpaid family work, unemployment and non-participation.

As is expected in developing and poor economies such as Mozambique, the proportion of men and women in self-employment is very high. As one can observe at the bottom of the table, 57% of men and 44% of women were self-employed. Wage employment in the public and private sectors, on the other hand, was very low, particularly among women. The proportion of women in the public and

private wage segments was below 2% for both sectors while for men it was 6% and 11%, respectively. Differently from men, women outside self-employment were mostly unpaid family workers (38%).

Turning to education, 16% of men and 43% of women had no formal schooling. For both genders, but more so for women, the figures were highest among the self-employed and unpaid family workers. For individuals who were wage employed, particularly those in the public sector, lack of formal schooling was very uncommon. Only 2% of men and 3% of women in the public sector had no formal schooling.

There was a major gap in primary school completion across genders. The proportion of women who had completed full primary education was about half (4%) that of men (9%). This gap was again more pronounced among self-employed and unpaid family workers, and almost non-existent within the wage employment segment of the working-age population.

The significance of schooling in determining the employment segment was more obvious at higher levels of education. While the proportion of all individuals with full secondary education completed was very low (2% for men and 1% for women), in the public sector it was as high as 14% for men and 20% for women. But in the self-employment and unpaid family worker segments, less than half a percent of men and women had completed full secondary school. This pattern is consistent with that at the tertiary education level, except that the proportion of women in the public sector with at least some higher education (8%) was well below that of men (12%). Accordingly, mean years of schooling varied widely among employment segments, pointing to large differences in skill requirements in different parts of the labour market. In the public sector, both men and women tended to be well educated, averaging 8.9 and 9.1 years of schooling, respectively. Among self-employed and unpaid family workers, in contrast, average educational attainment was very low – just, respectively, three and four years for men and two years for women in both sectors. Interestingly, mean years of education among the unemployed and non-participants was as high as was in the private sector, pointing to some queuing for wage public sector jobs, as mentioned before and further discussed below. The average age of the working-age population was 32 years for both men and women. Among employed men and women unpaid family workers were the youngest and public sector workers the oldest[30]. Private sector workers were much older than the unemployed and non-participants. The average age of men and women in the private sector was 32 years for both

[30] For men only. For the case of female self-employed workers seem to have been the oldest. But confidence intervals overlap, thus suggesting that age did not significantly differ between the two sectors (public sector and self-employment).

genders, but it was only 24 and 25 years among the unemployed, and 22 and 26 years among non-participants, respectively. Thus, though men in these three employment segments had equal average educational attainment, private sector workers were more experienced, on average.

The proportion of married men in the public, private and self-employment segments was significantly higher than was the case in the remainder of the segments. It averaged 87%, 66%, and 86% in the first three employment segments, but only 15%, 21%, and 11% among unpaid family workers, unemployed, and non-participants, respectively. This is consistent with the idea that married men are the household providers and therefore are more likely to be employed in sectors that bring external resources into the household. The causation goes either way, however, in the sense that often men might get married only when they are employed and therefore have the means to support a household. In contrast, the literature on female labour force participation finds that married women are less likely to work than unmarried women, presumably reflecting a higher reservation wage resulting from access to their spouses' incomes as well as due to childbearing and raising children (Glick & Sahn, 1997, p. 799). The data give some evidence of the above assertion. The proportion of married women was highest among unpaid family workers (81%) followed by self-employed women (61%). The proportion of married women in the public and private wage sectors was 54% and 37%, respectively. Married women were less likely to be wage employed.

The number of children in the household averaged two for both men and women and did not vary much across genders and across employment segments. The average number of adults in the household was three for both men and women and also there was not much variation across labour market segments. The proportion of household heads in the sample is highest among men. There is a clear divide between employment segments. For both men and women, the percentage of household heads in the wage and self-employment segments was significantly higher than in the remainder of the segments. This is consistent with the view that since household heads take the economic responsibility for the household they are more likely to work in employment segments that generate income. Again, it might have been the case that well employed household members end up becoming household heads. The proportion of rural individuals engaged in wage employment was relatively low. Only 30% of men and 16% of women in the public sector, and 29% and 16% of men and women, respectively, in the private sector were living in rural parts of the country. In contrast, for both men and women most self-employment and unpaid family work occurred in rural areas. Lastly, the share of wage employment was highest in Maputo City. This province alone employed 16% and 23% of the public sector men and women, respectively. The

unemployment share was also highest in Maputo City, as 26% and 29% of the unemployed men and women, respectively, were from Maputo City.

For the wage sectors, the survey contains information on the monetary value earned from the main occupation in the last 30 days. It also contains information on the number of hours worked in the last seven days. The measure of wage earnings used in this chapter was obtained by calculating the ratio of daily earnings to daily hours worked. The results were individual figures of hourly earnings from main occupation in Meticais (MTn[31]). For self-employed workers the survey reports weekly and monthly net income. This variable is calculated as the difference between the monetary amount obtained from the self-employment activities in question and the amount invested or spent to generate it. These net figures were converted to daily values and were divided by daily hours worked, thus obtaining figures for net hourly Meticais.

Table 3.2 on page 89 reports mean hourly earnings for men and women across three labour market sectors and by education level. To better describe the statistics this chapter also includes two figures (see Figure 3.1 and Figure 3.2 on page 90) illustrating mean hourly earnings as well as their respective 95% confidence intervals across employment sectors, for men and women, respectively. Within the wage sector men's hourly earnings were significantly higher within the public sector. While, on average, they earned 23 MTn per hour within the public sector, they earned only 9 MTn per hour within the private sector, and the confidence intervals do not overlap, thus suggesting high attractiveness of the civil service relative to private wage sector employment. Also for men, self-employment produced the highest pay. Self-employed men, on average, earned 29 MTn per hour. However, the confidence interval within this sector overlaps with the public sector´s confidence interval, thus self-employment hourly earnings were not significantly greater than public sector's, though they were significantly higher than private sector hourly earnings. For women, though mean hourly earnings seem to be highest within the public sector, the gap relative to hourly earnings within the private and self-employment sectors is not statistically significant.

[31] In 2006 three zeros were slashed from the Metical (MT), the Mozambican currency, without affecting its purchasing power. Between the 1st of January 2006 and the 1st of January 2007, which was the conversion period, the new Metical was denominated as *Metical da Noval Família* (MTn).

Table 3.1 – Summary statistics of variables used in the analysis

Employment Segment	Public Sector		Private Sector		Self-employment		Unpaid family work		Unemployment		Non-participation		All	
Variables	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Education level completed:														
No formal schooling	0.0176	0.0260	0.1091	0.0825	0.2075	0.5057	0.1432	0.4800	0.0427	0.1048	0.0666	0.2224	0.1555	0.4311
Some primary [1-6]	0.2233	0.1502	0.5406	0.4270	0.6685	0.4491	0.7195	0.4860	0.4463	0.5038	0.5089	0.4569	0.6059	0.4632
Full primary [7]	0.1250	0.1126	0.1221	0.1257	0.0679	0.0246	0.0751	0.0211	0.1524	0.1346	0.1263	0.1121	0.0887	0.0408
Some secondary [8-11]	0.3791	0.4374	0.1681	0.2148	0.0493	0.0196	0.0593	0.0127	0.2852	0.2204	0.2697	0.1848	0.1172	0.0534
Full secondary [12]	0.1388	0.1976	0.0354	0.0928	0.0047	0.0007	0.0028	0.0002	0.0523	0.0248	0.0163	0.0163	0.0197	0.0079
Tertiary education [+13]	0.1162	0.0762	0.0247	0.0573	0.0021	0.0003	0.0000	0.0000	0.0211	0.0117	0.0120	0.0075	0.0129	0.0036
Mean years of schooling	8.94	9.14	5.47	6.69	3.39	1.75	3.87	1.65	6.63	5.63	5.99	4.70	4.43	2.41
	(0.15)	(0.19)	(0.12)	(0.28)	(0.06)	(0.05)	(0.09)	(0.05)	(0.12)	(0.10)	(0.12)	(0.13)	(0.06)	(0.05)
Mean age	38.68	35.96	32.20	31.95	35.63	36.12	20.93	29.59	23.70	25.33	21.77	25.65	31.81	31.98
	(0.36)	(0.54)	(0.30)	(0.57)	(0.22)	(0.24)	(0.36)	(0.19)	(0.29)	(0.18)	(0.39)	(0.42)	(0.15)	(0.13)
Married	0.8675	0.5375	0.6565	0.3665	0.8645	0.6092	0.1533	0.8120	0.2077	0.5494	0.1141	0.2789	0.6544	0.6557
Mean no. of children in the household	2.29	2.12	2.21	2.15	2.22	2.23	2.68	2.60	2.22	2.46	2.32	2.33	2.28	2.40
	(0.08)	(0.10)	(0.06)	(0.13)	(0.03)	(0.04)	(0.17)	(0.09)	(0.07)	(0.06)	(0.11)	(0.07)	(0.04)	(0.04)
Mean no. of adults in the household	3.04	3.44	3.35	3.49	2.55	2.44	3.76	2.85	4.23	3.73	4.09	3.87	3.05	2.83
	(0.06)	(0.10)	(0.06)	(0.12)	(0.02)	(0.03)	(0.16)	(0.08)	(0.07)	(0.06)	(0.07)	(0.07)	(0.03)	(0.04)
Household head	0.8956	0.4402	0.6505	0.3464	0.8809	0.4499	0.0440	0.0083	0.1771	0.0969	0.1334	0.0928	0.6523	0.2260
Rural	0.2955	0.1648	0.2859	0.1643	0.7978	0.7138	0.8052	0.8646	0.1557	0.0866	0.3355	0.2752	0.6297	0.6757

Continues next page

Continued from previous page

Employment Segment	Public Sector		Private Sector		Self-employment		Unpaid family work		Unemployment		Non-participation		All	
Variables	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Provinces:														
Niassa	0.0428	0.0352	0.0155	0.0083	0.0549	0.0282	0.0645	0.0762	0.0138	0.0214	0.0424	0.0312	0.0470	0.0461
Cabo Delgado	0.0771	0.0494	0.0423	0.0139	0.1177	0.0949	0.0613	0.0954	0.0280	0.0364	0.1095	0.0797	0.0941	0.0875
Nampula	0.1577	0.0880	0.1239	0.0711	0.2355	0.1198	0.1879	0.2592	0.1316	0.1456	0.1530	0.1065	0.1994	0.1734
Zambezia	0.0843	0.0791	0.0741	0.0651	0.2415	0.2199	0.1885	0.2012	0.0628	0.0408	0.1353	0.1196	0.1870	0.1872
Tete	0.0668	0.0728	0.0559	0.0206	0.0790	0.0568	0.1357	0.1014	0.0157	0.0209	0.0451	0.0493	0.0746	0.0703
Manica	0.0686	0.0604	0.0484	0.0171	0.0643	0.0724	0.1041	0.0628	0.0275	0.0328	0.0716	0.0802	0.0652	0.0651
Sofala	0.1089	0.0466	0.1446	0.0879	0.0722	0.1211	0.0665	0.0434	0.0676	0.0735	0.1086	0.1164	0.0844	0.0854
Inhambane	0.0560	0.0879	0.0724	0.0642	0.0505	0.0984	0.1071	0.0855	0.0642	0.0394	0.0738	0.0604	0.0623	0.0855
Gaza	0.0387	0.1077	0.0891	0.1011	0.0379	0.1025	0.0546	0.0519	0.0896	0.0525	0.0706	0.0798	0.0518	0.0776
Maputo Province	0.1373	0.1444	0.1614	0.2197	0.0268	0.0518	0.0228	0.0176	0.2324	0.2443	0.0590	0.0808	0.0644	0.0599
Maputo City	0.1619	0.2285	0.1725	0.3311	0.0197	0.0344	0.0072	0.0053	0.2668	0.2924	0.1310	0.1959	0.0698	0.0620
No. of unweighted observations	1,219	505	2,481	483	7,226	8,546	1,529	6,052	1,660	2,602	1,800	1,962	15,915	20,150
Weighted proportion in each segment	0.0557	0.0157	0.1105	0.0149	0.5699	0.4380	0.1084	0.3849	0.0692	0.0798	0.0864	0.0667	1.0000	1.0000

Note: figures for the categorical variables correspond to proportions and should add up to 1; For all figures weighting and clustering effects were taken into account; linearised standard errors in parentheses. Source: Labour Force Survey 2004-2005

Table 3.2 – Mean hourly earnings across sectors and by education level completed

Employment Segment	Public Sector		Private Sector		Self-employment	
Variables	Men	Women	Men	Women	Men	Women
No formal schooling	9.32	6.59	5.18	3.81	12.95	16.81
	(3.33)	(0.62)	(0.82)	(0.81)	(2.39)	(8.80)
Some primary [1-6]	9.65	8.09	5.92	3.92	22.92	9.71
	(0.57)	(0.54)	(0.27)	(0.23)	(4.01)	(0.78)
Full primary [7]	11.27	10.91	7.17	5.32	21.20	25.14
	(0.91)	(1.03)	(0.44)	(0.63)	(2.36)	(10.50)
Some secondary [8-11]	16.55	16.33	11.37	12.92	72.20	20.95
	(1.03)	(0.88)	(0.79)	(2.08)	(34.18)	(3.47)
Full secondary [12]	29.75	26.64	35.26	31.88	60.71	174.13[32]
	(n.a.)	(2.62)	(n.a.)	(3.43)	(n.a.)	(149.95)
Tertiary education [+13]	74.93	64.35	52.80	82.15	232.44	31.61[33]
	(10.63)	(11.89)	(7.04)	(20.83)	(n.a.)	(18.60)
Mean hourly earnings	23.40	19.96	9.20	13.11	28.66	14.50
	(1.77)	(1.42)	(0.50)	(2.22)	(4.56)	(2.80)
No. of sample observations	1,053	438	2,177	440	2,455	1,655

Note: weighting and clustering effects were taken into account; linearised standard errors in parentheses; n.a. – missing standard error because of stratum with single sample unit; at the time of the survey USD1.00=25.00MTn; R1.00=3.50MTn; Source: Labour Force Survey 2004-2005

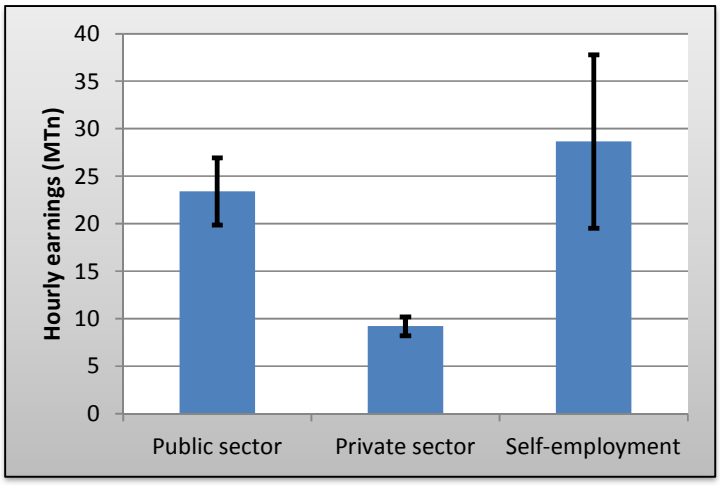
The premium for men's self-employment pay relative to men within the private sector, however, should be viewed with caution. For the sake of ensuring a large enough sample size this study lumped together two types of self-employed workers: self-employed workers with employees and self-employed workers without employees. The first group is rather small, constituting only 14% and 6% of self-employed men and women, respectively[34]. However, men and women were relatively better educated in this group. They averaged seven and six years of education completed, respectively, while the figures for self-employed men and women without employees were, correspondingly, five and four only. Therefore, mean hourly earnings within the first subgroup were significantly higher, in particular for men. Self-employment hourly earnings averaged 88 MTn and 50 MTn for men and women with employees, and 20 MTn and 13 MTn for men and women without employees, in that order. Nevertheless, men's average self-employment hourly earnings for those workers without employees were still double that of men's average private sector hourly earnings (20 MTn versus 9 MTn). Results such as these are not uncommon in the literature. Glick and Sahn found similar results for Guinea. In their view this does not mean that self-employment

[32] Only 9 observations in the sample.
[33] Only 2 observations in the sample.
[34] These percentages are in relation to those self-employed workers who reported earnings. Considering the entire sample of self-employed individuals reported in Table 3.1, self-employed men and women with employees correspond, respectively, to 7% and 2% only.

pays best but instead that self-employed workers underestimate the amount they spend or invest in their activities (Glick & Sahn, 1997, p. 821; Glick & Sahn, 1993).

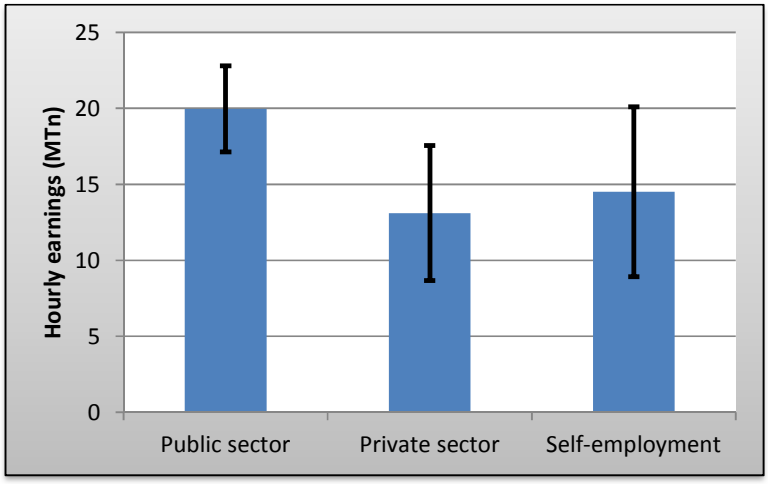
Across all sectors average hourly earnings grow with educational attainment. For those in self-employment the increase in average hourly earnings is not monotonic, as is the case for wage employment. In general, across all levels of education, men were better paid than women. But women with tertiary education in the private sector were much better paid on average than men with the same education level.

Figure 3.1 – Men's mean hourly earnings and 95% confidence intervals across sectors



Note: weighting and clustering effects were taken into account; at the time of the survey USD1.00=25.00MT; R1.00=3.50MT; Source: Labour Force Survey 2004-2005

Figure 3.2 – Women's mean hourly earnings and 95% confidence intervals across sectors



Note: weighting and clustering effects were taken into account; at the time of the survey USD1.00=25.00MT; R1.00=3.50MT; Source: Labour Force Survey 2004-2005

Table 3.3 and Table 3.4 on pages 92 and 94 use the IOF2008-09 data set to analyse the developments in the Mozambican labour market between 2004-2005 and 2008-2009, but most importantly to gauge the reliability of the main data source – the Labour Force Survey 2004-2005. The tables suggest that, overall, between the two surveys there was not much change in the Mozambican labour market. The distribution of the working-age population among the different segments, and the relationship between education and labour market outcomes remained quite similar. In other words, data from IOF2008-09 also suggest that educated men and women tended to work within the public wage segment of the labour market as well as tended to earn significantly better than uneducated men and women, respectively, notably in the public sector. These similar findings across the two surveys grant reliability to the main data set, particularly as four years is not a long enough time span for one to expect dramatic changes in a country's labour market.

Minor changes occurred, however. Firstly, women seem to be catching up with men concerning educational attainment. While the proportion of uneducated men only declined from 16% to 15% between the two surveys, the proportion of uneducated women decreased significantly from 43% to 36%. Analysing birth cohorts one concludes that most of the decline occurred amongst young women, indicating that the two surveys are not coherent on this matter. The proportion of uneducated women declined from 23% to 14% for women aged 15 to 19 years and from 34% to 24% for women aged 20 to 24 years. Secondly, the proportion of educated women in the public sector rose sharply, particularly at the tertiary education level. It increased from 8% in 2004-2005 to 14% in 2008-2009, whereas for the case of men it decreased from 12% to 11%. Thirdly, mean years of education completed increased by one year for both genders. Fourthly, the proportion of unemployed and non-participant men that were household heads declined from 18% and 13% in 2004-2005 to 10% and 6% in 2008-2009, respectively. Lastly, average hourly earnings for women in the public sector increased from 20 MTn to 31 MTn, whereas for men it increased only by 5 Meticais, from 23 MTn to 28 MTn. However, between the two surveys men's private sector average hourly earnings almost doubled. It changed from 9 MTn to 17 MTn while the women only had a 4 MTn increase from an initial pay of 13 MTn in 2004-2005. In IOF2008-09 mean hourly earnings for the same gender differ significantly across labour market segments, but the differences across genders within the same labour market segment are not statistically significant (see Table 3.4).

The differentials across employment segments in the men's and women's characteristics, particularly schooling differences, may explain the differentials in access to employment and also in earnings. This study next explores these issues further using multivariate regressions.

Table 3.3 – Summary statistics of variables used in the analysis

Employment Segment	Public Sector		Private Sector		Self-employment		Unpaid family work		Unemployment		Non-participation		All	
Variables	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Education level completed:														
No formal schooling	0.0185	0.0058	0.0852	0.1473	0.2094	0.4315	0.1376	0.4031	0.0319	0.0770	0.0583	0.1289	0.1455	0.3567
Some primary [1-6]	0.1788	0.0979	0.4449	0.3282	0.6329	0.4791	0.5828	0.5130	0.3155	0.3756	0.4087	0.3947	0.5356	0.4725
Full primary [7]	0.1524	0.0851	0.1649	0.1269	0.0793	0.0418	0.1212	0.0411	0.1588	0.1601	0.1483	0.1295	0.1128	0.0587
Some secondary [8-11]	0.3815	0.5092	0.2383	0.2756	0.0690	0.0421	0.1523	0.0406	0.3858	0.3339	0.3514	0.3209	0.1689	0.0966
Full secondary [12]	0.1590	0.1623	0.0426	0.0624	0.0069	0.0045	0.0061	0.0022	0.0709	0.0412	0.0167	0.0186	0.0235	0.0104
Tertiary education [+13]	0.1099	0.1398	0.0241	0.0596	0.0025	0.0011	0.0001	0.0001	0.0372	0.0121	0.0165	0.0073	0.0138	0.0051
Mean years of schooling	9.15	9.97	6.13	6.47	3.61	2.25	4.71	2.34	7.60	6.59	6.53	5.93	4.94	3.09
	(0.22)	(0.22)	(0.16)	(0.38)	(0.06)	(0.09)	(0.13)	(0.07)	(0.20)	(0.16)	(0.14)	(0.15)	(0.08)	(0.08)
Mean age	40.18	34.68	32.62	30.82	38.60	36.45	21.15	30.27	23.77	25.19	20.29	23.11	32.07	31.43
	(0.92)	(0.63)	(0.41)	(0.62)	(0.25)	(0.24)	(0.33)	(0.23)	(0.41)	(0.32)	(0.32)	(0.45)	(0.20)	(0.14)
Married	0.8623	0.6282	0.6775	0.3875	0.9223	0.6676	0.1822	0.7511	0.2048	0.4825	0.0743	0.2423	0.6269	0.6499
Mean no. of children in the household	3.59	2.74	3.20	2.96	3.26	3.54	3.26	3.62	3.27	3.20	3.53	3.42	3.30	3.53
	(0.33)	(0.15)	(0.15)	(0.19)	(0.05)	(0.14)	(0.11)	(0.09)	(0.22)	(0.12)	(0.15)	(0.16)	(0.06)	(0.08)
Mean no. of adults in the household	3.91	4.11	4.23	4.01	2.91	3.19	4.24	3.52	5.07	4.44	4.64	4.60	3.66	3.58
	(0.15)	(0.15)	(0.15)	(0.20)	(0.04)	(0.09)	(0.11)	(0.11)	(0.14)	(0.13)	(0.11)	(0.14)	(0.06)	(0.07)
Household head	0.8150	0.2784	0.5380	0.2537	0.8987	0.3842	0.0273	0.0056	0.0987	0.0737	0.0582	0.0429	0.5619	0.1482
Rural	0.3227	0.2103	0.2768	0.2018	0.7998	0.6355	0.7974	0.8438	0.1636	0.0575	0.4217	0.2779	0.6364	0.6539

Continues next page

Continued from previous page

Employment Segment	Public Sector		Private Sector		Self-employment		Unpaid family work		Unemployment		Non-participation		All	
Variables	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Provinces:														
Niassa	0.0734	0.0963	0.0198	0.0139	0.0614	0.0383	0.0712	0.0643	0.0099	0.0124	0.0336	0.0368	0.0531	0.0497
Cabo Delgado	0.0609	0.0687	0.0213	0.0032	0.1048	0.1244	0.0575	0.0454	0.0351	0.0477	0.0361	0.0354	0.0728	0.0706
Nampula	0.1412	0.0822	0.1112	0.0738	0.2226	0.1313	0.1992	0.1890	0.1505	0.0869	0.1568	0.1837	0.1896	0.1599
Zambezia	0.1676	0.0887	0.0665	0.0242	0.2232	0.1469	0.1883	0.2259	0.0577	0.0336	0.1572	0.1158	0.1797	0.1724
Tete	0.0376	0.0416	0.0455	0.0074	0.1132	0.0379	0.1164	0.1207	0.0397	0.0280	0.0846	0.0643	0.0947	0.0792
Manica	0.0755	0.0468	0.0385	0.0274	0.0760	0.0731	0.0760	0.0739	0.0213	0.0272	0.0600	0.0418	0.0670	0.0667
Sofala	0.1457	0.0904	0.1216	0.0875	0.0659	0.1345	0.1137	0.0700	0.0681	0.0475	0.1266	0.1275	0.0926	0.0966
Inhambane	0.0314	0.1026	0.0385	0.0710	0.0478	0.0642	0.0650	0.0838	0.0178	0.0123	0.0862	0.0964	0.0519	0.0747
Gaza	0.0523	0.0628	0.1036	0.1064	0.0358	0.1015	0.0807	0.0898	0.0959	0.0513	0.0753	0.0771	0.0602	0.0902
Maputo Province	0.0794	0.1185	0.2355	0.2948	0.0255	0.0845	0.0201	0.0302	0.2281	0.2545	0.0957	0.0993	0.0705	0.0733
Maputo City	0.1350	0.2014	0.1981	0.2904	0.0238	0.0633	0.0117	0.0071	0.2761	0.3987	0.0878	0.1220	0.0680	0.0666
No. of unweighted observations	779	350	1,889	369	5,112	4,979	1,505	5,125	661	1,097	1,423	1,594	11,369	13,514
Weighted proportion in each segment	0.0547	0.0167	0.1204	0.0166	0.4855	0.3342	0.1760	0.4787	0.0468	0.0546	0.1166	0.0993	1.0000	1.0000

Note: figures for the categorical variables correspond to proportions and should add up to 1; For all figures weighting and clustering effects were taken into account; linearised standard errors in parentheses; Source: Income and Expenditure Household Survey 2008-09

Table 3.4 – Mean hourly earnings across sectors and by education level completed

Employment Segment	Public Sector		Private Sector	
Variables	**Men**	**Women**	**Men**	**Women**
No formal schooling	7.57	6.22	13.47	15.81
	(3.76)	(1.81)	(3.07)	(5.08)
Some primary [1-6]	17.77	14.60	14.88	11.05
	(2.18)	(2.82)	(1.46)	(1.71)
Full primary [7]	15.70	16.10	12.97	7.26
	(5.03)	(2.38)	(12.97)	(1.06)
Some secondary [8-11]	22.17	26.65	16.40	14.81
	(1.29)	(2.85)	(1.51)	(3.95)
Full secondary [12]	31.84	45.09	35.66	24.68
	(3.07)	(11.57)	(7.26)	(2.77)
Tertiary education [+13]	78.25	51.60	69.31	75.93
	(12.88)	(4.85)	(14.08)	(15.96)
Mean hourly earnings[35]	27.87	30.82	17.19	16.87
	(2.38)	(2.68)	(1.01)	(2.08)
No. of sample observations	743	339	1,754	341

Note: weighting and clustering effects were taken into account; linearised standard errors in parentheses; at the time of the survey USD1.00=25.00MT; R1.00=3.50MT; Source: Income and Expenditure Household Survey 2008-09

3.4 Modelling Employment and Education

This part of the study estimates a model of labour market participation and determination of segment of employment. As discussed above, a plausible view of the labour market in Mozambique would allow for six distinct employment segment alternatives. These include wage employment in the private sector, wage employment in the public sector, self-employment, unpaid family work, non-participation and unemployment.

Given the number of discrete choices involved a multinomial logit model provides the most suitable econometric approach to estimate the probabilities that an individual will be found in each employment segment. Let y denote a random variable taking on the values {1, 2, ..., J}, where J is a positive integer representing the assignment of an individual to a particular labour market segment. Let x denote a set of conditioning variables such as the individual, household and regional characteristics that might have a bearing on the determination employment segment. We are interested in how changes in the elements of x affect the response probabilities, P(y=j|x), j=1, 2, ..., J. Given that probabilities must sum to one, P(y=1|x) is determined once we know the probabilities

[35] Differences across sectors are statistically significant, but not across genders within the same labour market segment.

for j=2, 3, ..., J (Wooldridge, 2002, p. 497). If x is a 1*K vector with first-element unity, the multinomial logit model has response probabilities

$$P(y = j|x) = \frac{e^{x\beta_j}}{1 + \sum_{h=1}^{J} e^{x\beta_h}}, j = 2, 3, \dots , \quad J \tag{1}$$

where β_j is K*1, j=2, ..., J. Because the response probabilities must sum to unity,

$$P(y = 1|x) = \frac{1}{1 + \sum_{h=1}^{J} e^{x\beta_h}}, j = 2, 3, \dots , \quad J \tag{2}$$

In this particular case modelling employment sector determination results in a six-way multinomial logit model. The way schooling and other factors determine the choice of employment sector is expected to vary by employment sector. The models are employed for men and women separately and control for place and province of residence, among other things. In order to ensure model identification the parameter vector associated with one of the employment sectors must be set to zero (Cameron & Trivedi, 2009, p. 484). Both the men and women regressions are normalised by setting the parameter vector associated with non-participation equal to zero. The coefficient estimates are then interpreted with respect to this category.

Parameter estimates from multinomial logit models of sector of employment determination for men and women are presented in Table 3.5 on page 97. Wald tests of the null hypothesis that all regression coefficients associated with each of the twelve pairs of employment segments are equal to zero were computed and rejected at the 0.0001 level. This suggests that the employment sectors need not be combined, i.e., the six-way split provides a useful view of the labour market structure in Mozambique.

The coefficient estimates in Table 3.5 do not indicate the impact of a change in an explanatory variable on the probability of entering a labour market segment. Instead, they are interpreted as the impact of the associated independent variable on the log odds of the particular labour market segment relative to non-participation (the base category). Because interpretability of the coefficients in this way is rather difficult, in this chapter marginal effects of changing the values of the covariates on the probability of observing a labour market segment choice, while the other covariates are kept fixed, were also computed. For categorical covariates such as level of education completed, a marginal effect shows how the probability of entry into a particular employment sector is predicted to change as the variable shifts from no formal schooling (the reference group) to, say, full primary school. For continuous variables (such as age) it measures the instantaneous rate of change in the probability of entry in an employment segment as age changes.

Very often studies in which the dependent variable is categorical and the samples are large report marginal effects at means (of other covariates). But whenever samples are small or moderate they can largely deviate from the marginal effects at values other than the mean. Therefore, according to Greene (2003, p. 668), current practice supports computing instead the sample average of individual marginal effects to obtain the overall marginal effect.

Table 3.6 and Table 3.7 on pages 100 and 101, respectively, report the average marginal effects from changes or shifts in the covariates for the men and women sample. The estimated marginal effects of schooling on employment sector allocation presented in the tables are broadly intuitive and in line with the descriptive statistics. Controlling for other determining factors of employment sector choice, for both men and women more education compared to no formal schooling increases the chances of employment in the public sector. For example, if a man has completed full primary education the average probability of public sector employment rises by 0.07, and it rises by 0.30 for a man with full secondary education. The corresponding figures for women are 0.06 and 0.53, respectively.

As the number of years of education completed rises women are also more likely to be employed in the private sector. However, the marginal effects for men within the private sector are negative and statistically significant, except for those with some primary education and full secondary. This suggests that men who have completed full primary school or further levels are less likely to work in the private sector than men without any formal schooling. This contrasting impact of education across genders might result from the occupational choices open to each gender. The data set indicates that more than half (54%) of men in the private wage sector were (blue collar) labourers, craftsmen or unskilled workers, while a very low proportion were, say, clerks (4%). In contrast, the corresponding proportions for women were 23% in the low skilled occupations and 14% in clerical occupations. Clearly, contrary to the case of men, entry into the type of occupations open to, or chosen by, women required relatively higher levels of education. But also higher education raises chances of getting public sector jobs for men.

Table 3.5 – Multinomial logit estimates of employment segment determination

	Men					Women				
Variables	**Public Sector**	**Private Sector**	**Self-Employment**	**Unpaid Family Work**	**Unemployment**	**Public Sector**	**Private Sector**	**Self-Employment**	**Unpaid Family Work**	**Unemployment**
Education level completed										
Some primary	1.0968***	–0.2822	–0.2209	–0.1454	0.1292	2.0482***	1.2402***	0.0788	–0.0747	0.5256***
Full primary	2.2175***	–0.5859**	–0.5792**	–0.5918**	0.0172	4.0180***	1.5815***	–0.5585***	–0.6178***	0.4860***
Some secondary	2.4141***	–1.3190***	–1.7347***	–1.6763***	–0.3053	4.9593***	1.4536***	–1.2263***	–1.4948***	0.3875**
Full secondary	3.2252***	–0.9474***	–1.9582***	–1.9703***	0.2210	6.3054***	2.5332***	–2.3384***	–3.1204***	0.2618
Tertiary education	2.8652***	–1.5882***	–2.5095***	–26.9029***	–0.9089**	5.5380***	2.4016***	–2.9574***	–30.6927***	0.2199
Age	0.7801***	0.6040***	0.5641***	0.2985***	0.4047***	0.6424***	0.5533***	0.3549***	0.2188***	0.2767***
Age squared	–0.0095***	–0.0080***	–0.0073***	–0.0040***	–0.0055***	–0.0076***	–0.0073***	–0.0044***	–0.0028***	–0.0044***
Rural	0.2804	–0.4330**	1.2264***	1.7165***	–0.6330***	0.6674***	–0.0122	1.1373***	2.2334***	–1.0812***
Household headship	1.4310***	0.9646***	1.5246***	–2.6147***	–0.1718	2.1505***	1.5135***	1.8247***	–2.9052***	0.8935***
No. of children in the household	0.0038	–0.0054	–0.0028	0.0311	0.0112	–0.0024	–0.0317	0.0536**	0.1257***	0.0283
No. of adults in the household	–0.1121***	–0.0568*	–0.1053***	–0.0570*	–0.0511**	0.0321	–0.0251	–0.1519***	–0.1405***	–0.0400*
Married	1.4788***	1.1632***	1.4715***	0.9614***	0.4305**	0.9134***	0.0089	1.3444***	1.2635***	1.0719***
Constant	–16.7794***	–8.2403***	–8.5490***	–5.9399***	–5.1095***	–18.2387***	–11.3890***	–6.0021***	–5.0401***	–3.8405***
No. of observations			15,915					20,150		
Pseudo R-squared			0.3841					0.3637		
Log-likelihood			–3,390,430					–4,109,178		
Wald χ^2 (d.f)			20,284.26 (110)					19,719.47 (110)		
Prob > χ^2			0.0000					0.0000		

Note: non-participation is the base category; model estimation accounts for survey design; * p<0.1, ** p<0.05, *** p<0.01; model includes dummies controlling for province of residence

For both genders schooling is significantly and negatively associated with self-employment choice, except for less than full primary completion. This means that for both men and women having completed full primary education or further schooling lowers the chances of being self-employed relative to men and women without any formal schooling, respectively, and the effects are stronger amongst women. Similar patterns are observed with respect to unpaid family work, and again education seems to have a much stronger effect on women than men. For instance, in the case of men, the coefficients on some primary and full primary school are not statistically different from zero. Further, having completed secondary education lowers the average probability of unpaid family work by 0.09 only. For women, however, even some primary education is associated with (statistically significant) lower chances of unpaid family work, and the corresponding fall in this average probability for full secondary education relative to no schooling is as high as 0.24.

Though significant (except for men at the tertiary education level) and positive the average marginal effects of schooling associated with unemployment are very small for both men and women. For instance, for both sexes having completed full secondary school increases the average probability of unemployment only by 0.07 for men and 0.04 for women. The positive relationship between schooling and the likelihood of unemployment might result from a mismatch between skills demanded by the private wage sector and self-employment activities, on the one hand, and the type of jobs and remuneration educated workers were hoping to get, on the other hand. They may perhaps rather stay unemployed than enter any of these employment segments while queuing for the possibility of entry into the wage public sector (Krishnan, Sellasie & Dercon, 1998, p. 17).

In sum, and as discussed in the literature review, developing country studies employing similar methodologies end up concluding that education, relative to no formal schooling, increases the chances of wage employment and lowers the chances of employment outside the wage labour segment. In the concluding section this chapter elaborates more on this.

The impact of age on employment segment allocation seems to be concave for both men and women and across most sectors. The older a man or woman is the more likely he or she is to be in public, private or self-employment, but the rate of change tends to flatten as he or she gets older. With respect to unpaid family work the reverse tends to occur, however. Individuals become less likely to fall into this employment segment as they grow older, but at a decreasing absolute rate.

Rural men are less likely to fall into non-participation, to be unemployed and to work in the public and private wage employment segments, but more likely to be self-employed and unpaid family workers relative to urban men. The same occurs with rural women, except that they do not significantly differ from urban women with respect to chances of public sector employment and self-employment.

Headship of household is positively associated with a mode of employment that brings resources into the household. In other words, for both men and women, being a household head relative to being an ordinary household member significantly increases the chances of working in the public, private (for women only) and self-employment segments, but significantly lowers the chances of falling into non-participation, unemployment (for men only) and unpaid family work. An interesting result is that household headship seems to drive both men and women into self-employment, as the average marginal effects are much larger for this segment than for the public and private sectors. Given the lower chances of entry into the public and private sectors, self-employment seems to be the measure of last resort for those with the responsibility of providing external resources to the household.

The number of children in the household does not have any significant impact on employment segment determination for men. But for women, though the average marginal effects are rather small, a greater number of children in the household makes them less likely to fall into non-participation and private wage segment, while it increases their chances of falling into unpaid family work. This makes evident the gender biases with respect to child rearing in such countries. Results such as these are not uncommon in the literature. For instance, Bhorat & Leibbrandt (2001) also found that the number of children in the household was insignificant in determining African men's labour force participation decisions in South Africa, but significantly deterred women from participation.

The marginal effects of a greater number of adults in the household are also very small for both sexes. Nevertheless, it significantly lowers the chances of self-employment for both men and women, and increases the likelihood of unemployment and public sector work for women. Lastly, married men are significantly more likely to fall either into public sector or self-employment segments of the labour market, and less likely to be unemployed or to fall into non-participation than unmarried men. Married women, in contrast, are significantly less likely to work in the private

sector segment, but more likely to be self-employed, unpaid family workers or unemployed. These results are consistent with African societies where women have to balance domestic responsibilities with the need to augment household income.

Now this section describes the model prediction by education level completed for men and women separately. It uses the method of predictive margins in which we vary education across the whole sample of men and women and then average the predictions within each gender sample. In other words, we first pretend that all people in our male (female) sample have no formal schooling but hold their other characteristics constant. Then the probability of falling into each employment segment is calculated. Next we pretend that all people in the male (female) sample have only some primary schooling, still holding their other characteristics constant, and calculate the probabilities of each outcome. This process is repeated up to tertiary education. The difference between those six sets of calculated probabilities for men and women within each labour market segment, then, is the difference associated with schooling, holding other characteristics constant, i.e., the difference corresponds to the average marginal effects of education presented in Table 3.6 and Table 3.7.

Table 3.6 – Multinomial logit average marginal effects of employment segment allocation for men

	Men					
Variables	**Public Sector**	**Private Sector**	**Self-Employment**	**Unpaid Family Work**	**Unemployment**	**Non-Participation**
Education level completed						
Some primary	0.0140***	–0.0165	–0.0198	0.0012	0.0144*	0.0067
Full primary	0.0686***	–0.0270**	–0.0646***	–0.0212	0.0208**	0.0233**
Some secondary	0.1791***	–0.0365***	–0.2032***	–0.0605***	0.0385***	0.0826***
Full secondary	0.3045***	–0.0276	–0.3182***	–0.0934***	0.0667***	0.0680***
Tertiary education	0.3570***	–0.0343*	–0.3165***	–0.1785***	0.0177	0.1546***
Age	0.0071***	0.0118***	0.0235***	–0.0172***	–0.0005	–0.0247***
Age squared	–0.0001***	–0.0002***	–0.0003***	0.0002***	0.0000	0.0003***
Rural	–0.0116*	–0.1457***	0.1584***	0.1179***	–0.0709***	–0.0482***
Household headship	0.0149***	0.0134	0.3095***	–0.2646***	–0.0469***	–0.0263***
No. of children in the household	0.0001	–0.0009	–0.0024	0.0033	0.0004	–0.0006
No. of adults in the household	–0.0009	0.0023	–0.0079**	0.0022	0.0005	0.0039***
Married	0.0087**	0.0067	0.0940***	–0.0184	–0.0335***	–0.0576***

Note: * p<0.1, ** p<0.05, *** p<0.01; Model includes dummies controlling for province of residence

Table 3.8 on page 102 reports the predicted probabilities of entry into each labour market segment. They mirror the average marginal effects of education. For both genders education is the key to public service employment. For those with less than primary education completed, the chances of public sector employment are extremely low. But those with completed full secondary or further

education are very likely to fall into this employment segment. Private sector average entry probabilities are very low, notably for women. For men they tend to fall with educational attainment while the opposite happens for women. But at all educational levels women are less likely to be employed in the private sector than men. Men and women with no formal schooling are very likely to be self-employed. The average probability of falling into this sector is 0.56 for both sexes. In contrast, for men and women with university education the probabilities of self-employment are, correspondingly, 0.24 and 0.10 only. The chances of doing unpaid family work also drop with educational attainment. For instance, both men and women with tertiary education have a zero average probability of falling into this segment.

As discussed before the chances of falling into unemployment tend to increase with schooling, notably for women, representing a possible mismatch between demand and supply of skills, and/or the preference for queuing for public sector jobs as educational attainment increases. The chances of non-participation for both men and women also rise with education. As shown in Figure 3.3 on page 103 this pattern might be due to the fact that the proportion of non-participant men and women that are still in school also tends to increase with educational attainment.

Table 3.7 – Multinomial logit average marginal effects of employment segment allocation for women

	Women					
Variables	**Public Sector**	**Private Sector**	**Self-Employment**	**Unpaid Family Work**	**Unemployment**	**Non-Participation**
Education level completed						
Some primary	0.0060***	0.0092***	–0.0076	–0.0225***	0.0230***	–0.0082
Full primary	0.0706***	0.0209***	–0.1042***	–0.0404***	0.0405***	0.0126
Some secondary	0.2149***	0.0226***	–0.2296***	–0.1049***	0.0537***	0.0434***
Full secondary	0.5309***	0.0587***	–0.4557***	–0.2386***	0.0381**	0.0666***
Tertiary education	0.4803***	0.0839***	–0.4551***	–0.3023***	0.0713***	0.1219***
Age	0.0061***	0.0037***	0.0145***	–0.0085***	0.0006	–0.0163***
Age squared	–0.0001***	–0.0000***	–0.0002***	0.0001***	–0.0001***	0.0002***
Rural	–0.0002	–0.0107***	0.0095	0.1902***	–0.1233***	–0.0656***
Household headship	0.0223***	0.0064*	0.4767***	–0.4551***	–0.0057	–0.0447***
No. of children in the household	–0.0008	–0.0011*	–0.0031	0.0091***	–0.0008	–0.0034***
No. of adults in the household	0.0026***	0.0008	–0.0088***	–0.0028	0.0026**	0.0056***
Married	–0.0014	–0.0187***	0.0525***	0.0180*	0.0119***	–0.0624***

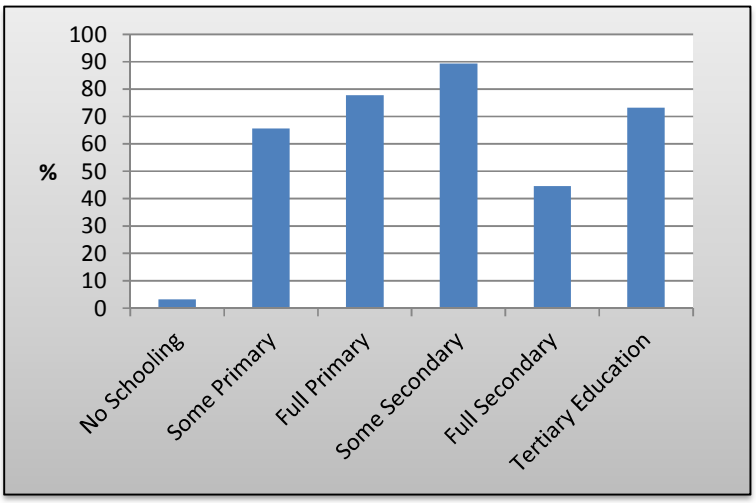
Note: * p<0.1, ** p<0.05, *** p<0.01; Model includes dummies controlling for province of residence

Table 3.8 – Predicted labour market sector entry probabilities, for men and women, by level of education completed

	Education Level Completed / Labour market sector	No Schooling	Some Primary	Full Primary	Some Secondary	Full Secondary	Tertiary Education
Men	Public Sector	0.0054	0.0194	0.0741	0.1845	0.3099	0.3624
	Private Sector	0.1410	0.1245	0.1140	0.1044	0.1134	0.1067
	Self-Employment	0.5605	0.5407	0.4959	0.3573	0.2423	0.2441
	Unpaid Family Work	0.1785	0.1798	0.1573	0.1181	0.0851	0.0000
	Unemployment	0.0535	0.0679	0.0743	0.0919	0.1202	0.0712
	Non-Participation	0.0611	0.0678	0.0844	0.1437	0.1291	0.2156
		1.000	1.000	1.000	1.000	1.000	1.000
Women	Public Sector	0.0010	0.0070	0.0716	0.2159	0.5320	0.4813
	Private Sector	0.0050	0.0142	0.0259	0.0275	0.0637	0.0889
	Self-Employment	0.5581	0.5505	0.4539	0.3285	0.1024	0.1030
	Unpaid Family Work	0.3023	0.2798	0.2619	0.1973	0.0636	0.0000
	Unemployment	0.0594	0.0824	0.0999	0.1131	0.0975	0.1307
	Non-Participation	0.0742	0.0660	0.0868	0.1176	0.1408	0.1961
		1.000	1.000	1.000	1.000	1.000	1.000

Note: all probabilities significant at the conventional levels of significance, 1%, 5%, or 10%

Figure 3.3 – Percentage of non-participants currently in school



3.5 Modelling Earnings and Education

Education is conventionally understood to bring many benefits. In this chapter we are interested in its economic benefits, specifically labour market benefits. As demonstrated in the foregoing section, one of the labour market benefits of education is to increase the chances of wage employment in the public and private sectors. Education is also believed to lead to increases in labour productivity and earnings (Cohn & Geske, 1990). Acknowledging these benefits, several methods have been developed to quantify their significance. The returns to education approach gained relevance and is based on the premise that education results in direct measurable returns to the individual and society.

Two methods stand out as the most important for calculating rates of returns to education (Psacharopoulos, 1981; Johnes, 1993, p. 28). The first method – the algebraic method – consists in finding the discount rate that equates the stream of education benefits to the stream of education costs, where these costs include only foregoing earnings, i.e., the opportunity costs of staying in school beyond the age of 18 rather than working in the labour market. In other words, if Y stands for labour earnings, c for cost years, n for benefit years, and h and s are subscripts for higher and secondary education, in that order, the rate of return (r) is found by solving the following equation for r:

$$\sum_{t=1}^{n}(Y_h - Y_s)_t(1+r)^{-t} = \sum_{t=1}^{c}(Y_s)_t\,(1+r)^t \qquad (3)$$

The problem with this approach is that it requires detailed data on age-earnings profiles by educational level, which may not be readily available, notably in the case of developing economies (Johnes, 1993, p. 29).

The second method of calculating returns to education is the earnings function approach. It started in the 1970s with the publication of Mincer's (1974) book and since then it became widely used (Psacharopoulos, 1981, p. 321). This approach relies on the human capital theory that links investment in education to higher labour productivity and thus higher earnings. In his model Mincer explains inequality in earnings by focusing on schooling and post-schooling investment as the central explanatory variables. Schooling investment is measured by years of education completed and post-schooling investment is given by labour experience, in the absence of more specific measures. Formally, his basic model is described by the following equation:

$$\ln Y = a + bS + cE + dE^2, \qquad (4)$$

where Y represents earnings, S stands for years of education, E stands for years of work experience, and a, b, c, and d are parameters to be estimated. Equation (4) is estimated using standard OLS techniques and cross-sectional data. Using U.S. census data of 1960 and different versions of Equation (4), Mincer found that schooling and post-schooling investment accounted for close to two-thirds of the inequality of earnings of adult, white, urban men (Mincer, 1974, p. 96). The model in its simplest form is a powerful tool for explaining earnings inequality, but it is often expanded to control for other factors that might have effects on worker pay, such as worker occupation, gender, sector of employment, and location, among other things.

3.5.1 Standard Earnings Functions

This study follows the Mincerian approach to analyse labour earnings outcomes associated with schooling across the different employment segments for men and women separately. In other words, we wish to know, separately for men and women, whether schooling and post-schooling investments yield equivalent returns in three employment segments for which reliable earnings data are available: the public wage sector, private wage sector and self-employment.

Aiming for more flexibility Equation (4) is amended in order to estimate separately the Mincerian returns to some primary, full primary, some secondary, full secondary and some tertiary or further education. In other words, rather than using continuous education this section employs dummy variables for the level of education completed. Experience is approximated by age. The dependent variable is log of hourly earnings. Apart from education and age, the regressions also include dummies to control for place and province of residence. The omitted categories are, respectively, urban and Maputo City. Control dummies for occupational choice and industry were also initially included in the regressions. However, due to the relatively small sample size resulting from splitting the labour market across sectors and gender, the number of individuals in some occupations and industries was too small to produce reliable estimation results. Therefore, these controls were dropped from the regressions.

Ordinary least squares (OLS) estimates from earnings equations for men and women in public wage employment, private wage employment, and self-employment are presented in Table 3.9 on page 107. The education point estimates are in line with the descriptive statistics reported above in Table 3.2. Overall they are positive and increase with attainment across all sectors.

To better scrutinise the impact of education on earnings the percentage changes in earnings associated with educational attainment are calculated and reported in Table 3.10 on page 107. In semilog specifications such as these where education is represented by dummies for each level of completion, the percentage change in earnings associated with schooling is calculated as $100*(e^c - 1)$, where c stands for the education dummy coefficient at each level (Glick & Sahn, 1997, p. 819). Across all labour market sectors and controlling for other factors, hourly earnings of men and women with only some primary education do not differ significantly from uneducated men's and women's hourly earnings, except for women in the public sector, whose hourly earnings increase by 24%. The effects of schooling on hourly earnings matter only from full primary completion onwards, except for women in the private sector, for whom primary school does not significantly increase hourly earnings, and self-employed women for whom secondary education and above does not put them in a significantly better off position compared to uneducated women. We should bear in mind, however, that only roughly 0.1% of women in the self-employment segment have completed secondary education or more. The small sample size might make it impossible to draw distinctions of that sort.

Returns to full secondary schooling are much higher in the private sector, particularly for women. Controlling for other determining factors of earnings, men and women in this sector earn, on average, five and nine times better than their uneducated counterparts, respectively. In contrast, both men and women with full secondary school in the public sector, and men in self-employment, earn only four times better than men and women without any formal schooling, in the respective sector. In the private sector, women with some or completed tertiary education[36] have higher returns to education than men at the same education level. While women's hourly earnings at this level are fifteen times that of women with no formal schooling, men's hourly earnings are just seven times that of uneducated men. In the public sector, things are in favour of men, but more balanced. Tertiary educated men and women get, respectively, eightfold and sevenfold the hourly earnings uneducated men and women get.

Going back to Table 3.9, the concavity of experience proxied by age is seen across all sectors, but it is statistically significant at the conventional levels only for men. This means that men's hourly earnings increase with experience but at a decreasing marginal rate. Concerning the control variables, the rural dummy is in most cases negative but statistically significant for men in the private sector only. Thus only in this sector, rural individuals (men) earn less than their urban counterpart, on average. The insignificant rural dummy in the public sector might be associated with the fact that in this sector jobs as well as pay scales are centrally fixed. The provincial dummies indicate that in general hourly earnings are significantly higher in Maputo City (the omitted category), except for women in the private wage sector and self-employment in Tete province, and self-employed women in Inhambane and Maputo Province who tend to earn significantly better than Maputo City's women. The R-squared is reasonably high in the private and public wage sectors, indicating good model fit, but it is relatively low in the self-employment segment.

[36] To achieve an acceptable sample size, the dummy on tertiary education was set equal to 1 for those who completed at least 13 years of education, i.e., more than secondary education completed. However, this has the disadvantage of not differentiating returns to, say, a completed degree from returns to just one year of university education.

Table 3.9 – OLS earnings functions

Employment Sectors	Public Sector		Private Sector		Self-Employment	
Variables	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Education Level Completed						
Some primary	0.2774	0.2171**	0.0874	0.1199	0.1428	–0.0072
Full primary	0.5258**	0.5998***	0.2882***	0.3144	0.3082**	0.4719***
Some secondary	0.9131***	0.9395***	0.6968***	0.9556***	0.5748***	0.7382***
Full secondary	1.3780***	1.4358***	1.6370***	2.2018***	1.3757***	0.9469
Tertiary education	2.0959***	2.0157***	1.8905***	2.7356***	1.9004***	0.9803
Age	0.0368**	–0.0029	0.1134***	0.0496*	0.0918***	0.0307
Age squared	–0.0003*	0.0002	–0.0012***	–0.0003	–0.0011***	–0.0004
Rural	–0.0582	0.0169	–0.1835**	–0.1957	–0.1317	0.1210
Province						
Niassa	–0.2352**	–0.0215	–0.4912***	–0.2241	–0.6949***	–0.2937
Cabo Delgado	–0.0046	–0.1854	–0.0772	0.2243	–0.4719***	–0.1254
Nampula	–0.0969	–0.0829	–0.7262***	–0.3237	–1.1367***	–0.8512***
Zambezia	–0.1855**	–0.1241	–0.4158***	–0.4657	–0.8890***	–0.1183
Tete	–0.1178	–0.0495	–0.3222***	0.4213*	0.0617	0.7098***
Manica	–0.2048**	0.0541	–0.2107**	0.3304	–0.5008***	–0.1960
Sofala	–0.1927**	0.0314	–0.3412***	–0.2016	–0.5610***	–0.1372
Inhambane	–0.0808	–0.0860	–0.0740	–0.5407***	–0.2599*	0.3124***
Gaza	–0.1497*	–0.0022	–0.1669*	–0.2097	–0.0355	0.1894
Maputo Province	–0.0726	–0.1742	–0.0357	0.0138	0.0347	0.8195***
Constant	0.9486***	1.6180***	–0.6707***	–0.0274	0.5824*	0.7829*
No. of observations	1,053	438	2,177	440	2,455	1,655
R-squared	0.4607	0.4249	0.4216	0.6351	0.1477	0.1476

Note: dependent variable is log of hourly earnings; model estimation accounts for survey design; * p<0.1, ** p<0.05, *** p<0.01

Table 3.10 – Percentage change in earnings associated with education compared to no education

Employment Sectors	Public Sector (%)		Private Sector (%)		Self-Employment (%)	
Categories	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Some primary	32	24**	9	13	15	-1
Full primary	69**	82***	33***	37	36**	60***
Some secondary	149***	156***	101***	160***	78***	109***
Full secondary	297***	320***	414***	804***	296***	158
Tertiary education	713***	651***	562***	1442***	569***	167

Note: calculated from the table above as $100*(e^{c} - 1)$, where c stands for the education dummy coefficient at each level; * p<0.1, ** p<0.05, *** p<0.01

3.5.2 Multinomial Logit Earnings Functions

The previous section assumed that workers were randomly assigned to each employment segment. If this is not the case, however, and OLS regression techniques are employed, there is potential for

sample selection bias. The coefficients of the earnings functions will be biased. This occurs because there might be unobserved worker characteristics that affect both the determination of employment sector and earnings, i.e. there is a correlation between the error terms of the earnings functions and the process that determines employment sector choice. If this correlation is statistically significant then the OLS returns to education reported in Table 3.9 will be biased.

To get around sample selection bias studies often employ Heckman's (1979) two-step procedure. The first step involves modelling the probability of observing the outcome variable, say, the probability of being wage employed (i.e., the selection equation). The wage employment estimates are used to derive estimates of the inverse Mills ratio (lambda λ). The lambda is then included in the right-hand side of the wage equation (earnings function), thus making it conditional on selection into wage employment. Statistical significance of lambda indicates the presence of sample selection bias and that the wage equation needed correction for this bias. Heckman's two-step procedure, however, allows for binomial outcomes only. The individual is either wage employed or not. Given the multinomial nature of our particular employment selection model a different procedure that takes this multitude of options into account must be applied instead.

Bourguignon, Fournier & Gurgand (2004), in work later formally published as Bourguignon, Fournier & Gurgand (2007) in the *Journal of Economic Surveys*, reviewed the set of methods available in the literature on selectivity bias correction using multinomial specifications and used Monte-Carlo experiments to test their performance. According to them, Lee (1983) proposed an approach whereby Heckman's (1979) two-step procedure is generalised and selectivity is modelled as a multinomial logit. In the first step a multinomial logit model is employed and used to estimate the probabilities of falling into each of the different employment types. Then inverse Mills ratios are calculated for each employment segment and are included in the wage equations. Lee's approach is simple and correction for selectivity requires the estimation of one parameter only for each employment type. But they point out that that simplicity comes at the cost of very strong assumptions (Bourguignon, Fournier & Gurgand, 2007, p. 177).

Dubin & McFadden (1984) also proposed sample selection correction based upon multinomial logit models, in which M – 1 correction parameters are added to each OLS earnings function, where M stands for the number of employment sector choices. In contrast to Lee (1983), they make no assumption regarding the correlation of the unobservable determinants of the wage and selection

equations, but their assumption regarding the linearity of the wage equation's error term restricts the class of allowed distributions for this error term (Bourguignon et al., 2004, p. 7). This might be problematic when the independence of irrelevant alternatives (IIA) property of the multinomial logit model is not satisfied (Koch & Ntege, 2008). The IIA assumption states that the ratio of the probabilities of any two alternatives being chosen is independent of the number of available alternatives.

According to Dahl (2002), whenever the multinomial logit model is warranted for sample selection correction, a large number of correction parameters has to be estimated, thus making practical implementation intractable. Therefore, he proposed a non-parametric approach wherein the set of probabilities used to correct the wage equation for selectivity bias is restricted to a chosen subset of particular interest. He went further and proposed a special case of his approach in which the probability of falling into, say, wage employment in the public sector, is the only information required to correct the public sector OLS earnings functions for sample selection bias. In contrast to Lee's (1983) approach, Dahl's allows for any sign structure between the correlation of the error terms of the wage and selection equations (Bourguignon et al., 2004, p. 9). But in this procedure the correction terms have no structural interpretation (Koch & Ntege, 2008, p. 18).

Bourguignon et al. (2004) suggested a modification of Dubin & McFadden's (1984) model which allows the error term of the wage equation to be normal. Their Monte-Carlo experiments comparing all of the surveyed approaches to correct for sample selection bias suggest that Dubin & McFadden's (1984) model performs relatively better. But, when the IIA assumption of the multinomial logit model is not satisfied, Bourguignon et al.'s (2004) version of the model is to be preferred (Bourguignon et al., 2004, p. 17).

3.5.2.1 The model

Our preferred approach to correct the OLS earnings functions for selectivity bias is the modified version of Dubin & McFadden's (1984) model. Below we formally describe the estimation problem and correction procedure. Consider the following two models:

$$y_1 = x\beta_1 + \mu_1 \tag{5}$$

$$y_j^* = z\gamma_j + \eta_j, \quad \text{j} = 1 \ldots \text{M} \tag{6}$$

where the vector x contains all determinants of the variable of interest y_1 (which in our case stands for hourly earnings within a specific sector), and the vector z denotes the observable exogenous

factors influencing the choice of the employment sector. The terms μ_1 and η_j are idiosyncratic shocks. The former satisfies the following conditions: $E(\mu_1|x,z)=0$ and $V(\mu_1|x,z)=\sigma^2$. The subscript j is a categorical variable describing the assignment of the workers among M labour market sector alternatives.

A worker falls into option j among M alternatives to maximise y_j^*, which is a latent variable capturing the discrete observation of whether he or she works in that particular sector or not. In other words, hourly earnings in, say, the public sector are only observed if the individual is assigned to work there, and this only happens if the individual utility derived from working in that sector is higher than the utility derived from working in alternative sectors. Formally, y_1 is only observed if:

$$y_1^* > \max_{j \neq 1} y_j^* \tag{7}$$

If we define:

$$\varepsilon_1 = \max_{j \neq 1}(y_j^* - y_1^*) \tag{8}$$

$$\varepsilon_1 = \max_{j \neq 1}(z\gamma_j + \eta_j - z\gamma_1 - \eta_1),$$

it implies that ε_1 should be negative. Assuming that the η_j's follow the IIA property, McFadden (1973) has shown that the discrete employment sector determination component can be consistently estimated with a multinomial logit model. Simply put:

$$P(\varepsilon_1 < 0|z) = \frac{e^{z\gamma_1}}{\sum_j e^{z\gamma_j}}$$

Bourguignon et al. (2004) stated that while estimating consistent parameters (γ_j) of the latent variable model (6) is straightforward, the problem lies in estimating the parameter β_1 of the wage equation (5). The disturbance terms of the wage equation (5) and selection equation (6) may not be independent. As a result the error term and explanatory variables of the wage equation (5) become correlated, leading to inconsistent OLS estimates of β_1 (Bourguignon et al., 2004, p. 4).

Given the above estimation problem, Bourguignon et al. (2004) proposed the following assumption regarding the correlation (r_j^*) between the error terms of the wage and selection equations:

$$E(\mu_1|\eta_1 \dots \eta_M) = \sigma \sum_{J=1 \dots M} r_j^* \eta_j^* \tag{9}$$

With the help of this assumption, they show that the parameter β_1 of the wage equation (5) conditional on choosing employment sectors j=1 can be consistently estimated by the following equation:

$$y_1 = x_1\beta_1 + \sigma\left[r_1^* m(P_1) + \sum_{J=2\ldots M} r_j^* m(P_j)\frac{P_j}{(P_j-1)}\right] + w_1, \quad (10)$$

where w_1 is an independent error term, $P_1$the probability that alternative 1 will be preferred, and $m(P_j) = \int J(v - log\, P_j)g(v)dv,\ \forall j.$ This method adds to each wage equation selection correction terms corresponding to the number of employment segment alternatives. Unlike Heckman's (1979) procedure whereby there is one choice involved only, in this case there are many employment segment alternatives, which result in as many selection correction terms as there are alternatives. These terms are consistent estimators of the conditional expected values of the residuals derived from the first stage multinomial selection equation. In the notation of the Heckman two-step procedure, the coefficients on these variables constitute the lambdas.

Parametric identification of the selection equation requires the addition of exclusion restrictions, i.e., variables that affect the chances of employment in a particular sector but not hourly earnings. This study uses four such variables: the number of children in the household, the number of adults in the household, marital status and household headship status of the individual.

3.5.2.2 Empirical results

Like in the case of the standard OLS earnings functions reported above this work estimates separate models for men and women in the public wage employment, private wage employment, and self-employment. In the first stage (selection) equations this study considers those working in the above mentioned sectors with or without earnings data[37] as well as unemployed individuals[38]. Once the selection correction terms are extracted they are included in the second stage (wage) equations, thus correcting them for sample selection bias.

The models are estimated using a Stata 12 (Statacorp, 2011) user-written ado-file named SELMLOG developed by Bourguignon et al. (2004), and results are reported in Table 3.11[39] on page 113. Within the public sector the selection correlations reported at the bottom of the table show that the unobserved determinants of hourly earnings are correlated in a statistically significant manner with the unobservables of both the process that determines selection into public sector and

[37] This introduces another type of sample selection bias into the regression estimates, since unobserved characteristics of those who reported earnings might differ from those who did not. Nevertheless, for the sake of avoiding adding further complications to the analysis, this chapter ignores this issue.

[38] Non-participants were excluded since they do not participate actively in the labour market. Unpaid family workers were excluded because none of them had tertiary education. Including them in the regressions would result in (unreliable) zeros for the coefficient estimates of the "tertiary education" dummy.

[39] Only the OLS selectivity corrected earnings functions are reported, since results and interpretations of the selection equations are qualitatively similar to those seen above in Table 3.5.

into unemployment, for men, and with the process that determines selection into self-employment, for women. Within the private sector, the earnings function's error term is significantly correlated with the self-employment equation's error term. For women the unobserved determinants of hourly earnings are significantly correlated also with the unobserved determinants of public sector employment. The selection correlation terms within the self-employment sector are not statistically significant at the conventional levels of significance, except for men, for whom the unobserved determinants of earnings are significantly correlated with the unobserved determinants of the process that determines selection into unemployment.

Thus, the model results indicate that across the three employment sectors and for both sexes, except for self-employed women, sample selection bias was indeed a problem and was corrected for by the inclusion of selectivity correction terms[40]. Sample selection bias was not found to be a major problem in the earnings function corresponding to self-employed women. Hence, the hourly earnings of a woman with average characteristics in the self-employment sector would not differ significantly from the hourly earnings of a woman randomly selected into the various sectors. Therefore, in this case the OLS estimates can be trusted.

The schooling estimates show a similar pattern as the OLS ones, whereby educated individuals earn significantly better than uneducated ones across all labour market sectors. As before, percentage changes in earnings associated with educational attainment are calculated and reported in Table 3.12 on page 114. Compared to a similar table in the OLS earnings functions case, this table shows that the correction for selectivity bias resulted in greater returns to education for men and women in the public sector and for men only in the private sector, but also resulted in slightly smaller returns to education for women in the private sector and for self-employed men. Further, for men even some primary education provides greater returns than no schooling. Furthermore, men's returns to education are greater than women's within the public sector, but the inverse occurs within the private sector, at least from secondary education onwards. Lastly, returns to education seem to be convex. Having fifteen years of education completed[41] compared to having no education is associated with greater earnings gains than having completed twelve years of education only

[40] The fact that the correction terms for men´s earnings functions within the private sector and within self-employment are not statistically significant might suggest that there selectivity bias was not too much of a problem. In fact, this is also supported by the selection correlations within the respective earnings functions which are only statistically significant at the 10% level of significance.
[41] At the tertiary level mean years of education completed is 15.

(secondary level). In turn, having twelve years of education completed compared to no education is associated with greater earnings than having completed seven years of education only (primary level). The benefits of tertiary education, however, can only be reaped if people have the chance to access and complete the lower levels of education.

Table 3.11 – Selectivity-corrected earnings functions

Employment Sectors	Public Sector		Private Sector		Self-Employment	
Variables	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Education Level Completed						
Some primary	0.7722***	0.3462	0.1458*	0.0944	0.2276**	0.0202
Full primary	1.4272***	0.9898***	0.3634***	0.3657*	0.5080***	0.1671
Some secondary	2.3630***	1.6164***	0.9070***	0.9414***	0.6176***	0.4087
Full secondary	3.0688***	2.3879***	1.9226***	2.1775***	1.2185***	0.4208
Tertiary education	3.7752***	3.0692***	2.2387***	2.7297***	1.5557***	0.1877
Age	0.0795**	– 0.0048	0.0975***	0.0468	0.0625***	0.0641***
Age squared	– 0.0006	0.0002	– 0.0010***	– 0.0003	– 0.0007**	– 0.0007**
Rural	– 0.2440	– 0.3590**	– 0.3530***	– 0.0873	– 0.2098	0.2802
Province						
Niassa	– 0.0577	– 0.2404	– 0.8014***	– 0.3018	– 0.4968	0.3190
Cabo Delgado	0.1131	– 0.2266	– 0.2540	– 0.0994	– 0.4126	0.3217
Nampula	– 0.1009	– 0.1921	– 0.9328***	– 0.4711**	– 1.0766***	– 0.6708***
Zambezia	– 0.1677	– 0.3426	– 0.7364***	– 0.3339	– 0.8012***	0.2540
Tete	– 0.1588	– 0.0842	– 0.4199***	– 0.0144	0.0206	0.7408***
Manica	– 0.1423	– 0.1659	– 0.4594***	0.0714	– 0.4176*	0.2963
Sofala	– 0.3089***	– 0.2129	– 0.4203***	– 0.3473**	– 0.6021***	0.0236
Inhambane	– 0.0235	– 0.1475	– 0.2344**	– 0.8638***	– 0.2113	0.6015***
Gaza	– 0.1097	– 0.0583	– 0.2120**	– 0.5200***	0.1257	0.5314***
Maputo Province	0.0238	– 0.1243	– 0.0229	– 0.1023	0.0097	0.8282***
Public sector correction term	0.9964***	0.2110	0.2742	– 0.8867*	0.1985	– 0.3612
Private sector correction term	– 0.6765	0.1392	0.1405	0.1604	0.4798	– 0.5464
Self-employment correction term	0.5290	– 0.8380**	– 0.7063	– 0.7825**	0.5153	0.3475
Unemployment correction term	2.0820***	0.6031	0.2352	– 0.2156	1.2449	– 0.4262
Constant	– 2.2262	0.8200	– 0.6877	– 0.5869	1.0905**	– 0.5702
Selection Correlations						
Public sector	0.7655***	0.3018	0.3213	– 1.2041**	0.1357	– 0.2764
Private sector	– 0.5198	0.1991	0.1647	0.2178	0.3281	– 0.4182
Self-employment	0.4065	– 1.1984**	– 0.8278*	– 1.0626**	0.3524	0.2659
Unemployment	1.5997***	0.8626	0.2757	– 0.2928	0.8513*	– 0.3262

Note: dependent variable is log of hourly earnings; Statistical significance based on bootstrapped standard errors (1,000 repetitions); * p<0.1, ** p<0.05, *** p<0.01

Table 3.12 – Percentage change in earnings associated with education

Employment Sectors	Public Sector (%)		Private Sector (%)		Self-Employment (%)	
Categories	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Some primary	116***	41	16*	10	26**	-1
Full primary	317***	169***	44***	44*	66***	60
Some secondary	962***	403***	148***	156***	85***	109
Full secondary	2,052***	989***	584***	782***	238***	158
Tertiary education	4,261***	2,052***	838***	1,433***	374***	167

Note: Figures derived from selectivity-corrected earnings functions are underscored and are calculated from the table above as $100*(e^c - 1)$, where c stands for the education dummy coefficient at each level; the remainder of the figures are the same as in Table 3.10, calculated from the standard OLS earnings functions; * p<0.1, ** p<0.05, *** p<0.01

3.6 Conclusion

The objectives of this chapter were twofold. First it aimed to identify the impact of schooling and other characteristics on the chances of falling into the different labour market employment segments in Mozambique. Then, it aimed to link education to earnings in a Mincerian type approach.

Mozambique is no different from other poor developing countries such as Kenya, Ethiopia, Guinea and others, where a big slice of the working-age population works outside the wage sector. In such countries, self-employment, unpaid family work, informal sector employment, subsistence agriculture employment and the like are more common than employment in the public and private wage sectors, where the latter two in normal circumstances are rationed and tend to be the preferred employment. Therefore analysing the impact of schooling on the chances of employment and earnings in the wage sector only is nonsensical. Instead, the analysis done in this chapter focused in a number of labour market sectors for which reliable data are available. Indeed, as Vijverberg (1993, p. 937) postulated, "when measuring market returns to education, one should consider how the labour market operates".

Using the Mozambican Labour Force Survey conducted over the years 2004 and 2005, and employing a multinomial logit model, firstly, it was found that for both men and women schooling relative to no formal education increases the chances of employment in the public wage sector. Within the private wage sector schooling was found to be important for women only, though men's chances of employment were found to be greater at all education levels. For both genders schooling was found to lower the chances of unpaid family work and self-employment. With regards to the

latter, the poorly educated generally find it difficult to access formal employment in the public and private sectors and hence choose self-employment, the only or most likely alternative option available to those wanting to work so as to earn income and hence improve their living standards.

The above findings are not uncommon to what is found in similar studies in the sub-Saharan African context. All of the surveyed studies concluded that schooling increased the chances of wage employment, particularly in the public sector, and lowered the chances of self-employment/informal/agricultural jobs and unemployment, but it could not always guard against unemployment. Glick & Sahn (1997) and Wambugu (2002) also found that education increased the chances of private sector jobs for women only.

Secondly, education has a bearing on earnings across all sectors for which reliable earnings data are available, and its impact on earnings is convex, i.e., returns to education are greatest at the tertiary education level, followed by secondary schooling and then primary schooling. As pointed out by Kuepie, Nordman & Roubaud (2007), this means that investment in primary schooling in these situations might be worthwhile to reduce poverty and inequality if those investing in it can continue their studies up to secondary or even tertiary education. In other words, the role of education in improving well-being grows with educational attainment.

Within the public wage sector, private wage sector, and self-employment sector, men and women with high levels of educational attainment are paid significantly better than uneducated men and women, respectively. Returns to education are greater for women than men in the private wage sector, but greater for men than women in the public sector. For individuals in self-employment secondary education or further schooling matters significantly for men only, suggesting that self-employed women engage in activities for which advanced education attainment does not bring much more income. We should bear in mind, however, that sample sizes for self-employed women with secondary education completed or more are very small and this deficiency might be driving the findings.

The reviewed studies were also generally consistent with these findings. For instance, very often returns to education were found to be greater for women than men. Thus, obstacles constraining women's participation in the labour market should be addressed by means of appropriate policy interventions, in this way addressing the issues of poverty and inequality in Mozambique. One example of such policy interventions includes programmes targeted at women, aimed to increase

access as well as quality of education. This could help address the existing gender imbalances with regards to skills valued by the labour market.

Programmes aimed at fostering the education of women are already taking place in Mozambique. For instance, *Fundação para o Desenvolvimento da Comunidade* (FDC), a civil society organisation which aims to promote community development, democracy and social justice, has developed a scholarship programme targeted at women. The scholarship holders are selected based on financial need and academic merit.

Thus, education is important since it increases the chances of wage employment in the public sector, which in developing countries such as Mozambique can be considered the preferred employment. Educated men and women would rather not participate or stay unemployed while queuing for the possibility of getting into the public wage sector. However, only a very small sample of the working-age population is educated enough to enter this sector. Only those few who can get enough schooling end up benefiting from it. The remainder of the working-age population resorts to activities outside the wage labour market. But even within these less attractive sectors education leads to greater earnings.

The above paragraphs summarise this study's role in promoting the addition of new knowledge. Education does indeed impact differently on the chances of employment and earnings within the different labour market segments that characterise sub-Saharan African developing economies such as Mozambique.

While in this chapter we get around sample selection bias in the face of multiple employment statuses, the earnings functions estimates might be prone to another type of bias originating from the omission of some important variables – the omitted variable bias. Among these variables the related literature devotes much attention to ability and quality of education. Ability, for instance, is believed to affect both education and earnings positively, therefore creating an upward bias in the estimates of returns to education. These variables are very uncommon in the type of data sets used in this study, notably in the context of developing countries. Correcting these earnings functions in particular for ability bias following an Instrumental Variable (IV) approach is one avenue for further research.

If appropriate actions are not taken the potential for education to provide access to better jobs and better earnings is likely to decline over time. The Education For All movement which, among other things, fosters the increase in access to education, had significant impacts on the Mozambican education system. As will be seen in the next chapter, the expansion of access to education in the last decade has been a success story. For instance, between the year 2000 and the year 2007 the number of grade 6 pupils in the Mozambican schools increased from about 120,000 to about 320,000. However, it might have come at the cost of education quality. If education quality declines over time, educational attainment becomes insufficient to secure better jobs and earnings, as employers will prioritise better skilled individuals. Further, if education quality deteriorates disproportionally across regions it might exacerbate the patterns of poverty and inequality. Thus, understanding the factors behind the change in the quality of education over time is important to strengthen the relationship between education, labour market, and poverty and inequality alleviation in Mozambique. This issue is therefore addressed in Chapter 4.

Chapter 4

Understanding Cognitive Achievement at the Primary Level of Education in Mozambique

4.1 Chapter Introduction

In 1975 Mozambique became politically independent from Portugal, but subsequently experienced a civil war that extended until 1992, when the belligerent parties signed a peace agreement. The civil war worsened the poverty and inequality of the Mozambican population and also made impracticable any initiatives to make access to education equal across the different regions of the country, as the war was more intense in the Central and Northern provinces, with widespread destruction of infrastructures, including the school network. For instance, as indicated in Table 4.1, respectively 95%, 88%, and 69% of the existing schools in 1983 in the provinces of Tete, Zambezia and Niassa were destroyed or closed by the end of the civil war, while in the capital (Maputo City) all schools remained operational. Nationally, only about 42% of the school network existing in 1983 was still operational in 1992.

Table 4.1 – The impact of the civil war on the destruction of the school network

Province	Schools		
	Existing in 1983	Destroyed or closed by 1992	
		Number	%
Tete	479	454	94.8
Zambezia	1,130	997	88.2
Niassa	508	352	69.3
Sofala	386	253	65.5
Maputo Province	339	204	60.2
Manica	225	109	48.4
Nampula	1,116	535	47.9
Inhambane	506	220	43.5
Gaza	546	169	31.0
Cabo Delgado	542	109	20.1
Maputo City	109	0	0.0
Total	**5,886**	**3,402**	**57.8**

Source: Adapted from Ministry of Education (1996, p. 40)

Various stakeholders, however, argue that education increases the chances to escape poverty and if equally distributed amongst the population diminishes the inequality gaps across the various socio-economic groups. For example, education impacts on poverty through the labour market. Educated people are more likely to participate in the labour force than those without formal education. Education permits accessing better jobs and better earnings. In developing countries such as

Mozambique the best jobs are in the public sector and the more educated have better chances of accessing these jobs. Even within the self-employment segment, which is often negatively associated with education (i.e., schooling lowers the chances of falling into self-employment activities such as subsistence agriculture and informal activities), education has the potential to produce better earnings. Therefore, if disadvantaged groups such as people in the civil war's most affected provinces have fewer opportunities to access quality education they will also have fewer opportunities of escaping poverty. Those in Maputo City, however, who were barely affected by the war and who have access to the best schools, have better chances of escaping poverty.

With the end of the civil conflict in 1992, the Mozambican government has made large efforts to expand the coverage of the school network in order to increase access to education. For instance, between 2000 and 2007 the number of grade 6 pupils in Mozambican schools increased from 122,338 to 319,243, an increase of about 161% (see Table A 6 in the Appendix). Those provinces that due to the civil conflict were underserved experienced a substantial increase in enrolments. In Niassa, for instance, grade 6 enrolments increased by about 355%, from 4,660 to 21,184 pupils, and in Zambezia enrolments increased by 238%, from 13,722 to 46,396 pupils. In Maputo City, however, which is the least poor province and the only province that escaped destruction and closure of schools, the number of grade 6 pupils grew during the same period from 24,287 to 30,429, corresponding only to a 23% increase. These figures suggest that the expansion of access to education in Mozambique in the last decade has been a success story.

Handa & Simler (2005) used the first income and household survey (IAF1996-97) to analyse the quantity versus quality of education trade-off in Mozambique. They interrogated whether budget-constrained governments such as the Mozambican should focus on increasing the number of schools or alternatively on improving the quality of those schools already open. They argued that when access to education is low (as was the case in Mozambique) a stronger focus should be placed on quantity. They concluded that "when cost information is considered the main policy implication is that the expansion of school quantity through well-targeted placement of new schools will provide the most cost-effective increase in educational outcomes at this time" (Handa & Simler, 2005, p. 59). By 'this time' the authors meant the late 1990s. This issue will be further discussed in this chapter's conclusion.

While the increased access to education is in itself positive, the rapid expansion of the education system brought costs in terms of average pupil performance. The average quality of education has deteriorated. Thus this chapter studies the performance of the national education system in terms of pupil educational outcomes.

The study of the performance of the national system of education requires the use of data representative of the entire country. For the case of Mozambique, the Southern and Eastern Africa Consortium for Monitoring Education Quality (SACMEQ) is so far the only entity that collected relevant data representative for the country as a whole. SACMEQ consists of 15 ministries of education in 14 countries located in the Southern and Eastern Africa, namely, Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Uganda, Zambia and Zimbabwe. Its mission, among other things, is to undertake integrated research and training activities in order to generate information that can be used by policymakers to plan and improve the quality of education. The SACMEQ study basically involves the collection of a variety of data at the pupil (grade 6), teacher and school levels, including for instance, information on background characteristics, access to school resources, and access to public services. As part of the study both pupils and teachers were required to write three standardised tests on reading, numeracy and HIV/AIDS knowledge[42].

Three international surveys were already implemented by the SACMEQ consortium, but Mozambique participated only in the last two (hereafter SACMEQ II and SACMEQ III), which took place in the years 2000 and 2007. In 2000 the Mozambican pupils' performance was above the SACMEQ average (which was set to 500 with a standard deviation of 100). They scored 517 on the reading test and 530 on the numeracy test, and occupied the seventh and fourth positions in the SACMEQ ranking. In the latest survey, however, the average performance of the Mozambican grade 6 pupils fell to 476 and 484 in reading and numeracy, pulling down the country´s average performance to the bottom places of the country ranking, signalling deterioration in average education quality of those pupils tested.

This study hypothesises that such quality deterioration was caused by a massive influx of pupils from disadvantaged socio-economic backgrounds as well as the consequent pressure on existing school resources. Poor pupils are likely to fare worse than their rich counterparts since they often

[42] The HIV/AIDS test was only included as from the last survey (SACMEQ III).

have to walk long distances to school, have poor health and nutrition, lack learning materials and get limited help with school work at home. The increased number of pupils causes pressure on existing resources, given that the supply of school factors such as school infrastructures and teachers grows slower than school demand. The new school resources such as teachers are likely to be of lower quality themselves.

Another plausible hypothesis, but difficult to test with the data used in this work, is that standards have declined over time. In other words, efforts aimed at the rapid expansion of the education system resulted in lenient promotion of lower ability pupils. Pupils that previously would have been retained at lower grades unless they demonstrated command of the relevant materials for those grades were allowed to progress to higher grades.

This chapter aims to isolate the school, teacher and pupil characteristics that consistently predict education outcomes in Mozambique, and then to identify the factors behind the sharp decline in average reading and numeracy test scores over time. To identify the factors associated with school effectiveness, this work employs the production function approach that looks at the school as a production facility responsible for generating educational outputs. Inputs are the pupil, teacher and school characteristics related to educational outcomes. Concerning the factors behind the sharp decline in average pupil reading and numeracy test scores over time this work makes use of the Oaxaca-Blinder decomposition technique often employed to study labour market outcomes by decomposing earnings across groups (e.g. gender, race) into a component explained by productive differences and an unexplained component normally attributed to discrimination (Oaxaca, 1973; Blinder, 1973). In the particular case of this study we are interested in explaining the differential in pupil mean scores across surveys. Therefore this work partitions the score differential into its constituent components using the SACMEQ surveys as the group variable. This method will in principle indicate whether the decline in education quality over time was caused by the decline in the endowments of productive characteristics related to achievement or by the decline in how these characteristics are "remunerated" over time[43].

[43] The terms "productive characteristics" and "remuneration" of these characteristics are more appropriate for the labour literature. This chapter nevertheless is going to transfer that phrasing to the education research.

4.2 School Effectiveness Research

The research on school effectiveness was pioneered in the industrialised countries and can be split in three waves, according to Riddell (2008, p. 10). The first wave, which is traced back to the 1960s and early 1970s, utilised production functions to measure the impact of inputs on outputs using ordinary least squares (OLS) regression models. Inputs included individual variables such as pupil's family background and school factors, among other things, while outputs included variables related to pupil academic achievement. The general belief was that school factors had no discernible impact on pupil achievement. Achievement was determined by pupils' family background.

A second wave of the school effectiveness research took place from the late 1980s and focused on process variables. It moved away from the economist's perspective fixed on inputs that generate outputs to the study of what happens inside the classroom, i.e., what teaching styles generate better school outcomes. The move towards approaches more based on educational theory was justified by the fact that education production functions failed to ascertain statistically significant relationships between pupil academic achievement and common sense variables such as class size, school resources and teacher's education attainment, among other things (Riddell, 2008, p. 10).

The third wave also begun in the late 1980s and in a way went back to the education production function approach. But in contrast to the first wave that used OLS regression models, the third wave employed multilevel regression models thought to be more realistic in exploring the factors determining academic achievement. Unlike the OLS models, multilevel regression models take into account the clustering existing in any education system. Pupils of varying backgrounds are taught in different classes, by different teachers, in different schools, with different resources, and in different parts of the country. While OLS assumes that such clustering does not influence academic achievement, the multilevel approach assumes that clustering does have differential impact on educational outcomes.

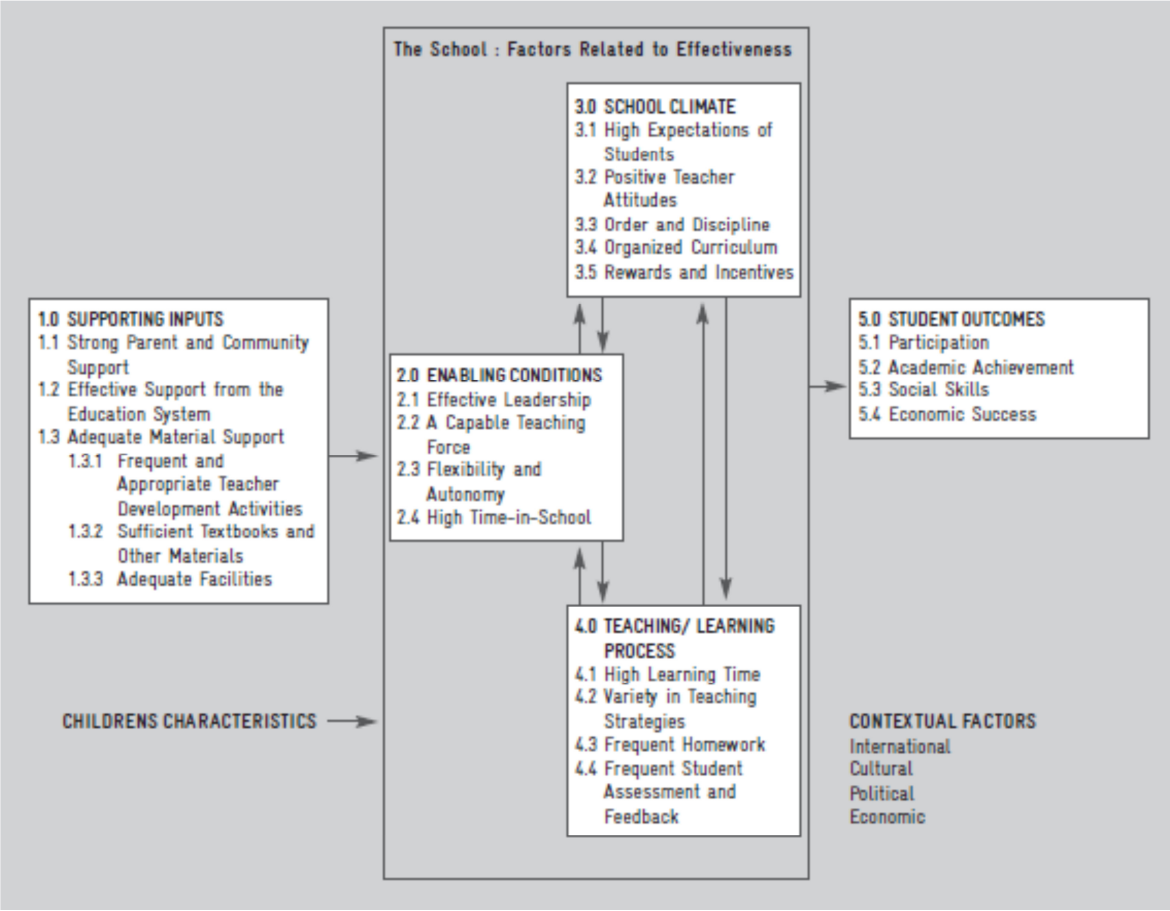
In developing countries, however, the bulk of the research on education effectiveness focused on the first and third wave approaches. In contrast to the research results from the industrialised world, in developing countries the education production function approach led to the conclusion that pupil achievement could be influenced by policy, i.e., school factors did have a discernible impact on pupils' academic achievement.

Heneveld (1994) developed a conceptual framework of factors determining school effectiveness. It is reproduced below in Figure 4.1 on page 124. On the one hand there are pupil schooling outcomes which include participation and academic achievement, among other things. On the other hand there are supporting inputs, enabling conditions, school climate, teaching and learning processes, contextual factors as well as pupil characteristics believed to have a bearing on pupil schooling outcomes.

Riddell (2008) reviewed the research into educational effectiveness in the developing world. She based her work on Heneveld's (1994) model and presented a list of factors that the variety of studies in the developing world found to determine pupil school outcomes, grouping them according to Heneveld's conceptual framework. The list is reproduced in Table 4.2 on page 124. Out of Riddell's list of factors determining school effectiveness in developing countries, the data sets used in this chapter contain information on access to textbooks, class sizes, classroom and school amenities, pupils' nutrition, parental involvement, teacher supervision, teacher subject knowledge, teacher training, education and experience, gender of teacher, school head training, community involvement, school discipline, homework, grade repetition, assessment and feedback. Due to high levels of multicolinearity which lead to inconsistent and unexpected results, not all these variables were included in the models. Table A 7 in the Appendix lists and summarises the variables selected for this study.

Since this chapter also aims to explain the reasons behind the average deterioration in education quality over time, the next section reviews the methods employed to decompose the average fall in education quality.

Figure 4.1 – Conceptual framework of factors determining school effectiveness



Source: Heneveld (1994, p. 6)

Table 4.2 – Factors determining school effectiveness in the developing world

Supporting Inputs	Enabling Conditions	School Climate	Teaching/Learning Process
Textbooks	Teachers:	Community	Time on task
Class size	Subject knowledge Verbal	Teachers´ presence	Pedagogy
Distance	ability	Commitment	Maternal language
Class room/school	Language	Incentives	Reading
amenities	Pre-service and in-service	Status	Homework
Pre-school	education Pedagogical	Order	Assessment and feedback
Children´s health and	repertoire Experience	Discipline	Multi-grade
nutrition	Proximity	Goals	Ability grouping
Parental involvement	Gender	Improvement	Repetition
Community involvement		Curriculum	
Teacher supervision and	School head:	Standards	
involvement	Leadership	Expectations	
Standards – institutional	Supervision		
guidelines	Training time		

Source: Adapted from Riddell (2008, p. 48)

4.3 The Oaxaca-Blinder Decompositions

In his seminal paper Oaxaca (1973) studied the gender wage differentials in the United States' urban labour markets, as many previous studies had demonstrated that wage discrimination against women was widespread. Oaxaca defined a discrimination coefficient as

$$D = \frac{W_m/W_f - \left(W_m/W_f\right)^0}{\left(W_m/W_f\right)^0}, \tag{1}$$

where W_m/W_f was the observed male-female wage ratio and $\left(W_m/W_f\right)^0$ the male-female wage ratio in the absence of discrimination. In logarithm format Expression (1) is equivalent to

$$ln(D+1) = ln\left(W_m/W_f\right) - ln\left(W_m/W_f\right)^0 \tag{2}$$

In a non-discriminating labour market employers would minimise costs by equalising the male-female wage ratio in the absence of discrimination to the ratio of the male-female marginal products.

Estimating the male-female wage ratio existing in the absence of discrimination requires the satisfaction of either one of two assumptions: a) the wage structure currently faced by females would also apply to males; or b) the wage structure currently faced by males would also apply to females. The wage structure applicable to each gender is estimated by an OLS wage equation for each group and takes the form:

$$ln(W_i) = Z_i'\beta + u_i\,, \tag{3}$$

where W_i is the hourly wage rate of the i-th worker; Z_i' stands for a vector of individual characteristics; β is a vector of coefficients; and u_i is a disturbance term.

Oaxaca showed that his formulation of the coefficient of discrimination given by Expression (2) in conjunction with the two alternative assumptions a) and b) permitted decomposing the wage differential across male and female workers into a component attributable to differences in

individual characteristics and an unexplained component deemed discrimination. For that he defined the gender wage discrimination coefficient (G) as:

$$G = \frac{\bar{W}_m - \bar{W}_f}{\bar{W}_f} \tag{4}$$

then in natural logarithm format Expression (4) is equivalent to

$$ln(G+1) = ln(\bar{W}_m) - ln(\bar{W}_f) \tag{5}$$

$\bar{W}_m$ and $\bar{W}_f$ stand for the average hourly earnings for males and females, and are estimated using the least squares equation specified by Expression (3). Therefore, Expression (5) is equivalent to Expression (6) below:

$$ln(G+1) = \bar{Z}_m'\hat{\beta}_m - \bar{Z}_f'\hat{\beta}_f \tag{6}$$

After a couple of manipulations Oaxaca showed that

$$ln(G+1) = (\bar{Z}_m' - \bar{Z}_f')\,\hat{\beta}_f - \bar{Z}_m'(\hat{\beta}_f - \hat{\beta}_m) \tag{7}$$

Or alternatively

$$ln(G+1) = (\bar{Z}_m' - \bar{Z}_f')\,\hat{\beta}_m - \bar{Z}_f'(\hat{\beta}_f - \hat{\beta}_m) \tag{8}$$

In Expression (7) one assumes that in a non-discriminating situation the current female wage structure would apply to both sexes. The first part of the right-hand side of the expression, that is, $(\bar{Z}_m' - \bar{Z}_f')\,\hat{\beta}_f$ represents the component of the differential in wages associated with differences in endowments of individual characteristics between sexes, and the second part, $-\bar{Z}_m'(\hat{\beta}_f - \hat{\beta}_m)$, stands for the estimated effect of discrimination.

Expression (8) the wage differential between groups is decomposed by assuming that the current male wage structure would apply to both males and females in the absence of discrimination. Thus $(\bar{Z}_m' - \bar{Z}_f')\,\hat{\beta}_m$ corresponds to the part of the wage differential associated with differences in characteristics across sexes, and $-\bar{Z}_f'(\hat{\beta}_f - \hat{\beta}_m)$ stands for differences on how the various characteristics are remunerated across sexes, or simply put, corresponds to discrimination.

At about the same period Blinder (1973) was also interested in the factors behind wage differentials in the United States. In his study he investigated the sources of wage differentials between white and black males and between white males and white females. To decompose the dispersion in

individual wages he started by estimating a wage equation for each of the two demographic groups he was interested in comparing:

$$Y_i^H = \beta_0^H + \sum_{j=1}^{n} \beta_j^H X_{ji}^H + \mu_i^H \tag{9}$$

$$Y_i^L = \beta_0^L + \sum_{j=1}^{n} \beta_j^L X_{ji}^L + \mu_i^L \tag{10}$$

where Y_i^H is the logarithm of wage rate for the high-wage group, Y_i^L the logarithm of wage rate for the low-wage group, X_{ji}^H and X_{ji}^L stand for the wage rate predictors for the high-wage and low-wage groups, and μ_i^H and μ_i^L the respective disturbance terms.

Unlike Oaxaca (1973) who employed a twofold decomposition, Blinder (1973) showed that the raw differential in the wage rate across, say, white males and black males could be split into three parts according to the following expression:

$$R = \beta_0^H - \beta_0^L + \sum_J \bar{X}_j^L (\beta_j^H - \beta_j^L) + \sum_j \beta_j^H (\bar{X}_j^H - \bar{X}_j^L) \tag{11}$$

In the above expression the first component $\beta_0^H - \beta_0^L$ stands for the unexplained portion of the differential in wages due to group membership per se (Jones & Kelley, 1984, p. 327), or the part of the differential in wages due to the "shift" coefficient, or due to differences between intercepts. The second component, $\sum_J \bar{X}_j^L (\beta_j^H - \beta_j^L)$, corresponds to the part attributable to differing coefficients. Together with the first component they are deemed the discrimination effect. The third component $\sum_j \beta_j^H (\bar{X}_j^H - \bar{X}_j^L)$ is the portion of the differential in the wage rate across groups attributable to differing endowments of productive characteristics. This portion of the wage differential, that is, the value of the advantage in endowments of the high-wage group, was evaluated from the viewpoint of the high-wage group. Blinder (1973, p. 438) acknowledged that the decomposition of the wage differentials across groups would run into the index-number problem, in the sense that alternative decompositions were also possible. For example, the wage differential could have been alternatively decomposed into:

$$\mathrm{R} = \beta_0^H - \beta_0^L + \sum_J \bar{X}_j^L \left(\beta_j^H - \beta_j^L\right) + \sum_j \beta_j^L \left(\bar{X}_j^H - \bar{X}_j^L\right) + \sum_j \left(\bar{X}_j^H - \bar{X}_j^L\right) \left(\beta_j^H - \beta_j^L\right) \tag{12}$$

In Expression (12), the portion $\sum_j \beta_j^L \left(\bar{X}_j^H - \bar{X}_j^L\right)$, the value of the advantage in endowments possessed by the high-wage group is instead evaluated using the wage equation of the low-wage group. Blinder (1973) disliked this alternative decomposition because while the first two

components of Expression (12) $\beta_0^H - \beta_0^L$ and $\sum_J \bar{X}_j^L \left(\beta_j^H - \beta_j^L\right)$ could be attributable to differing coefficients between the two groups of workers the interaction component $\sum_j \left(\bar{X}_j^H - \bar{X}_j^L\right)\left(\beta_j^H - \beta_j^L\right)$ had no easy interpretation (Blinder, 1973, p. 438).

Jones (1983), however, argued that the threefold decomposition proposed by Blinder (1973) was inappropriate. He argued that partitioning the discrimination residual term $\beta_0^H - \beta_0^L + \sum_J \bar{X}_j^L (\beta_j^H - \beta_j^L)$ into a portion due to differences in the coefficients (i.e., $\sum_J \bar{X}_j^L (\beta_j^H - \beta_j^L)$) and another due to differing intercepts across groups (i.e., $\beta_0^H - \beta_0^L$) produced arbitrary and uninterpretable results (Jones, 1983, p. 126). Using dummies for level of schooling he demonstrated that alternative choices of the omitted base category of schooling led to different values for each portion of the discrimination residual proposed by Blinder (1973). The issue still remains even when continuous variables are used as predictors. He showed that two decomposition models with equal predictive power of individual earnings, but with the only difference that one used age left school as the predictor of earnings across sexes and the other used years of education completed (which is equivalent to age left school minus 6), also led to significantly different values for each portion of the discrimination residual.

Jones & Kelley (1984) also raised a cautionary note on how the variety of studies in the related literature measured wages discrimination employing decomposition techniques. According to them Oaxaca's (1973) model was inappropriate because it excluded the interaction term as an individual component of the differential in the outcome variable, and attributed its effect to the difference in the coefficients. Blinder (1973), on the other hand, attributed the effect of the interaction term to the different endowments of productive characteristics. Jones and Kelley argued that neither of these two approaches were appropriate because "there is no unambiguous way of allocating [the interaction term] to either rates-of-return or to endowments" (Jones & Kelley, 1984, p. 329).

Jones and Kelley also examined a model by Winsborough & Dickenson (1971) which decomposed the earnings differentials into four components, namely, the part of the differential in earnings due to group membership per se, the portion due to differing coefficients, the part due to different endowments of productive characteristics, and the portion due to the interaction of differing productive characteristics and different returns to those productive characteristics across groups. The innapropriateness of this model rests, as discussed in Jones (1983), in decomposing the unexplained part of the differential in the outcome variable into a portion due to group membership,

i.e., different intercepts, and a portion due to different coefficients. "The problem is that the relative size of the terms (but not their joint contribution) depends on the location of the zero point of each independent variable of the model" (Jones & Kelley, 1984, p. 334).

Oaxaca & Ransom (1999) showed that the issue raised above goes beyond the identification of the intercept component. Researchers are often interested in the separate contributions of each of the regressors to the endowments and coefficients effect, which is known in the literature as the detailed decomposition. For instance, researchers could be interested in decomposing the differential in wage rates across genders associated with different schooling levels (differing endowments of schooling between male and female workers) and how the schooling levels are remunerated between these groups (i.e., the difference in schooling coefficients between male and female wage regressions). This is appealing as it makes possible identifying the ultimate sources of observed wage differences between males and females. Oaxaca & Ransom (1999) showed that this is valid only for the case of the (detailed) endowments effect. For the case of the (detailed) coefficients effect, the contribution of a categorical schooling variable to the overall discrimination component is not invariant to the choice of the left-out reference category. According to them, the issue still remains with certain affine transformations of continuous regressors. In other words, if for instance schooling entered the wage regressions as a continuous variable (e.g., years of education) any transformation of the specification of this variable (e.g., transforming it into age left school) would yield different estimates of the discrimination that is attributable to schooling, without changing the estimated overall level of discrimination. The authors, therefore, concluded that the detailed decomposition of the coefficients effect was misleading (Oaxaca & Ransom, 1999, p. 154).

Yun (2005) proposed a solution for the issue that occurs with the detailed decomposition when the regressors of the wage equations are categorical[44]. The identification problem occurs because the choice of the omitted reference category is arbitrary, i.e., there is no agreement on which category is to be chosen as the reference group. Yun therefore suggested using the average of the coefficients effects with varying reference groups as the contribution of individual variables to the overall coefficients effect. For instance, if education is a categorical regressor of the wage equations and has three categories, the researcher should calculate the contribution of categories 2 and 3 when category 1 is omitted, the contribution of categories 1 and 2 when category 3 is omitted, and the

[44] He argued that the identification problem related to continuous regressors could not be resolved due to the fact that there are infinitely many transformations possible. The only alternative in specifying continuous regressors is to rely on the conventions in the field (Yun, 2005, p. 766).

contribution of categories 1 and 3 when category 2 is omitted. The average of these three contributions is the overall contribution of the schooling variable to the overall coefficients effect. Interpretation of the average coefficient, however, becomes arbitrary.

Given the issues raised above, Jones & Kelley (1984) favoured a threefold decomposition that included (i) the unexplained difference between groups, which puts together the part of the differential in the outcome variable due to group membership and the portion due to differing coefficients; (ii) the portion due to different endowments of productive characteristics and; (iii) the interaction between different coefficients and different endowments. Any attempt to go beyond this threefold decomposition leads to misleading conclusions (Jones & Kelley, 1984, p. 338). This is the decomposition approach followed in this chapter.

4.4 Methods

The first part of this section describes the specification of the cognitive achievement model that will help identify the best predictors of pupil educational outcomes. The second part presents the specification of the Oaxaca-Blinder decomposition model that is appropriate for the research question addressed in this study, i.e., the factors behind the sharp decline in mean reading and numeracy test scores over time.

The ideal cognitive achievement model is one that identifies the factors that in a cumulative process describe educational outcomes and is formally presented as:

$$Pupil\ outcomes_{it} = f(Family\ background_i^{(t)},\ Peer\ influences_i^{(t)},\ School\ inputs_i^{(t)},\ Teacher\ inputs_i^{(t)},\ Puil\ innate\ ability_i) \tag{13}$$

where the subscript “i” refers to a particular pupil, the subscript “t” stands for a particular time and the superscript “(t)” means cumulative to that particular time.

As in most education production function studies, given the cross-sectional nature of the data at hand we are unable to account for the cumulative process that inputs have on educational outcomes, and we are therefore forced to resort to a more simplified version of the model where only contemporaneous school outcomes and educational inputs are considered. It is assumed that current input measures capture the entire history of inputs and therefore all references to time are dropped from the initial model, which becomes:

$$Pupil\ outcomes_i = f(Family\ background_i,\ Peer\ influences_i,\ School\ inputs_i,\ Teacher\ inputs_i,\ Pupil\ innate\ ability_i) \quad (14)$$

The two SACMEQ surveys (SACMEQ II from 2000 and SACMEQ III from 2007) are pooled together and the models are run using the OLS technique. A number of model specifications are run having reading and numeracy pupil scores as two separate dependent variables. The regressors include among other things provincial and location dummies as well as a SACMEQ III dummy.

An issue with the school effectiveness regressions is that pupil innate ability was not controlled for since the data do not include any variable one could use for that purpose. If pupil ability is positively correlated with, say, family background variables, omitting innate ability will bias the estimates of family background variables upwards. It is therefore assumed that innate ability is unrelated to contemporaneous inputs or it will only affect the coefficients for socio-economic status.

Given the criticisms raised about the Oaxaca (1973) and Blinder (1973) models regarding the treatment of the interaction term or the further splitting of the part attributable to discrimination into a component attributable to differing intercepts and another attributable to different returns to productive characteristics, this work employs the decomposition version suggested by Jones & Kelley (1984). The latter is a threefold decomposition of the differences in the outcome variable into a portion due to different endowments of productive characteristics, a portion due to different coefficients, and an interaction term accounting for the simultaneous effects of the endowments of productive characteristics and coefficients effect.

Formally, the specification of the decomposition technique is as follows: the reading and numeracy test scores are the outcome variables Y in the reading and numeracy decompositions, respectively; the set of predictors are variables on family background, and school and teacher inputs; there are two groups of pupils, those from SACMEQ II and those from SACMEQ III. These groups are represented with the subscripts 2000 and 2007, which correspond to the year the respective surveys were carried out. This technique aims to help explain what was behind the difference (D) in mean test scores over time, i.e., how much of the mean outcome difference

$$D = E(Y_{2007}) - E(Y_{2000}), \quad (15)$$

where $E(Y)$ stands for the expected value of the outcome variable, is accounted for by group differences in the set of predictors. The mean scores difference can be expressed as the difference in

the linear prediction at the group-specific means of the education production function's predictors, as follows:

$$D = E(Y_{2007}) - E(Y_{2000}) = E(X_{2007})'\beta_{2007} - E(X_{2000})'\beta_{2000}, \quad (16)$$

where X is a vector containing pupil test scores' predictors and a constant, and β contains the slope parameters and the intercept. Identifying the contribution of each of the productive characteristics to the overall difference in test scores over time simply entails rearranging Equation (16) as follows:

$$\begin{aligned} D = {} & [E(X_{2007}) - E(X_{2000})]'\beta_{2000} + E(X_{2000})'(\beta_{2007} - \beta_{2000}) \\ & + [E(X_{2007}) - E(X_{2000})]'(\beta_{2007} - \beta_{2000}) \end{aligned} \quad (17)$$

The first summand of this three-fold decomposition represents the expected change in SACMEQ II pupils´ mean test scores if they had SACMEQ III pupils' endowments of productive characteristics. In other words, it amounts to the part of the differential in test scores that is attributable to group differences in the predictors, what is usually called the endowments effect. The second summand represents the contribution of differences in the coefficients (including differences in the intercept), i.e., the expected change in SACMEQ II pupils' test scores when applying the SACMEQ III pupils' coefficients to the SACMEQ II pupils (i.e., the coefficients effect). The last component is an interaction term accounting for the simultaneous effects of existing differences in endowments and coefficients between the two groups of pupils.

The decomposition shown in Equation (17) is formulated from the viewpoint of SACMEQ II pupils. That is, the group differences in the predictors are weighted by the coefficients of SACMEQ II pupils to determine the endowments effect. One could have decomposed the score differential from the viewpoint of SACMEQ III pupils instead, as illustrated in Equation (18). The endowments effect would measure the expected change in SACMEQ III pupils' mean test scores if they had SACMEQ II pupils' predictor levels. By the same token, the second component would quantify the change in SACMEQ III pupils' scores if they had SACMEQ II pupils' coefficients. In this study it makes sense to use Equation (17) instead because we would like to know what "went wrong" that made the "current" pupils perform worse on average on the SACMEQ reading and numeracy tests.

$$\begin{aligned} D = {} & [E(X_{2007}) - E(X_{2000})]'\beta_{2007} + E(X_{2007})'(\beta_{2007} - \beta_{2000}) \\ & - [E(X_{2007}) - E(X_{2000})]'(\beta_{2007} - \beta_{2000}) \end{aligned} \quad (18)$$

4.5 Data

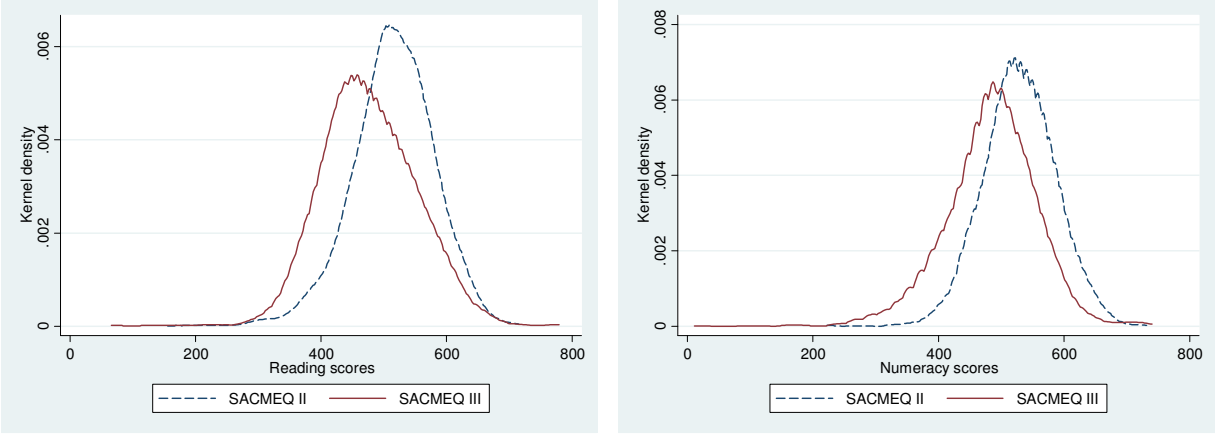
The main source of data this study uses is the Mozambican sample of the SACMEQ archives, the 2000 and 2007 versions, named SACMEQ II and SACMEQ III. Both projects surveyed a stratified random sample of pupils, taking the 11 Mozambican provinces as the strata, and the schools as the primary sample units. In the year 2000 3,177 grade 6 pupils across 176 schools nationwide were surveyed, while the corresponding figures for SACMEQ III were, respectively, 3,360 and 183.

In this work the two surveys were pooled and initially there were 6,537 observations. But missing data relevant to the study reduced the sample size. The problem of missing data at the pupil level was simple to deal with. For instance, 21 pupils did not report the type of floor of the houses they were living in. Another 949 and 1,211 pupils did not report the education level of their mother and father. To avoid dropping these observations and lose important sources of variation, the mode of the primary sample unit, i.e., the mode of the school was imputed, for each variable. With respect to non-response at the school level, however, there was no other alternative but to drop the missing observations from the analysis. For instance, 223 pupils (159 from SACMEQ II and 64 from SACMEQ III) that were in schools whose school head did not report information on participation in management training were dropped from the analysis. Another 172 observations corresponding to the number of pupils in classrooms whose teachers did not report their gender were dropped from SACMEQ II. An alternative in the latter case was to include the missing observations as an additional category of the teacher gender variables when running the regressions. This option did not significantly change the results while making interpretability of the teacher gender variables difficult. Therefore this study kept to the initial approach of dropping the observations. Lastly, 58 pupils (41 from SACMEQ II and 17 from SACMEQ III) did not write the numeracy test, and were therefore also excluded from the analysis. As a result roughly 6% of the observations were dropped from the analysis, leaving us with a sample of 6,142 pupils for the reading regressions and 6,089 for the numeracy regressions.

As previously indicated, mean reading and numeracy test scores fell sharply across the two surveys. Figure 4.2 on page 134, for instance, shows that the SACMEQ III distributions of pupil reading and numeracy scores are to the left of the SACMEQ II distributions, indicating two different data generating processes across the two surveys. In the Appendix, Table A 7 contains weighted summary statistics of the variables included in the cognitive achievement regression models. The table reports their means and standard errors across the two surveys. To select the covariates for

inclusion in the model this study looked at the school level, classroom level, pupil level as well as pupil family background variables that theory suggests should have an effect on pupil educational outcomes.

Figure 4.2 – Overall fall in test scores across surveys



Pupil SES is a variable measuring the socio-economic status of pupils´ household. The derivation of this variable required the use of a technique named Multiple Correspondence Analysis (MCA). The method uses the correlations of household ownership of a variety of assets to compute a latent variable that proxies for household wealth. This technique was the one applied in Chapter 2 to compute the asset index for welfare analysis. In this chapter, the variables included were the same across the two surveys: ownership of a variety of assets such as car, motorcycle, refrigerator, among other things, mother's and father's education, and housing quality (type of roof, floor, walls, and access to water and electricity). Assets such as mobile phone, video cassette recorder, cassette player and the like were excluded from this index, since while access to those assets generally increases over time, they are a poor indication of change in wealth. In other words, due to technological advances the proportion of poor households owning such items is likely to increase while the proportion of rich households possessing them is likely to decline. But if rich households get rid of, say, a video cassette recorder (VCR) because they acquired, say, a DVD player, including VCR in the wealth index would wrongly suggest that richer households became poorer and poor households became richer.

After creating the variable it was normalised so as to have a mean equal to zero and standard deviation equal to one. As the table shows, mean values of pupils socio-economic status seem to have dropped. But this decline across surveys is not statistically significant. This result immediately

affects the initial hypothesis raised in this chapter, i.e., that the average deterioration in education quality over time resulted from the increase in the proportion of pupils from low SES. It is possible that the result we got was influenced by the strong growth rates that characterised the Mozambican economy during the last decade. It also may be that there was measurement error in capturing the assets people buy.

Thus, the wealth index might not be a good indicator of the evolution of pupil socio-economic status over time, even though this study attempted to control for that by excluding certain assets when creating the wealth index. This suspicion is based on the behaviour of other variables in the data. These variables seem to suggest an increase over time in the proportion of grade 6 pupils from lower socio-economic backgrounds. For example, the proportion of pupils with a mother without formal education increased significantly over time. In SACMEQ II 12% of pupils had a mother without formal schooling. By 2007 the figure had risen to 17% of pupils. In terms of actual figures[45] this corresponds to an increase from 14,813 pupils in SACMEQ II to 54,480 pupils in SACMEQ III[46]. There was also a 10 percentage point increase in the number of pupils living in households that used candles as source of lighting, which in actual terms corresponds to an increase from 18,304 pupils in 2000 to 78,170 in 2007. This is indication of an increase over time in the proportion of pupils from low socio-economic status. This work nevertheless uses the wealth index since at the cross-sectional level it does distinguish wealthier pupils from their less advantaged counterparts.

The average pupil was aged fourteen and a half years in SACMEQ II, but only fourteen in SACMEQ III. The proportion of girls in Mozambican schools increased significantly over time, from 41% to 46%, an indication of a trend to gender equalisation in access to education. A relatively high proportion of pupils reported getting help with school work at home. But there was not a significant change over time, which could be associated with the increase in the proportion of parents without schooling. Uneducated parents would face more difficulties helping their children with school work.

[45] The figures are calculated using the raising factor provided in each data set. Unlike the normal sample weight included in the surveys, this special weight provides a mechanism for calculating population totals for our variables of interest.

[46] The proportions are from different totals. 17% of SACMEQ III pupils is roughly four times 12% of SACMEQ II pupils on account of the sharp increase in enrolments over time.

Until 2004 teachers had the discretion to deny promotion to those pupils who failed to reach the minimum learning requirements for the grade. In 2004, however, the government introduced a new education policy regarding grade repetition. Since then the primary level of education is split into three learning cycles within which teachers are not allowed to fail pupils, irrespective of pupil learning performance. Technically, after completing the academic year in grades one, three, four, and six pupils are automatically promoted to the subsequent grade, whether or not they meet the minimum requirements for passing the grade. They can only be retained in grades two, five and seven if the learning requirements are not met. As Table A 7 illustrates, the proportion of pupils that never repeated a grade rose dramatically over time from 22% to 40%. This is in all likelihood associated with the mentioned policy of automatic grade promotion, and might have facilitated the promotion of a great number of lower ability pupils.

About half of pupils reported that they spoke the language of instruction outside school most of the time. The figure remained stable across the two surveys. Most pupils reported having the evening meal regularly, and this has improved significantly over time, which can be associated with the strong growth of the Mozambican economy over the last decade. As indicated by the online database of the Mozambican National Statistics Institute (INE), between 2000 and 2007 the economy grew in real terms by an average of 8% per year. Nevertheless, the proportion of pupils without any books at home increased significantly from 20% to 31%, another indication of a fall in the average socio-economic status of the pool of pupils attending grade 6.

Pupils' reports on their own absenteeism suggest it was quite low in SACMEQ II, averaging about 3 days a year. In SACMEQ III it declined further to 1 day a year. In SACMEQ II only 6% of pupils were in schools whose school head had a degree and had received school management training. The figure did not change significantly over time.

The percentage of pupils from urban schools seems to have declined, but the fall was statistically insignificant. The proportion of pupils coming from the various provinces was also fairly stable over time, as was the proportion of pupils in government (public) schools. In both surveys roughly 98% of pupils were in public schools. There were advances with respect to gender equalisation at the pupil level, but not at the teacher level. Roughly only one third of pupils had female teachers both in SACMEQ II and in SACMEQ III.

An immediate consequence of fast increases in enrolments is that the deployment of teachers does not grow at the same pace as the number of pupils. Class sizes, therefore, tend to get bigger. For the case of Mozambique, which already had large classes in SACMEQ II (53 pupils per class on average), in the latest survey things got worse, with the average pupil studying in a class of 59 pupils. In a situation where roughly only half of pupils have their own reading and maths textbooks, the limited access to teachers is very alarming, as it is likely to translate into poor educational outcomes.

Nevertheless, some teacher practices thought to have a positive impact on pupils' academic achievement improved. For instance, the number of hours teachers spent per week preparing lessons and marking increased significantly, from about 7 hours a week to 10 hours a week. Further, teachers are normally required to write school reports for each pupil at the end of each term. In SACMEQ II roughly 40% of pupils were in classes whose teacher's reports included comments. The figure increased to 60% in SACMEQ III.

4.6 Reporting on the Education Production Functions

This section reports the results of the multivariate regressions, i.e., the effect of each explanatory variable on test scores taking into account other determining factors of scores. The section starts by discussing the reading regressions and then goes on to discuss the numeracy regressions.

Below on page 140 Table 4.3 shows four different specifications of reading education production functions. The first specification includes family background and pupil level variables only. Pupil socio-economic status is positively associated with reading scores. Controlling for other factors that have a bearing on reading test scores, on average, pupils from more affluent backgrounds performed significantly better than their less advantaged counterparts. An increase by one standard deviation in pupil SES is associated with an increase in reading scores by 18 points or by 0.24 of a standard deviation[47]. The coefficient on age is negative and statistically significant, suggesting that overage pupils scored less than younger pupils. Girls scored less than boys by about 13 score points.

The model also shows that parents have a role in improving the education performance of their offspring. Pupils that got help with their school work at home performed better than those who reported otherwise. Extra tuition, however, has a negative relationship with reading scores, i.e.,

[47] In the pooled data the reading scores have mean 495 and standard deviation of 73. The coefficient on Pupil SES (17.63) divided by 73.46 (the standard deviation) is equal to 0.24 standard deviations.

pupils taking extra tuition reading classes scored less on average than those not taking these extra lessons. This suggests the presence of endogeneity in the sense that pupils took extra classes because they were performing badly. In other words, the extra tuition classes could have been a remedial measure for already weaker students.

Pupils that repeated grades twice scored better on average than pupils who never repeated a grade. Pupils with greater absenteeism levels also scored better than more regular pupils. These two results are unexpected and very likely caused by misspecification of the model. In the preferred model specification (which is the fourth one), these variables have the expected results. Grade repetition and pupil absenteeism become negatively associated with scores.

In Mozambique Portuguese is the language of instruction. In the pooled data set 7% of pupils indicated that they never spoke Portuguese outside school. The regression results suggest that speaking the language of instruction also outside school environment produced better results than not speaking it at all. For instance, pupils who reported speaking Portuguese outside school most of the time scored 40 points more than those who never spoke it outside school. In fact, even speaking it sometimes has an effect of similar magnitude on pupils' reading scores.

The regression results also show that pupils' nutrition has a bearing on their academic performance. Pupils that had supper every day performed 19 score points higher than those that had the evening meal less often. Having a considerable number of books at home is beneficial to pupil literacy since they can practice their reading and writing skills. It is not surprising then that pupils reporting having books at home, in particular those with more than 10 books at home, performed better on the reading test than those not having books at home.

The second specification extended the model to include school level variables. The signs, magnitude and statistical significance of the family background and pupil level variables remained roughly the same, except that grade repetition changed signs and became associated with lower reading scores. School SES, which measures the average pupil socio-economic status in the school, was found to be positively associated with reading scores, suggesting that pupils in more affluent schools performed better than pupils in less affluent schools. But for this particular specification school socio-economic status was statistically insignificant.

The coefficient on the urban dummy is small and statistically insignificant, indicating that on average pupil performance on the reading test was similar across locations. The provincial dummies suggest that controlling for pupil and school factors determining pupil academic achievement, Niassa alongside Cabo Delgado and Tete were, on average, the worst performing provinces and Maputo City the best performing one. School type as well as education and training of school head did not enter the regression significantly.

The third specification added teacher or classroom level variables. Pupils with female reading teachers scored less than pupils with male teachers. Class size is negatively related to scores and is statistically significant, indicating that after controlling for other determining factors of pupil reading scores, pupils studying in bigger classes performed worse on average than pupils studying in smaller ones. Nevertheless, the magnitude of the effect is rather small, suggesting small pupil performance gains for the costly reduction of class sizes. A reduction in class sizes by one pupil increases reading scores by 0.63 points only. Pupils that had their own textbooks or even pupils that shared them with other pupils performed better than those that did not have any kind of textbook access. There is, therefore, a role for policy on providing textbooks, in particular for those pupils from low socio-economic backgrounds. These pupils are more likely to have limited access to textbooks.

The last specification included the variables described in the previous model specifications plus added a SACMEQ dummy variable and its interaction with pupil socio-economic status. Because most variables were statistically significant and presented the expected signs this is the preferred model specification. Apart from that this specification has the best model fit ($R^2 = 0.30$). In contrast with previous model specifications, pupils in schools whose school head had a university degree and received management training performed better (by 13 points) than pupils in schools without such school heads. Pupil absenteeism now has the expected negative sign. Each day of absenteeism is associated with a decline by 0.85 points in pupil scores. All dummies on grade repetition are negative and statistically significant, which suggests that on average and controlling for other factors, repeaters performed worse than non-repeaters. The number of hours the reading teacher spent preparing lessons and marking entered the regression as a concave function but was not statistically significant. Also, teacher practices of including comments in pupils' reports, which could be proxying for monitoring of pupil performance or for whether pupils and their parents were provided with teacher feedback, was not significant, despite the positive sign.

The SACMEQ III dummy enters the regression with a negative and statistically significant coefficient, suggesting that controlling for the (included) factors thought to determine pupil scores, pupils performed worse in SACMEQ III than in SACMEQ II by an average of 39 score points. The coefficient on the interaction of the SACMEQ III dummy with pupil SES is positive and statistically significant. It therefore indicates that while pupil performance was weaker in SACMEQ III relative to SACMEQ II, the average drop in reading scores was larger for low SES pupils than for those from more affluent backgrounds.

Figure 4.3 on page 142, though not controlling for other determining factors of pupil achievement, reiterates this finding. On its left panel it shows the distribution of quintile 1 pupils' scores across surveys, and on the right panel the distribution of quintile 5 pupils' scores also across surveys. The distribution of scores amongst the poorest group of pupils in SACMEQ III is clearly to the left of the SACMEQ II distribution. For the richest group of pupils, however, the two SACMEQ distributions almost overlap, therefore indicating that the average drop in pupil performance was quite small.

Table 4.3 – Reading education production functions

	(1) *Reading Scores*	(2) *Reading Scores*	(3) *Reading Scores*	(4) *Reading Scores*
Pupil SES	17.63***	7.96***	7.50***	2.42
Age	-2.93***	-1.77***	-2.27***	-2.89***
Pupil sex (ref: Boy)				
Girl	-12.56***	-13.96***	-13.07***	-11.68***
Pupil gets help with school work at home	7.20**	5.78*	4.58	6.03**
Pupil takes extra tuition	-9.17***	-5.81*	-6.63**	-15.07***
Repeated grades (ref: Never)				
Once	2.38	-1.61	-2.43	-5.87**
Twice	8.28***	-0.87	-2.11	-7.68***
3+	1.36	-9.02**	-11.69***	-15.07***
Pupil speaks Portuguese at home (ref: Never)				
Sometimes	40.17***	31.96***	30.96***	29.09***
Most of the time	39.69***	34.19***	33.00***	30.38***
Pupil has supper every day	19.11***	19.07***	19.43***	23.74***
Number of books at home (ref: No book)				
1 -10	6.79*	7.60**	7.54**	3.71
11+	15.31***	13.28***	11.45***	6.83*
Number of days absent from school	1.80***	1.59***	0.89**	-0.85**
School head has degree and received management training		9.67	10.45	13.49*
School SES		4.73	10.77***	9.47***
School type (ref: Government)				
Private		-3.02	-4.99	0.97

Continues next page

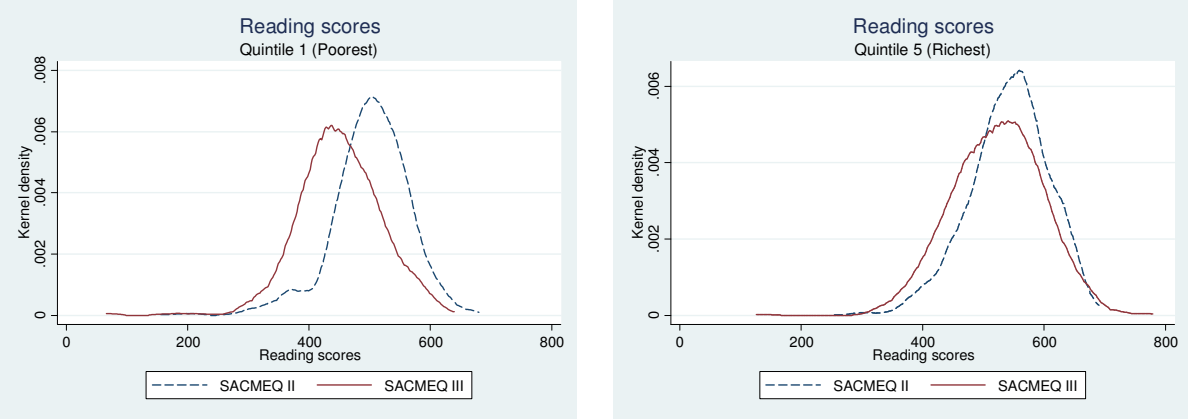
Continued from previous page

School location (ref: Rural)				
Urban		5.28	7.61	1.74
Province (ref: Niassa)				
Cabo Delgado		12.04	10.52	14.49
Nampula		47.65***	41.39***	41.17***
Zambezia		43.89***	37.69***	39.40***
Tete		12.62	8.49	6.83
Manica		30.26***	18.42**	22.30**
Sofala		21.96**	19.13**	19.73**
Inhambane		51.06***	42.60***	41.08***
Gaza		43.45***	31.25***	33.53***
Maputo Province		53.39***	40.25***	41.91***
Maputo City		75.27***	63.28***	62.86***
Teacher sex (ref: Male)				
Female			-11.71***	-6.04*
Class size			-0.63***	-0.33**
Reading textbooks (ref: No textbooks)				
Only teacher has			5.59	-0.39
Share with 2+			28.51***	23.92***
Share with 1			36.00***	28.01***
Use by myself			34.61***	29.31***
Number of hours of teacher preparation per week			-0.85	0.89
Squared number of hours of teacher preparation per week			0.02	-0.03
Pupil school report includes comments			-4.25	1.82
Survey (ref: SACMEQ II)				
SACMEQ III				-39.31***
SACMEQ III x Pupil SES				8.44***
Constant	469.75***	420.18***	453.03***	467.98***
Model Statistics				
Observations	6,142	6,142	6,142	6,142
F-Statistic	28.20	19.71	23.40	29.15
Prob > F	0.00	0.00	0.00	0.00
R^2	0.14	0.22	0.25	0.30

Note: dependent variable is reading test scores; model estimation accounts for complex survey design; * p<0.1, **p<0.05, ***p<0.01

The following paragraphs discuss the results of the numeracy education production functions. Like was done in the previous discussion, we ran four different model specifications. The first specification included family background and pupil level variables only. The second one added school level variables, and teacher or classroom level variables were added in the third specification. The last specification added the SACMEQ III dummy as well as its interaction with pupil socio-economic status.

Figure 4.3 – Distributions of reading scores across surveys for the poorest and richest quintiles



The output of the numeracy regressions is presented in Table 4.4 on page 143. With a few exceptions the findings are similar to what was found with the case of the reading regressions, and therefore this discussion does not go into too much detail. The first exception is that getting help with homework is still positively associated with pupil scores, but its effect is statistically insignificant. The second exception concerns the coefficient on extra tuition, which is now positive and statistically significant in the first three model specifications, but insignificant in the fourth one. This is an indication that extra classes in maths improved pupils' numeracy scores, but after all potential variables are included in the regression, nothing can be said about the effect of this variable on numeracy scores. Another exception is related to teacher gender. Unlike pupil reading scores, pupil numeracy scores did not differ significantly by teacher gender. Lastly, in Model 4, which is the preferred specification, the number of hours of teacher preparation per week and its square enter the regression concavely but now they are statistically significant. This suggests that the more hours teachers put in preparing lessons and marking, the better pupils perform on the numeracy test, but the increase in scores fades out the longer the time teachers spend.

As was the case with the reading regressions, in the fourth model specification the SACMEQ III dummy and its interaction with pupil SES suggest that in SACMEQ III pupils performed worse than was the case in SACMEQ II. Again, the fall in numeracy scores was larger for low SES pupils than for their more affluent counterparts. It is important to note, however, that in both regressions the magnitude of the coefficient on the interaction term (SACMEQ III x Pupil SES) is rather small. This suggests that the large influx of low SES pupils played only a marginal role in explaining the fall in mean scores over time.

Table 4.4 – Numeracy education production functions

	(1) *Numeracy Scores*	(2) *Numeracy Scores*	(3) *Numeracy Scores*	(4) *Numeracy Scores*
Pupil SES	8.70***	1.67	1.24	-1.30
Age	-1.34**	-0.14	-0.64	-1.33**
Pupil sex (ref: Boy)				
Girl	-17.60***	-19.25***	-18.38***	-16.55***
Pupil gets help with school work at home	3.24	1.88	1.04	3.54
Pupil takes extra tuition	12.35***	13.43***	11.88***	-1.53
Repeated grades (ref: Never)				
Once	-0.26	-3.50	-3.67*	-8.25***
Twice	5.68**	-1.45	-2.18	-9.62***
3+	2.96	-4.30	-5.38	-10.04***
Pupil speaks Portuguese at home (ref: Never)				
Sometimes	29.23***	22.45***	22.01***	20.19***
Most of the time	27.85***	24.02***	23.01***	20.79***
Pupil has supper every day	12.82***	11.80***	12.41***	17.54***
Number of books at home (ref: No book)				
1 -10	9.70***	9.63***	10.19***	5.60**
11+	12.97***	11.03***	9.91***	4.32
Number of days absent from school	1.66***	1.81***	1.13**	-0.88**
School head has degree and received management training		3.14	4.03	8.30*
School SES		2.87	6.11	6.14*
School type (ref: Government)				
Private		-8.47	-11.07	-3.22
School location (ref: Rural)				
Urban		2.14	4.58	-2.33
Province (ref: Niassa)				
Cabo Delgado		11.60**	11.96**	13.06**
Nampula		39.24***	34.27***	33.62***
Zambezia		32.26***	26.93***	30.26***
Tete		19.25***	14.82**	13.62**
Manica		39.16***	27.50***	30.88***
Sofala		25.75***	21.02***	20.26***
Inhambane		60.53***	50.83***	46.90***
Gaza		51.51***	40.76***	38.57***
Maputo Province		53.00***	43.25***	42.09***
Maputo City		63.10***	54.78***	50.42***
Teacher sex (ref: Male)				
Female			1.59	5.22
Class size			-0.79***	-0.44***
Maths textbooks (ref: No textbooks)				
Only teacher has			-3.59	-3.02
Share with 2+			20.35***	18.51***
Share with 1			21.02***	17.20***
Use by myself			22.39***	17.90***
Squared number of hours of teacher preparation per week			0.00	-0.04***
Pupil school report includes comments			-5.47	0.74

Continues next page

Continued from previous page

Survey (ref: SACMEQ II)				
SACMEQ III				-46.51***
SACMEQ III x Pupil SES				4.85**
Constant	475.08***	427.59***	471.40***	490.54***
Model Statistics				
Observations	6,089	6,089	6,089	6,089
F-Statistic	14.80	16.91	15.44	26.36
Prob > F	0.00	0.00	0.00	0.00
R^2	0.08	0.14	0.17	0.23

Note: dependent variable is numeracy test scores; model estimation accounts for complex survey design; * p<0.1, **p<0.05, ***p<0.01

4.7 Reporting on the Oaxaca-Blinder Decompositions

The purpose of doing the Oaxaca-Blinder decompositions in this study was to investigate what changes may have occurred over time that would help explain the reasons behind the dramatic fall in the SACMEQ average reading and numeracy test scores in Mozambique between the years 2000 and 2007. Table 4.5 on page 146 shows the decomposition of reading and numeracy test scores over time into three parts, as discussed in Section 4.3: a part explained by differences in endowments of the productive characteristics (the endowments effect); a part reflecting differences in the returns to the productive characteristics (the coefficients effect); and the joint impact of the differential in characteristics and differential in coefficients (the interaction effect). As stated previously, compared to their SACMEQ II counterparts the Mozambican pupils performed worse in SACMEQ III both on the reading test and on the numeracy test. The average fall in scores was 38.5 in reading and 45.5 in numeracy, corresponding to 0.59 and 0.79 of Mozambican standard deviations[48].

In both decompositions the portion $[E(X_{2007}) - E(X_{2000})]'\beta_{2000}$ which stands for the explained component (i.e., the endowments effect) is negative, indicating that on average SACMEQ II pupils had better endowments of productive characteristics than SACMEQ III pupils. If for instance, only one productive characteristic was included in the decomposition and it was, say, pupil nutrition (found to be positively related to pupil scores), the endowments effect would always be negative if the proportion of pupils that had supper regularly was higher in SACMEQ II than in SACMEQ III. If instead the included productive characteristic was class size (found to be negatively related with pupil scores), the endowments effect would also always be negative if SACMEQ II pupils were in

[48] In SACMEQ II the standard deviation of the mean test scores in Mozambique was 65.14 for the reading test and 57.30 for the numeracy test.

smaller classes relative to SACMEQ III pupils. Thus, a negative endowments effect would always indicate superior endowments of productive characteristics for SACMEQ II pupils.

Looking only at their superior endowments of productive characteristics it makes sense that SACMEQ II pupils have scored better on average than SACMEQ III pupils. In other words, the average deterioration in endowments of productive characteristics over time contributed to the score gap existing between SACMEQ II and SACMEQ III pupils. The contribution to the fall in scores over time, however, was small in magnitude and statistically insignificant both in reading (-5% of the differential in scores) and in numeracy (-1% of the differential in scores), suggesting that there is no evidence to state that the average fall in scores over time had much to do with changes in endowments of productive characteristics[49] (see Table 4.5 on page 146).

SACMEQ III pupils had on average smaller returns to their productive characteristics than their SACMEQ II counterparts both in reading and numeracy. This is indicated by a negative and statistically significant coefficients effect in both decompositions (see Table 4.5). It is simple to conclude from its formula, which is given by $E(X_{2000})'(\beta_{2007} - \beta_{2000})$, that the coefficients effect will always be negative whenever the coefficients of the regressions involving SACMEQ II pupils are greater than the coefficients of SACMEQ III's regressions. So, respectively 104% and 100% of the gap in reading and numeracy scores over time is explained by the coefficients effect. The interaction term in the reading and numeracy decompositions suggests that -9% and -3% of the score differential respectively across surveys is explained jointly by the change in pupils' endowments of productive characteristics as well as by the change in the returns to those productive characteristics. This effect, however, is statistically insignificant, which indicates that there is not enough evidence to attribute part of the fall in scores to the interaction effect.

Thus, for both reading and numeracy, the bulk of the overall fall in scores over time resulted from changes in the coefficients. This suggests that, in terms of learning outcomes, for the included set of pupil productive characteristics schools became worse in converting those characteristics into cognitive achievement. Thus there is no evidence that it was the deterioration in the socio-economic background of the pool of pupils that caused the fall in mean scores over time, as the endowments

[49] Note that because the endowments effect is negative the reported sum of the proportions attributable to the coefficients effect and interaction effect goes beyond 100%, but the sum of the proportions of the three effects should add up to 100%.

effect had no significant role. The evidence points instead towards a deterioration in school efficiency between SACMEQ II and SACMEQ III.

Table 4.5 – Decomposition of reading and numeracy scores over time

	Reading scores		Numeracy scores	
SACMEQ III (2007)	476.9***		484.1***	
SACMEQ II (2000)	515.4***		529.5***	
Difference	-38.5***		-45.5***	
	Scores	**Proportion**	**Scores**	**Proportion**
Endowments	-1.8	0.05	-0.5	0.01
Coefficients	-40.3***	1.04	-46.2***	1.00
Interaction	3.6	-0.09	1.2	-0.03
Endowments				
Family background	0.9		1.3**	
Pupil level	5.0***		2.7*	
School level	-4.4		-3.4	
Classroom level	-3.4		-1.1	
Coefficients				
Family background	1.6		-4.8	
Pupil level	33.7**		45.9**	
School level	-39.0***		-22.5**	
Classroom level	34.9**		13.6	
Constant	-71.6**		-78.3***	
Interaction				
Family background	-1.2		-1.7*	
Pupil level	1.7		0.1	
School level	-0.6		1.7	
Classroom level	3.7		1.1	
Observations	6,142		6,089	

Note: decompositions account for complex survey design; * p<0.1, **p<0.05, ***p<0.01

Table 4.5 also shows within each of the three components of the decomposition which group of variables contributed the most for the score gaps. With respect to the endowments effect, predictors falling under the pupil level category contributed had a positive effect on tests scores over time, both in reading (5.0 score points) and in numeracy (2.7 score points). In other words, between SACMEQ II and SACMEQ III pupil level productive characteristics improved. For instance, pupils became younger, grade repetition became less frequent, and school absenteeism fell. This means that considering only pupil level variables, scores should have increased between surveys by 5.0 score points in reading and by 2.7 score points in numeracy.

The detailed decompositions showing the contribution of each variable to the reading and numeracy score gaps over time are presented respectively in Table A 8 and Table A 9 in the Appendix. The specific pupil level productive characteristic that improved the most and therefore contributed the most against the average fall in reading scores over time was extra tuition. Extra tuition was found to be negatively related to reading scores and the proportion of pupils taking reading extra tuition

declined between surveys from 0.28 to 0.11. The endowments effect associated with this variable is 3.2 (see Table A 8). For the case of the numeracy test, age is the variable that contributed the most against the average fall in scores. The average pupil became younger between surveys. The average age declined from 14.51 years to 13.95 years. The endowments effect associated with this variable is 1.72 score points (see Table A 9).

Despite the slight improvements in pupil level productive characteristics between SACMEQ II and SACMEQ III, school level productive characteristics became worse. For instance (ignoring statistical significance), school socio-economic status worsened and the proportion of grade 6 pupils coming from Maputo City (the Capital) also declined. This therefore contributed to a change in scores by -4.4 score points in reading and by -3.4 score points in numeracy, thus offsetting the advantage provided by pupil level variables. In the end, and as is shown in Table 4.5, after controlling for the variables that potentially had a bearing on pupil academic achievement, the endowments effect was negative for both reading and numeracy, but small in magnitude and statistically insignificant.

As mentioned earlier, the detailed decomposition of the coefficients effects is criticised in the literature. Despite those criticisms the results of these decompositions are also discussed here. The coefficients effect, i.e., the effect of differing returns to productive characteristics between SACMEQ II and SACMEQ III pupils, indicates that school level variables contributed the most to the score differential in both reading and numeracy (see Table 4.5 on page 146). This means that the coefficients of school level variables in SACMEQ III are smaller than SACMEQ II coefficients. For both the reading and numeracy tests it was the returns to school level variables jointly with the intercept that contributed the most to the negative coefficients effect. In the reading decomposition they contributed respectively -39.0 and -71.6 score points, respectively, but were offset by a substantial advantage to SACMEQ III pupils due to the coefficients on the remainder of the variables, especially the classroom level variables (34.9 score points), therefore reducing the coefficients effect to -40.3 score points. In numeracy things were similar. The coefficient on pupil level (45.9) and classroom level (13.6) variables had a positive effect on scores. Family background variables (-4.8), school level variables (-22.5) and the intercept (-78.3) contributed to a fall in scores over time. It is simple to conclude that, in both cases, without the contribution of the intercept to the fall in scores over time, the coefficients effect would in fact have been positive. This means that what caused the large negative coefficients effects was not the fall in coefficients of the productive

characteristics between SACMEQ II and SACMEQ III, but instead the large fall in the intercept between the two surveys.

The negative, large, and statistically significant intercept effects can be interpreted following the literature, as the advantage of SACMEQ II pupils over SACMEQ III pupils on aspects unrelated either to differentials in endowments of productive characteristics or to different returns to those productive characteristics. For instance, dropout rates were relatively high for the earliest cohort (SACMEQ II pupils). According to UNESCO's Global Monitoring Report (2011, p. 316) the dropout rate for all grades in Mozambique was 72% in 1999, but only 56% in 2008. Assuming that these figures accurately represent dropout rates of grade 6 pupils in 2000 and 2007, and that the pupils dropped out before the reading and numeracy tests were conducted, if dropouts would have scored below average on the SACMEQ tests, then the significantly greater proportion of early school leavers in SACMEQ II might have upwardly biased the SACMEQ II average results.

Another plausible explanation for the results found here could be that standards have declined between the two surveys. The rapid expansion of the school system between the years 2000 and 2007, added to the fact that since 2004 teachers and schools institutionally have less discretion to fail underperforming pupils, might have resulted in the promotion of a great proportion of low ability pupils. The greater proportion of weak pupils in SACMEQ III compared to SACMEQ II might have pulled down the average reading and numeracy test scores over time. Unfortunately, the data used in this study do not allow testing this hypothesis, as the data do not include measures of innate ability.

It was attempted to include in the regressions and decompositions variables related to teacher quality and classroom dynamics such as teacher experience and education, homework frequency and homework correction frequency, but the results were statistically insignificant and in some cases counter-intuitive. With the pressure on the education system teacher quality and classroom dynamics very likely deteriorated over time. Though statistical inferences cannot be made on this, practices such as some pupils having access to exam papers before actually writing them are becoming more frequent in Mozambican schools. This might indicate that the education system became so large that its administration worsened.

As discussed above, however, partitioning the coefficients effect[50] into a portion attributable to differing returns to productive characteristics between groups and another portion associated with different intercepts (or group membership per se) is considered inappropriate by some since the size of each portion depends on the choice of the omitted base category of the regressions' categorical predictors. To check the impact of choosing alternative base categories for the categorical predictors this work employed Yun's (2005) solution, which is to use the average of the coefficients effects with varying reference groups as the contribution of individual variables to the overall coefficients effect. This solution, however, did not improve the results. It only made the intercept effects change from -71.6 to -115.3 score points in reading and from -78.3 to -92.6 in numeracy[51]. This solution, however, leaves the estimation and decomposition without a simple meaningful interpretation.

Thus, if the detailed decomposition of the coefficients effect is valid, then neither changes in endowments of productive characteristics nor changes in their returns explain the average fall in pupils' scores over time. Another factor unrelated to pupils' characteristics, possibly the relatively higher dropout rates in SACMEQ II or the massive promotion of low ability pupils through the school system, could explain the fall in test scores over time. If, however, one is sceptical about the detailed decomposition of the coefficients effect or to the solution proposed by Yun (2005) and therefore looks only at the (broad) decomposition of the score gap over time into three portions, one should conclude that for both reading and numeracy the bulk of the overall difference in scores over time resulted from changes in the returns to productive characteristics between SACMEQ II and SACMEQ III pupils. In other words, there was a deterioration in the efficiency of the school system over time.

4.8 Conclusion

As noted in the introduction, in the last two decades the expansion of the school network to previously underserved regions of Mozambique resulted in large increases in pupil enrolments, in particular at the primary level of education. For instance, the SACMEQ data indicate that between 2000 and 2007, the number of grade 6 pupils in previously underserved provinces such as Niassa

[50] As discussed before, some authors argue against the detailed decomposition of the coefficients effect. For instance, Jann (2008) argues that "the detailed decomposition results for the unexplained part only have meaningful interpretation for variables for which scale shifts are not allowed, that is, for variables that have a natural zero point" (Jann, 2008, p. 9).

[51] Because of interpretability issues concerning Yun's (2005) solution the results are not reported here, but they are available from the author upon request.

and Zambezia increased by 355% and 238% respectively, an indication of great success of the Mozambican government in expanding the coverage of the education system across the country.

Despite the successes in expanding the coverage of the school network over time and the resulting improvement in access to education, there are indications that the average quality of education has deteriorated. SACMEQ tested two cohorts of grade 6 pupils, one in 2000 and another in 2007. The Mozambican pupils performed above the SACMEQ average in 2000 (SACMEQ II) both on the reading test and on the numeracy test. In SACMEQ III, however, their mean reading and numeracy scores fell dramatically, by 38.5 and 45.5 score points, corresponding to 0.59 and 0.79 Mozambican standard deviations, respectively.

The objectives of this chapter were twofold. First, it aimed to isolate the school, teacher and pupil characteristics that consistently predict educational outcomes in Mozambique. For that this work employed the education production function approach that looks at the school as a production facility responsible for generating educational outputs. The second objective was to identify the factors behind the sharp decline in pupil mean reading and numeracy test scores over time. Initially it was hypothesised that it was the massive influx of pupils from low socio-economic backgrounds that pulled down the mean tests scores over time. To corroborate this hypothesis this work made use of the Oaxaca-Blinder decomposition technique often employed in the literature to study labour market outcomes by decomposing earnings across groups into a component explained by productive differences (i.e., endowments effect) and an unexplained component (i.e., coefficients effect) normally attributed to discrimination. This work adapted that technique and decomposed the average fall in reading and numeracy scores, separately, over time into a portion attributable to different endowments of productive characteristics, a part associated with different returns to those characteristics and a component due to the joint effect of the first two.

Employing the cognitive achievement regression models this study found that most of the variables suggested by the literature on school effectiveness were indeed significantly related to educational outputs (reading and numeracy test scores) in Mozambique. At the level of the pupil, socio-economic status, younger pupils compared to older ones, boys compared to girls, the number of books at home, parents help with school work, pupil nutrition, minimal absenteeism and grade repetition levels, and speaking the language of instruction also outside school versus never speaking it outside school are the variables found to be positively associated with reading and numeracy test

scores. At the school level it was found that more educated school heads who received specialised training in school management were associated with better pupil performance in their respective schools. This work also found that average school socio-economic status, a measure of school wealth or of peer effects, was positively related with tests scores. At the classroom level access to textbooks was found to be positively associated with pupil test scores. The number of hours teachers spent marking and preparing classes per week also had a positive effect on pupils' scores. Class size was found to have a bearing on pupil educational outcomes. Pupils in bigger classes scored less on average. But in both cases, the magnitude of the coefficient on class size was rather small, suggesting that while there are gains from reducing current class sizes these effects may not be large enough to justify policy action given costs. The coefficient on the SACMEQ III dummy, which takes the value of 1 for SACMEQ III and 0 otherwise, was found to be negative, large and statistically significant, both for the reading and numeracy tests. It indicates that controlling for other factors, on average SACMEQ III pupils performed worse than SACMEQ II pupils. The coefficients on the interaction of the SACMEQ dummy with pupil SES were positive and statistically significant, therefore indicating that while pupil performance was weaker in SACMEQ III relative to SACMEQ II, the average drop in reading and numeracy scores over time was larger for low SES pupils than for those from more affluent backgrounds.

The results of the Oaxaca-Blinder decompositions indicate that the endowments of productive characteristics did not change much over time. The endowments effect was 5% for the reading test and 1% for the numeracy test, but it was statistically insignificant in both cases. The returns to productive characteristics (i.e., the coefficients effect), on the other hand, seem to have declined a lot. The data suggest that respectively 104% and 100% of the gap in mean reading and numeracy tests scores over time was attributable to differing returns to productive characteristics between SACMEQ II and SACMEQ III. The detailed decomposition of this component, however, suggests that it was not the returns to productive characteristics per se that fell over time, but instead the effect of group membership per se, i.e., the intercept. This chapter raised the hypothesis of ability bias, i.e., that the sharp fall in pupil dropout rates across surveys gave advantage to SACMEQ II pupils, assuming that early school leavers dropped out before the SACMEQ test took place and would have scored below average had they written the test. This would have biased the SACMEQ II mean scores upwards, an effect possibly large enough to cause test scores to fall over time. Another plausible hypothesis was that the education system grew so large in a short time that

standards declined over time. For instance, between 2000 and 2007 the lenient promotion of lower ability pupils might have increased sufficiently to pull down the average test scores.

The detailed decomposition of the coefficients effect is criticised in the literature in the sense that meaningful interpretation is only possible for variables that have a natural zero point. Most of the predictors of pupil test scores were dummy variables, and therefore there was some arbitrariness about the choice of the omitted base category. One could have chosen alternative base categories and the contribution of each variable (including intercepts) to the total coefficients effect would have changed. Yun's (2005) solution tries to get around this issue but not without creating further arbitrariness. The task therefore rests on the reader to decide which "story" is more plausible: interpreting the coefficients effect as a whole and therefore concluding that because the coefficients effect worsened over time it indicates a worsening of school efficiency, or alternatively, that the fall in test scores between SACMEQ II and SACMEQ III was caused by the fall in the intercept effect, reflecting ability bias or the fall in standards over time.

Either way, the findings from the empirical work presented in this chapter suggest that the issue resides in the education system. The Mozambican authorities seem to be utterly focused on providing full primary education coverage, even if it comes at the expense of quality deterioration. In fact, a few researchers have suggested such approach. Handa & Simler (2005) argued that when access to education is low (as was the case in Mozambique) and cost information is considered, a stronger focus should be placed on education quantity. But if attention is diverted from quality and focused only on quantity, Mozambique runs the risk of having a huge mass of functionally innumerate and illiterate primary school graduates. The funds saved by providing more access and less quality schools will ultimately be in vain, since the skills the funds aimed to impart in pupils will not be assimilated by them, despite having completed seven years of education.

As mentioned in the introductory chapter, enrolment rates at the primary level of education have been increasing fast in the past few years. In particular, net enrolment rates at the lower primary level of education (grades 1 to 5) were at 95% by 2007 (see Figure 1.3 on page 20). This suggests that even if Handa and Simler as well as the Mozambican authorities were right in focusing on education quantity first, the time has arrived to start giving a greater emphasis to education quality. Issues of school efficiency and of lenient grade promotions should be addressed without delay.

Chapter 5
Conclusion

As noted in Chapter 1, this thesis had three general groups of research objectives, which were dealt with sequentially in separate chapters. In other words, the empirical work was largely divided into three chapters. In order to accommodate this structure they were discussed separately in chapters 2, 3 and 4. The next few paragraphs present the thesis objectives and summarise the main findings. Then follow the conclusion with respect to the findings and the theses statements, next a summary of contributions follows and lastly suggestions for further research are presented.

5.1 Objectives and Summary of Findings

Chapter 2 aimed to improve the understanding of well-being in Mozambique. That was justified on the grounds that earlier studies based on money-metric analysis drew counter-intuitive conclusions about the spatial distribution of poverty. This deficiency is detrimental for policy targeting. To attain the mentioned objective this chapter proposed the use of an asset index that would capture another dimension of well-being.

In this chapter, the main findings were the following. Employing the money-metric dimension of well-being and the variable poverty lines defined by DNPO et al. (2004), in the period 2002-2003 about 54% of the Mozambican population was below the poverty line. People living in households whose head had completed at least the primary level of education tended to be less poor than people living in households with less educated heads. Rural regions were poorer than urban regions but the gap across locations was very narrow. The Southern region of Mozambique was a great deal poorer than the Central and the Northern regions, but the ranking was not robust to the choice of the poverty line. On the provincial ranking, Maputo City was found to have greater incidence of poverty than most of the Central and Northern provinces. Only Cabo Delgado in the North experienced a higher incidence of poverty than Maputo City. Inhambane in the South was the poorest province.

Using poverty lines which do not vary by geographical regions as well as focusing the analysis on the non-money-metric dimension of well-being led to somewhat different findings. The gap in the incidence of poverty between rural and urban regions was much wider than found before. The South was the least poor region, and in particular Maputo Province and Maputo City were the least poor

provinces. Zambezia was one of the poorest provinces when the constant money-metric poverty lines were employed, but it was the poorest province for the case of the non-money-metric asset index. As discussed in Chapter 2, this would suggest that anti-poverty programmes should be targeted at Zambezia, in particular given that it is one of the most populous provinces of Mozambique.

With regards to the correlates of poverty, age-dependency ratio, household size, rural areas, female heads of household and illness of household head were factors positively associated with poverty. In other words, controlling for other factors, the ones listed above tended to be associated with a higher incidence of poverty. Education of head, on the other hand, was associated with lower levels of poverty amongst the members of the household. In fact, it came as the most important poverty marker, followed by living in rural areas. A policy implication from these findings is that concerted efforts should be made at targeting appropriate programmes and policy at poverty in rural areas. In turn, such programmes and policies should foster and create enabling conditions for a greater access to quality education for all, but in particular in rural areas.

The national level of Gini inequality was 0.48 employing the consumption aggregate, and 0.30 employing the wealth index. Urban regions were more unequal than the rural in the former case, but inequality rankings could not be made in the latter case because there was no Lorenz dominance. The money-metric approach which employs the consumption aggregate suggested that the South was the most unequal region, with Maputo City being the most unequal province. But the asset index suggested that the South was the least unequal region, with Maputo Province and Maputo City being the least unequal provinces in terms of asset holdings. Most of the inequality across the various groups, e.g., across provinces, locations and regions, was within-group inequality. Very small proportions of the overall inequality were attributable to between-group inequality.

Chapter 3 aimed to demonstrate that schooling is associated with poverty alleviation through the labour market. In particular, the objective was to show that in developing and poor countries such as Mozambique, which are characterised by a very segmented labour market, schooling influences the type of job one gets, and conditional on that, schooling is associated with higher earnings. The methods applied included a multinomial logit model and a modified version of Dubin & McFadden's (1984) model.

In this chapter it was found that an individual's education increased the chances that he or she would work in the wage public sector. Schooling also tended to be associated with private wage employment, but for women only. In contrast, schooling was negatively associated with jobs outside the wage segment. Schooling tended to be negatively associated with self-employment activities and unpaid family work. Another interesting finding was that schooling could not always guard against unemployment. It was also found that education was positively associated with greater earnings in the public, private and self-employment segments. Returns to education were found to be greater for women than men in the private sector, but were greater for men than women in the public sector. In self-employment activities education mattered for men only, in particular secondary education and above. Returns to education were found to be convex, and correcting the earnings functions for sample selection bias in a multinomial fashion did improve the earnings functions compared to those based on ordinary least squares (OLS) regressions. Compared to the OLS results, the corrected earnings functions presented greater returns to education for men and women in the public sector and for men only in the private sector.

Because education quality has a bearing on an individual's performance in the labour market and thus affects the role of education in alleviating poverty, **Chapter 4** aimed to identify the correlates of education quality in Mozambique. In particular, given the global agenda of Education For All which fostered a rapid expansion of enrolments in the education system, this chapter aimed to separate the effects of a change in education quantity from those of a change in education quality over time. In this chapter the methods employed were education production functions based on OLS multivariate regressions and the Oaxaca-Blinder decomposition technique (Oaxaca, 1973; Blinder, 1973).

With regards to the findings, this chapter found that most of the correlates of cognitive educational achievement suggested by the literature were indeed associated with educational outputs. Belonging to a high socio-economic status (SES), being a young pupil, being a boy, having a greater number of books at home, having parents that help with school work, having supper regularly, lower absenteeism and lower grade repetition levels, and speaking the language of instruction were factors positively associated with pupil test scores. In addition, being in a school headed by a principal with high levels of education and with training in specialised school management, having peers from high SES, having access to textbooks and having teachers that spend long hours marking and preparing classes were characteristics also positively associated with pupil performance. Pupils in

bigger classes, however, tended to perform worse on average compared to those in smaller classes. Another important finding was that while SACMEQ III pupils performed worse than SACMEQ II pupils on average, the average fall in reading and numeracy test scores over time was larger for low SES pupils than for high SES pupils.

With respect to the reasons behind the average deterioration in education quality over time, the Oaxaca-Blinder decomposition technique showed that the endowments of productive characteristics did not change much between SACMEQ II and SACMEQ III. The coefficients effect, however, was very large, thus explaining most of the average deterioration in education quality. The detailed decomposition of the coefficients effect, while criticised in the literature, suggested that it was not the coefficients of the correlates of educational achievement that declined between surveys, but instead the intercept of the respective education production functions.

Either way, the findings from the empirical work presented in Chapter 4 suggest that the issue with the fall in average education quality resides in the education system. Given the relatively high net enrolment rates observed lately at the primary level of education, it is time to start putting greater emphasis on education quality. Issues of school efficiency and of lenient grade promotions should be addressed immediately, otherwise Mozambique runs the risk of having a huge mass of functionally innumerate and illiterate primary school graduates.

The next section summarises what we can conclude from the main findings as well as from the theses statements.

5.2 Conclusion on the Theses Statements

This thesis has demonstrated in **Chapter 2** that indeed the poverty and inequality profiles suggested by earlier studies are inappropriate to inform policy. Using constant money-metric poverty lines as well as using a wealth index based on asset holdings, the spatial distributions of poverty and inequality differ from that suggested by earlier studies. One of the earlier studies is DNPO et al. (2004), a government report on poverty and well-being in Mozambique that estimated poverty for the period 2002-2003. PARPA II (*Plano de Acção para a Redução da Pobreza Absoluta,* the second Poverty Reduction Strategy Paper – PRSP) may have been wrongly informed. If efforts aimed at the reduction of poverty were based on DNPO et al. (2004), they may have been concentrated on the wrong regions. This was detrimental to the poorest, in particular those living in regions mistakenly labelled as relatively better off.

In **Chapter 3** this study demonstrated that modelling the Mozambican labour market as one made of wage employment only is not appropriate. Mozambique but also other low income economies have a structure of labour markets which leads to segmentation. The wage segment is very small. A huge proportion of the working-age population is engaged in non-wage activities and would thus fall out of the analysis. This would bias the employment estimates. Using a multinomial logit model this study demonstrated that schooling has an influence on the choice of employment segment. For instance, the effect of schooling on an individual's chances of having a public sector job were found to be different from the effect of schooling on his or her chances of engaging in self-employment activities. It also demonstrated that, on analysing the relationship between education and earnings, we should correct the earnings functions for sample selection bias, in particular accounting for the multiple employment segments that characterise this type of economy.

In **Chapter 4** the thesis statement with regards to the reasons behind the average fall in education quality over time could not be verified. Thus, there is no evidence that it was the deterioration in the socio-economic background of the pool of pupils that caused the average fall in mean scores over time, as the endowments effect was small and statistically insignificant. In other words, the average characteristics of pupils did not change much between the years 2000 and 2007. The likely reasons behind the mentioned deterioration in education quality are summarised in the following paragraphs.

First, there is some evidence pointing towards a decline in the manner in which pupil learning characteristics were valued, which represents a decline in the efficiency of the education system over time. The evidence on this matter was the fall in the coefficients between SACMEQ II and SACMEQ III. This conclusion is appropriate if the interpretation of the Oaxaca-Blinder decompositions does not go beyond the threefold "aggregate" decomposition, as mentioned. But if the detailed decomposition of the coefficients effect is valid, a potential conclusion is that efforts aimed at a rapid expansion of access to education in Mozambique were associated with the mentioned average deterioration in education quality. Standards could perhaps have declined because the education system tended to increase the promotion of lower ability pupils over time. For instance, in 2004 an education policy of "automatic" grade promotion was approved. The primary level of education was split into cycles, within which pupils are automatically promoted to the following grade, irrespective of whether or not they meet the learning requirements for promotion. Only in the last year of each cycle could they be held back.

Another potential conclusion emanating from the detailed decomposition is ability bias. Dropout rates in the year 2000 were much higher than dropout rates in 2007. If dropouts would have scored below average in the cognitive tests, then the greater proportion of dropouts in SACMEQ II might have upwardly biased the tests scores, thus making it look as if average test scores had fallen over time.

5.3 Contributions

This thesis has made three main contributions to the accumulation of new knowledge. In **Chapter 2** the main contribution was to show that employing conventional or best practice approaches to analyse well-being does not always lead to reliable results. Existing analyses that did this lead to further urban bias in well-being and therefore exacerbate the existing patterns of poverty and inequality. If the intention is to alleviate poverty and inequality, future studies that analyse these phenomena in Mozambique should look also at other dimensions of well-being, in the manner undertaken in this study. If results are not also confirmed by analysis of other indicators of well-being, one should be sceptical of simply applying best practice approaches. In future studies one should also be more careful on unquestioningly using variable poverty lines. In other words, results should be considered against other evidence such as non-money-metric indicators. This is more important if the collection of food expenditure and price data is subject to the usual measurement errors. This is likely to be the case in Mozambique, judging by the provincial samples which were not very large.

The main contribution made in **Chapter 3** was to present the first empirical estimation of returns to education in Mozambique, in particular accounting for the existing multiple employment segments that characterise the labour market, as well as accounting for sample selection bias. This has not yet been done before.

Lastly, in **Chapter 4** the main contribution was to learn that the average deterioration in education quality in Mozambique between the years 2000 and 2007 was not necessarily the consequence of the massive influx of pupils from disadvantaged socio-economic backgrounds. Showing empirically that the mentioned average deterioration in education quality could be associated with a decline in the efficiency of the Mozambican education system or with more lenient pupil promotion policies are other contributions of this chapter. No studies in the SACMEQ region have yet tried to address

the issue of fast increases in enrolments and that of a change in education quality, particularly using the methods employed in this thesis.

5.4 Suggestions for Further Research

The third Mozambican income and expenditure household survey which collected data between September 2008 and August 2009 was recently released for public use. In 2010 the government released a new report on poverty and well-being in Mozambique using this survey (Direcção Nacional de Estudos e Análise de Políticas, 2010). An avenue for further research would be to do similar work as was done in Chapter 2 of this thesis, but employing the new data set as well as the new report. This would indicate whether the government analysis strengthens the urban bias or whether it moved closer to identifying the poorest better.

Another avenue for further research is to use the exiting data to try to identify causal effects in the returns to education in Mozambique, but still accounting for the segmentation of the labour market as well as for sample selection bias. The instrumental variables (IV) approach is a suitable method for this. If causal effects of education on earnings could be made, this would strengthen the case for education in alleviating poverty and inequality in Mozambique.

List of References

Anand, S. (1977). Aspects of Poverty in Malaysia. *Review of Income and Wealth , 23*, 1-16.

Asselin, L. (2002). Multidimensional Poverty: Composite Indicator of Multidimensional Poverty. Québec: Institut de Mathématique Gauss.

Banco de Moçambique. (2012, Julho). Conjuntura Económica e Perspectivas de Inflação. Maputo.

Bata, R. (2006). O Discurso Oficial Sobre o Desemprego no Contexto do Programa de Reabilitação Económica e Social (PRES) - 1989 a 2004. *Universidade Eduardo Mondlane, Unpublished Undergraduate Thesis*.

Becker, G. (1975). *Human Capital.* Chicago: University of Chicago Press.

Bhorat, H., & Leibbrandt, M. (2001). Modelling Vulnerability and Low Earnings in the South African Labour Market. In H. Bhorat, M. Leibbrandt, M. Maziya, S. Van der Berg, & I. Woolard (Eds.), *Fighting Poverty: Labour Markets and Inequality in South Africa* (pp. 107-129). Lansdowne: UCT Press.

Bhorat, H., & McCord, A. (2003). Employment and labour market trends in South Africa. In HSRC (Ed.), *Human Resources Development Review.* Cape Town: HSRC Press.

Blasius, J., & Greenacre, M. (2006). Correspondence Analysis and related methods in practice. In J. Blasius, & M. Greenacre (Eds.), *Multiple correspondence analysis and related methods.* London: Chapman & Hall.

Blinder, A. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources, 8* (4), 436-455.

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

Bourguignon, F., Fournier, M., & Gurgand, M. (2004). Selection Bias Corrections based on the Multinomial Logit Model: Monte-Carlo Comparisons. *Département et Laboratoire d'Économie Théorique et Appliquée, Working Paper No. 20*.

Bourguignon, F., Fournier, M., & Gurgand, M. (2007). Selection bias corrections based on multinomial logit model: Monte-Carlo comparisons. *Journal of Economic Surveys, 21* (1), 174-205.

Cameron, A., & Trivedi, P. (2009). *Microeconometrics Using Stata.* Texas: Stata Press.

Chichava, S. (2007). Uma Província Rebelde: O Significado do Voto Zambeziano a Favor da RENAMO. *Inaugural Conference of the Instituto de Estudos Sociais e Económicos (IESE).* Maputo.

Clark, S., Hemming, R., & Ulph, D. (1981). On indices for the measurement of poverty. *Economic Journal, 91*, 515-526.

Cohn, E., & Geske, T. (1990). *Economics of Education.* Oxford: Pergamon Press.

Conceição, P., & Ferreira, P. (2000). The Young Person's Guide to the Theil Index: Suggesting Intuitive Interpretations and Exploring Analytical Applications. *UTIP* .

Cowell, F. (1985). Measures of Distributional Change: An Axiomatic Approach. *Review of Economic Studies, 52*, 135-151.

Da Maia, C., & Van der Berg, S. (2010). When the remedy is worse than the disease: adjusting survey income data for price differentials, with special reference to Mozambique. *Department of Economics, Stellenbosch University* .

Dahl, G. (2008). Mobility and the returns to education: Testing a Roy Model with multiple markets. *Econometrica, 70* (6), 2367-2420.

Deaton, A. (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Baltimore: Johns Hopkins University Press for the World Bank.

Deaton, A., & Grosh, M. (1998). The Consumption module in the Living Standards Measurement Survey. In P. Glewwe, & M. Grosh (Eds.), *Designing household survey questionnaires for developing countries.* Washington, D.C.: World Bank.

Deaton, A., & Zaidi, S. (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis. *Living Standards Measurement Study* . Washington, D.C.: World Bank.

Demery, L., & Grootaert, C. (1993). Priority Analysis. In L. Demery, M. Ferroni, C. Grootaert, & J. Wong-Valle (Eds.), *Understanding the Social Effects of Policy Reform.* Washington D.C.: The World Bank.

Direcção Nacional de Estudos e Análise de Políticas. (2010). *Pobreza e Bem-estar em Moçambique: Terceira Avaliação Nacional.* Maputo.

Dubin, J., & McFadden, D. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica, 52* (2), 345-362.

Falco, P., Kerr, A., Rankin, N., Sandefur, J., & Teal, F. (2009). The Returns to Formality and Informality in Urban Africa. *CSAE WPS/2010-03*, 1-30.

Fields, G. (2000). *Distribution and Development: A new look at the developing world.* Cambridge: MIT Press.

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

Foster, J., Greer, J., & Thorbecke, E. (1984). A Class of Decomposable Poverty Measures. *Econometrica, 52* (3), 761-766.

Fox, L., Bardasi, E., & Van den Broeck, K. (2005). *Poverty in Mozambique: Unravelling Changes and Determinants.* World Bank. Washington, D.C.: Africa Region Working Paper Series No. 87.

Francisco, A. (2002). Evolução da Economia de Moçambique da Colónia à Transição para a Economia de Mercado. In C. Rolim, A. Franco, B. Bolnick, P.-A. Anderson, & M. d. Gabinete de Estudos (Ed.), *A Economia Moçambicana Contemporânea: Ensaios* (pp. 15-43). Maputo: Imprensa Universitária, UEM.

Glewwe, P. (1991). *Schooling, Skills, and the Returns to Government Investment in Education: An Exploration Using Data from Ghana.* Washington, D. C.: World Bank.

Glick, P., & Sahn, D. (1993). Labor Force Participation, Sectoral Choice, and Earnings in Conakry, Guinea. *Cornell Food and Nutrition Policy Program Working Paper Series , Working Paper no. 43*.

Glick, P., & Sahn, D. (1997). Gender and Education Impacts on Employment and Earnings in West Africa: Evidence from Guinea. *Economic Development and Cultural Change, 45* (4), 793-823.

Greenacre, M. (2007). *Correspondence Analysis in Practice.* New York: Chapman & Hall/CRC.

Greene, W. (2003). *Econometric Analysis.* New Jersey: Prentice Hall.

Handa, S. & Simler, K. (2005). Quality or Quantity? The Supply-side Determinants of Primary Schooling in a Poor Rural Economy. *Journal of African Economies*, *15* (1), 59-90.

Hanlon, J. (1997). *Paz sem benefícios: como o FMI bloqueia a reconstrução de Moçambique.* Maputo: Centro de Estudos Africanos.

Hanlon, J. (2007). Is Poverty Decreasing in Mozambique? *Inaugural Conference of the Instituto de Estudos Sociais e Económicos (IESE).* Maputo.

Hanlon, J. (2010). Mozambique: 'the war ended 17 years ago, but we are still poor'. *Conflict, Security & Development, 10* (1), 77-101.

Haughton, J., & Khandker, S. (2009). *Handbook on Poverty and Inequality.* Washington, DC: The World Bank.

Heavily Indebted Poor Countries (HIPC) Initiative. (s.d.). Retrieved on 07 August 2012, from The World Bank: http://go.worldbank.org/6RGY6ATR40

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica , 47* (1), 153-161.

Heneveld, W. (1994). *Planning and Monitoring the Quality of Primary Education in Sub-Saharan Africa.* Washington, D.C.: The World Bank.

Heston, A., Summers, R., & Aten, B. (2006, September). *Penn World Table Version 6.2*. (Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania) Retrieved March 3, 2010, from http://pwt.econ.upenn.edu/php_site/pwt62/pwt62_form.php

Hofstee, E. (2006). *Constructing a Good Dissertation: A Practical Guide to Finishing a Masters, MBA or PhD on Schedule.* Johannesburg, South Africa: EPE.

Instituto Nacional de Estatística (INE). (n.d.). Retrieved March 15, 2010, from http://www.ine.gov.mz/

International Monetary Fund. (April de 2012). Retrieved August 08, 2012, from World Economic Outlook Database:

http://www.imf.org/external/pubs/ft/weo/2012/01/weodata/weorept.aspx?sy=1980&ey=2012&scsm=1&ssd=1&sort=country&ds=.&br=1&pr1.x=75&pr1.y=12&c=688&s=NGDP_RPCH&grp=0&a=#cs1.

Jann, B. (2008). The Blinder-Oaxaca decomposition of linear regression models. *The Stata Journal, 8* (4), 453-479.

Johnes, G. (1993). *The Economics of Education.* London: Macmillan Press.

Jones, F. (1983). On Decomposing the Wage Gap: A Critical Comment on Blinder's Method. *The Journal of Human Resources, 18* (1), 126-130.

Jones, F., & Kelley, J. (1984). Decomposing Differences Between Groups: A Cautionary Note on Measuring Discrimination. *Sociological Methods & Research, 12* (3), 323-343.

Kakwani, N. (1977). Measurement of Poverty and the Negative Income Tax. *Australian Economic Papers, 16*, 237-248.

Kakwani, N. (1980). On a Class of Poverty Measures. *Econometrica, 48* (2), 437-466.

Kingdon, G., & Knight, J. (2003). Well-being poverty versus income poverty and capabilities poverty. Centre for the Study of African Economies: Oxford.

Koch, S., & Ntege, S. (2008). Returns to Schooling: Skill Accumulation or Information Revelation? *University of Pretoria, Department of Economics Working Paper Series, Working Paper No. 12*.

Krishnan, P., Sellasie, T., & Dercon, S. (1998). The urban labor market during structural adjustment: Ethiopia 1990-1997. *Oxford University, Centre for the Study of African Economies, CSAE WPS/98-99*.

Kuepie, M., Nordman, C., & Roubaud, F. (2007). Education and Labour Market Outcomes in Sub-Saharan West Africa.

Lanjouw, P. (2009). Constructing a Consumption Aggregate for the Purpose of Welfare Analysis: Principles, Issues and Recommendations Arising from the Case of Brazil. *Measuring Poverty, Income Inequality and Social Exclusion: Lessons from Europe.* Paris.

Lee, L. (1983). Generalized econometric models with selectivity. *Econometrica , 51* (2), 507-512.

Litchfield, J. (1999). Inequality: Methods and Tools. Text for World Bank's Web Site on Inequality, Poverty, and Socio-economic Performance: http://www.worldbank.org/poverty/inequal/index.htm.

MacFarquhar, E. (2 de May de 1988). The killing fields of Mozambique. *U.S. News & World Report, 104*, p. 45.

Madden, D., & Smith, F. (2000). Poverty in Ireland, 1987-1994: A Stochastic Dominance Approach. *The Economic and Social Review, 31* (3), 187-214.

Mansur, E., Mendelsohn, R., & Morrison, W. (2008). Climate Change Adaptation: A study of fuel choice and consumption in the US energy sector. *Journal of Environmental Economics and Management, 55*, 175-193.

Maximiano, N., Arndt, C., & Simler, K. (2005). Qual foi a Dinâmica dos Determinantes da Pobreza em Moçambique? Maputo, Moçambique: Direcção Nacional do Plano e Orçamento, Ministério da Planificação e Desenvolvimento.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics.* New York: Academic Press.

Mincer, J. (1974). *Schooling, Experience and Earnings.* New York: National Bureau of Economic Research.

Ministério do Plano e Finanças. (1998). *Pobreza e Bem-estar em Moçambique: Primeira Avaliação Nacional (1996-1997).* Maputo.

Ministry of Education. (1996). *Development of Education.* Maputo: Directorate of Planning.

Ministry of Education. (1996). *Development of Education.* Maputo: Directorate of Planning.

Mole, P. (1994). O Impacto Social do Programa de Ajustamento Económico. In C. Castel-Branco, *Moçambique: Perspectivas Económicas* (pp. 158-174). Maputo: Imprensa Universitária, UEM.

Mouton, J. (2001). *How to succeed in your Master's & Doctoral Studies.* Pretoria: Van Schaik Publishers.

Mozambique Education Statistics 1998 - 2009, Ministério da Educação e Cultura (MEC), retrieved on 21st April 2012, from http://www.mec.gov.mz/img/documentos/20100319020325.pdf

National Directorate of Planning and Budget; Ministry of Planning and Finance; Economic Research Bureau; International Food Policy Research Institute; Purdue University. (2004). *Poverty and Well-Being in Mozambique: The Second National Assessment.* Discussion paper No. 3E, Maputo - Mozambique.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review, 14* (3), 693-709.

Oaxaca, R., & Ransom, M. (1999). Identification in Detailed Wage Decompositions. *The Review of Economics and Statistics, 81* (1), 154-157.

PNUD. (1998). *Moçambique: Paz, crescimento económico, oportunidades para o desenvolvimento humano.* Maputo: Relatório Nacional do desenvolvimento Humano.

PreventionWeb. (2012). Obtido em 08 de August de 2012, de Mozambique - Disaster Statistics: http://www.preventionweb.net/english/

Psacharopoulos, G. (1981). Returns to Education: An Updated International Comparison. *Comparative Education, 17* (3), 321-341.

Psacharopoulos, G., & Patrinos, H. (2004). Returns to Investment in Education: A Further Update. *Education Economics, 12* (2), 111-134.

Rankin, N., Sandefur, J., & Teal, F. (2010). Learning & Earning in Africa: Where are the Returns to Education High? *CSAE WPS/2010-02*, 1-24.

Ravallion, M. (1996). Issues in Measuring and Modelling Poverty. *The Economic Journal , 106* (438), 1328-1343.

Ravallion, M., & Bidani, B. (1994). How Robust Is a Poverty Profile? *The World Bank Economic Review, 8* (1), 75-102.

República de Moçambique. (2001). *Plano de Acção para a Redução da Probreza Absoluta, 2001-2005.* Maputo.

República de Moçambique. (2006). *Plano de Acção para a Redução da Pobreza Absoluta, 2006-2009.* Maputo.

Riddell, A. (2008). *Factors Influencing Educational Quality and Effectiveness in Developing Countries: A Review of Research.* Eschborn: GTZ.

Sahn, D., & Stifel, D. (2000). Poverty Comparisons Over Time and Across Countries in Africa. *World Development, 28* (12), 2123-2155.

Schmertmann, C. (1994). Selectivity bias correction methods in polychotomous sample selection models. *Journal of Econometrics, 60* (1), 101-132.

Sen, A. (1976). An Ordinal Approach to Measurement. *Econometrica, 44* (2), 219-231.

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

Sen, A. (1987). *The Standard of Living.* Cambridge: Cambridge University Press.

Sen, A. (2001). *Development as freedom.* Oxford: Oxford University Press.

Simione, F. (2011). *Labor Market Returns to Education in Mozambique: Learning From a Quase-Experiment.* Maputo: National Directorate of Studies and Policy Analysis, Ministry of Planning and Development.

Statacorp. (2009). Stata Statistical Software: Release 11.0. Texas: Stata Corporation.

Statacorp. (2011). Stata Statistical Software: Release 12.0. Texas: Stata Corporation.

Taylor, S., & Yu, D. (2009). The importance of socio-economic status in determining educational achievement in South Africa. Stellenbosch working Paper Series No. WP01/2009.

Theil, H. (1967). *Economics and Information Theory.* Chicago: Rand McNally and Company.

UNESCO. (2009). *Overcoming inequality: why governance matters.* Paris and Oxford: United Nations Educational Scientific and Cultural Organization (UNESCO) and Oxford University Press.

UNESCO. (2010). *EFA Global Monitoring Report: Reaching the Marginalized.* Paris & Oxford: United Nations Educational & Oxford University Press.

Van der Berg, S. (2008). *Poverty and education.* Paris and Brussels: The International Institute for Education Planning (IIEP) and The International Academy of Education (IAE).

Van Ginneken, W. (1980). Some Methods of Poverty Analysis: An Application to Iranian Data, 1975-76. *World Development, 8*, 639-646.

Vijverberg, W. (1992). Dual Selection Criteria with Multiple Alternatives: Migration, Work Status, and Wages. *Living Standards Measurement Study WP 78*. The World Bank.

Vijverberg, W. (1993). Educational Investments and Returns for Women and Men in Côte d'Ivoire. *Journal of Human Resources, 28* (4), 933-974.

Vijverberg, W. (1995). Dual Selection Criteria with Multiple Alternatives: Migration, Work Status, and Wages. *International Economic Review , 36* (1), 159-185.

Vyas, S., & Kumaranayake, L. (2006). Constructing socio-economic indices: how to use principal components analysis. 459-68. Oxford University Press in association with The London School of Hygiene and Tropical Medicine.

Wambugu, A. (2002). Education, Employment and Earnings in Kenya. *Göteborg University, Department of Economics*.

Watts, H. (1968). An economic definition of poverty. In *On Understanding Poverty* (D. P. Moynihan ed.). New York: Basic Books.

Winsborough, H., & Dickenson, P. (1971). Components of negro-white income differences. *American Statistical Association*, (pp. Section 6-8).

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

World Bank. (2003). Measuring living standards: Household consumption and wealth indices. Quantitative techniques for health equity analysis. Washington, D.C.: World Bank.

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

World Bank. (2007). *Beating the Odds: Sustaining Inclusion in a Growing Economy - A Mozambique Poverty, Gender and Social Assessment.* Poverty Reduction and Economic Management. Washington, D.C.: World Bank.

Yun, M. (2005). A Simple Solution to the Identification Problem in Detailed Wage Decompositions. *Economic Inquiry, 43*, 766-772.

Zimmerman, C. (1932). Ernst Engel's Law of Expenditures for Food. *The Quarterly Journal of Economics, 47* (1), 78-101.

Appendix A – Avoiding Double-Counting

When computing the consumption aggregate, there was the risk of double-counting products. There were similar products present in both the second and the third questionnaires. While the second questionnaire gathered information on household daily expenses on food, the third obtained information with respect to general expenditures and revenues of the household, which included yearly and monthly information.

The following table shows a list of goods appearing in both questionnaires 2 and 3, and respective codes. With an "X" it also shows from which questionnaire the products were deleted. The approach followed to deal with this issue was to delete the data for the product which was being double-counted from the questionnaire that reported the lowest frequency for that product. For example, if new matches were reported to be consumed twice a week in the second questionnaire, but four times a week in the third questionnaire, the consumption data on new matches was deleted from the second questionnaire. Doing it in this manner avoided biasing the consumption aggregate downwards.

Table A 1 – Products creating potential for double-counting

Product	Product Code	Monthly Expenses	Daily Expenses
Wild animal meat	011272	X	
Sausages	011286		X
Accessories for clothing	031340		X
Piped water	044101		X
Self-transported water	044102	X	
Electricity from grid	045101		X
Petroleum	045301	X	
Wood	045401	X	
Charcoal	045402	X	
Electronic home accessories	055206		X
Cleaning products	056110	X	
Liquid detergents	056112		X
Powder detergents	056114		X
Soap and bleach	056116		X
Disinfectants	056117		X
Matches	056125		X
Candles	056128	X	
Milling services	056230		X
Petrol for private cars	072201		X
Diesel for private cars	072202	X	
Other lubricants	072207		X
Transportation by train	073100		X
Transportation by car	073200		X
Transportation by boat	073401	X	
Seeds of food crops	093306		X
Bathing soap	121331		X
Pomade	321325		X
Razors	321341		X

Appendix B – Computing the Asset Index

Vyas & Kumaranayake (2006, p. 461) argued that there is no best practice approach in selecting the quantity and the right variables to include in the computation of the wealth index. In order to compute the asset index this study benefited from the first and the third questionnaires. The first questionnaire provided information on the material walls and roofs were made of. Information on the sanitation system and households' access to public services (water and electricity) was also of great importance. Table A 2 describes this information.

The third questionnaire contains information on ownership of durable private assets by each of the households. For instance, from a list of 23 durable private assets, it tells whether or not a particular household owns, among other items, a radio, a bicycle, a bed, a computer and/or a cellular phone. This information was also used in the compilation of the wealth index.

To derive the asset index, firstly all private durable assets were transformed into binary categorical variables where 1 stands for "yes the household owns the durable private asset", and 0 for "no". The housing characteristics variables and the variables on access to water and electricity were not transformed into dummy variables, as Stata, the software employed to derive the asset index, does that automatically for categorical variables. This resulted in the indicator matrix.

The next step was to calculate the household profile with respect to the categories of asset ownership. Fortunately, Stata11 (Statacorp, 2009) has a routine to estimate these profiles. The *mca* command applies MCA on the created matrix. Table A 4 shows the results of MCA applied on the matrix. According to it, total inertia equals 0.0298 and the first dimension alone explains 68.1% of this inertia. That means that by sacrificing the remaining dimensions only 31.9% of the inertia of the profile points has been lost. Another way of interpreting this result is that many of the sacrificed row profiles lie very close to the plane of representation, meaning that, when exploring their relative positions, their distance from the plane can effectively be ignored (Greenacre, 2007, p. 66).

Table A 2 – Households housing characteristics and weights in the asset index

Housing Characteristics	Freq.	%	Cum %	*Weights*
Wall material				
Adobe blocks	*2,181*	*25.07*	*25.07*	*0.870*
Cement blocks	*1,657*	*19.05*	*44.11*	*-3.276*
Bricks	*395*	*4.54*	*48.66*	*-2.325*
Wattle and daub	*2,681*	*30.82*	*79.47*	*1.015*
Wood/sticks/zinc	*207*	*2.38*	*81.85*	*-1.094*
Reed/bamboo/palm	*1,473*	*16.93*	*98.78*	*0.332*
Sacks/cans/box cards	*12*	*0.14*	*98.92*	*0.976*
Others	*73*	*0.84*	*99.76*	*0.586*
Missing Observations	*21*	*0.24*	*100.00*	-
Total	*8,700*	*100.00*		-
Roof material				
Concrete	*249*	*2.86*	*2.86*	*-6.748*
Tile	*63*	*0.72*	*3.59*	*-7.853*
Lusalite (fibre-cement foils)	*391*	*4.49*	*8.08*	*-4.002*
zinc foils	*2,813*	*32.33*	*40.41*	*-1.583*
Grass/palm	*5,109*	*58.72*	*99.14*	*1.104*
Other	*71*	*0.82*	*99.95*	*0.511*
Missing Observations	*4*	*0.05*	*100.00*	-
Total	*8,700*	*100.00*		-
Sanitation system				
None (bushes)	*3,365*	*38.68*	*38.68*	*1.190*
Flush toilet with sewerage	*237*	*2.72*	*41.4*	*-8.061*
Flush toilet with septic tank	*424*	*4.87*	*46.28*	*-6.136*
Improved pit latrine	*1,055*	*12.13*	*58.4*	*-2.485*
Unimproved pit latrine	*3,558*	*40.90*	*99.30*	*0.157*
Other	*59*	*0.68*	*99.98*	*0.070*
Missing Observations	*2*	*0.02*	*100.00*	-
Total	*8,700*	*100.00*		-
Water source				
Piped (internal or in the yard)	*813*	*9.34*	*9.34*	*-6.003*
Public tap	*954*	*10.97*	*20.31*	*-0.617*
Protected water well/hole	*1,942*	*22.32*	*42.63*	*0.442*
Unprotected well, rain water	*2,941*	*33.80*	*76.44*	*1.010*
River/lake/lagoon	*1,158*	*13.31*	*89.75*	*1.307*
Buy from others/cistern	*299*	*3.44*	*93.18*	*-1.497*
Other	*557*	*6.40*	*99.59*	*-2.327*
Missing Observations	*36*	*0.41*	*100.00*	-
Total	*8,700*	*100.00*		-
Lighting Energy Source				
Oil/Paraffin	*4,778*	*54.92*	*54.92*	*0.350*
Gas	*3*	*0.03*	*54.95*	*-1.052*

Electricity form grid	*1,064*	*12.23*	*67.18*	*-5.614*
Generator/Solar	*13*	*0.15*	*67.33*	*-2.393*
Battery	*6*	*0.07*	*67.4*	*-1.470*
Candles	*441*	*5.07*	*72.47*	*-1.432*
Firewood	*2,158*	*24.80*	*97.28*	*1.410*
Other	*234*	*2.69*	*99.97*	*1.383*
Missing Observations	*3*	*0.03*	*100.00*	-
Total	*8,700*	*100.00*		-
Cooking Energy Source				
Firewood	*6,615*	*76.03*	*76.03*	*0.850*
Charcoal	*1,627*	*18.70*	*94.74*	*-2.694*
Oil/Paraffin	*111*	*1.28*	*96.01*	*-2.976*
Gas	*207*	*2.38*	*98.39*	*-8.622*
Electricity from grid	*109*	*1.25*	*99.64*	*-8.399*
Sawdust	*16*	*0.18*	*99.83*	*-0.165*
Other	*12*	*0.14*	*99.97*	*0.195*
Missing Observations	*3*	*0.03*	*100.00*	-
Total	*8,700*	*100.00*		-

Table A 2 and Table A 3 report the first dimension MCA weights applied to the categories. These tables show that those components that reflect higher socio-economic status contribute negatively to the asset index, while components that reflect lower standards of living contribute positively to the index. For instance, on the one hand, owning a durable private asset, having brick or cement block walls, having a concrete or tile roof, having a flush toilet, having access to piped water, or having electricity as a source of energy decreases a household's asset index score. On the other hand, not owning a durable private asset, having sacks, cans or box cards for walls, having a grass or palm roof, having only access to rain or well water, using firewood as a source of energy, or simply not having a sanitation system, increase a household's asset index score. These weights suggest that this MCA asset index is, in fact, a poverty index rather than a wealth index.

Again, this is not reason for concern since it can still be used in the present analysis. Its interpretation is simply reversed. Simply put, if the index is left unchanged, households that form part of the 1st quintile have the highest socio-economic status, while the ones in the 5th quintile have the lowest possible socio-economic status. To make the analysis easier, a simple and logically obvious way to get around this problem is to transform the poverty index into a wealth index just by multiplying it by –1. This is well supported by Greenacre (2007, p. 140) who, for instance, states that "it is always possible to reverse an axis (i.e., multiply all coordinates by –1)".

Table A 3 – Assets and weights in the asset index

Variable and Categories	Weights
Bed	
no	*0.149*
yes	*-2.599*
Car new	
no	*0.010*
yes	*-10.768*
Second hand car	
no	*0.033*
yes	*-9.328*
Motorcycle	
no	*0.000*
yes	*-0.109*
Bicycle	
no	*-0.099*
yes	*0.601*
Radio	
no	*-0.076*
yes	*0.338*
Sound system	
no	*0.075*
yes	*-2.373*
TV	
no	*0.118*
yes	*-6.711*
Laundry Machine	
no	*0.003*
yes	*-14.159*
Air Conditioner	
no	*0.003*
yes	*-14.832*
Sewing Machine	
no	*0.006*
yes	*-1.161*
Refrigerator	
no	*0.054*
yes	*-12.630*
Freezer	
no	*0.088*
yes	*-9.226*
Ironing Machine	
no	*0.125*
yes	*-9.506*

Fan	
no	*0.102*
yes	*-8.203*
Telephone	
no	*0.021*
yes	*-8.298*
Mobile Phone	
no	*0.170*
yes	*-7.295*
Computer	
no	*0.012*
yes	*-15.809*
Printer	
no	*0.007*
yes	*-11.100*
Clock	
no	*0.145*
yes	*-0.740*
Electrical Stove	
no	*0.056*
yes	*-9.159*
Gas stove	
no	*0.034*
yes	*-11.784*
Mixed stove	
no	*0.006*
yes	*-2.874*

Table A 5 shows the descriptive statistics for the poverty index, the wealth index and for the transformed wealth index. According to this table, the wealth index is simply the negative opposite of the poverty index. This means that the interpretation of the asset index (wealth index) is now the usual one: households that form part of the 1st quintile have the lowest socio-economic status, while the ones in the 5th quintile have the highest socio-economic status.

Booysen et al. (2008, p. 1118) quoted the World Bank (2003) arguing that the correlation coefficients between this type of indices and expenditure usually range between 0.20 and 0.40. In this study it was found that the correlation between the wealth index and total consumption expenditure was 0.55, therefore suggesting that this index might be a very good indicator of household socio-economic status. However, the index does show signs of bias against rural households since, though positive, its correlation with total consumption expenditure in urban zones

(0.56) is much higher than in the rural parts of the country (0.29). Lastly, as negative and zero values might complicate the poverty analysis, the wealth index was transformed in order to include only positive values, following Booysen et al. (2008, p. 1117). According to the authors, a number slightly higher than the greatest absolute negative value is added to each household's wealth index score. According to Table A 5, the greatest absolute negative value for the wealth index is *2.76166*. Thus the approach followed in this study was to add *2.76167* to each of the households' wealth index score, making it positive for all the surveyed households (see this transformed wealth index in Table A 5). This index was then used for the poverty and inequality analyses.

Table A 4 – MCA applied to the matrix

Dimension	*Principal Inertia*	*%*	*Cumul %*
dim 1	*0.0203249*	*68.13*	*68.13*
dim 2	*0.0028397*	*9.52*	*77.64*
dim 3	*0.0011298*	*3.79*	*81.43*
dim 4	*0.0005011*	*1.68*	*83.11*
dim ...	...	...	...
dim 23	*0.0000000738*	*0.00*	*85.61*
Total	*0.029834*	*100.00*	

Number of observations: 6562; Number of axes: 2

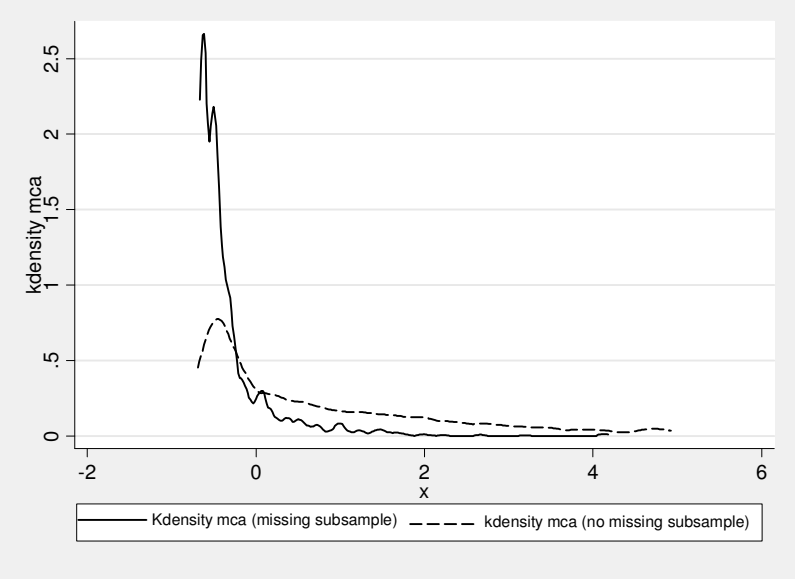
Table A 5 – Descriptive Statistics for the Poverty and Wealth Indices

Percentiles	**Poverty Index**	**Wealth Index**	**Transformed Wealth Index**
1%	*-4.54325*	*-2.76166*	*0.00001*
5%	*-2.82009*	*-2.64228*	*0.11939*
10%	*-1.68232*	*-2.24213*	*0.51954*
25%	*-0.42659*	*-0.63777*	*2.12390*
50% (Median)	*0.40023*	*-0.40023*	*2.36144*
75%	*0.63777*	*0.42659*	*3.18826*
90%	*2.24213*	*1.68232*	*4.44399*
95%	*2.64228*	*2.82009*	*5.58176*
99%	*2.76166*	*4.54325*	*7.30492*
Smallest	*-13.19478*	*-2.76166*	*0.00001*
Largest	*2.76166*	*13.19478*	*15.95645*
Mean	*0.17303*	*-0.17303*	*2.58864*
Standard Deviation	*1.543394*	*1.543394*	*1.54339*
Variance	*2.382066*	*2.382066*	*2.38207*
Skewness	*-1.13398*	*1.13398*	*1.13398*
Kurtosis	*7.211688*	*7.211688*	*7.21169*
Observations	*8700*	*8700*	*8700*

Appendix C – Ensuring the Quality of Data

The cumulative density functions (CDFs) for the wealth index have a small bump at the lower end of the distribution. This reflects the fact that there were about 21.03% missing observations on household ownership of durable private assets. Since these same households had no missing information on the remaining variables and to avoid having a much smaller data set, these households were included in the calculation of the wealth index.

Figure A 1 – Kernel density curves for the missing and no missing subsamples



The above paragraph means that Stata11 (Statacorp, 2009) placed the households with missing information at the bottom end of the wealth index, thus assuming that they did not possess those assets. This might suggest that the index is further biased. However, Figure A 1 suggests differently. This figure illustrates two Kernel density curves drawn making use of a wealth index that contains only household housing characteristics[52]. According to it, those households with missing data on household durable private assets were already concentrated at the bottom end of the distribution. In addition, the correlation between this reduced index and the extended one is 0.91 when all observations are considered and 0.90 when the missing observations are excluded.

[52] These include variables on the type of walls and roofs the houses are made of, the type of sanitation systems, the source of energy used in the houses, and on the type of access to water. The household private durable assets were excluded from this wealth index.

Therefore, the poverty and inequality analyses will not be biased by the manner in which the wealth index was computed.

Appendix D – Grade 6 Pupils in SACMEQ II and SACMEQ III

Table A 6 – Number of grade 6 pupils in Mozambique

Provinces	SACMEQ II (Year 2000)	SACMEQ III (Year 2007)	%-change
Niassa	5,317	13,944	162
Cabo Delgado	4,660	21,184	355
Nampula	14,514	40,692	180
Zambezia	13,722	46,396	238
Tete	7,269	23,561	224
Manica	7,986	23,786	198
Sofala	9,192	29,861	225
Inhambane	12,531	29,935	139
Gaza	12,063	27,059	124
Maputo Province	10,257	32,397	216
Maputo City	24,827	30,429	23
Total	**122,338**	**319,243**	**161**

Source: SACMEQ data; Note: totals were calculated using the raising factor provided in the data set

Appendix E – Summary Statistics of Education Data

Table A 7 – Summary statistics

Survey	SACMEQ II		SACMEQ III	
Variable	**Mean**	**SE**	**Mean**	**SE**
Pupil SES	0.09	0.05	-0.02	0.05
Age	14.50	0.06	13.95	0.07
Girl	0.41	0.01	0.46	0.01
Gets help with homework sometimes or often	0.81	0.01	0.83	0.01
Reading extra tuition	0.28	0.02	0.11	0.01
Numeracy extra tuition	0.42	0.02	0.12	0.01
Pupil repeated grades				
Never	0.22	0.01	0.40	0.01
Once	0.36	0.01	0.35	0.01
Twice	0.33	0.01	0.19	0.01
Three or more times	0.09	0.01	0.05	0.01
Portuguese spoken at home				
Never	0.06	0.01	0.08	0.01
Sometimes	0.41	0.02	0.42	0.01
Most of the time	0.53	0.02	0.50	0.02
Has supper every day	0.85	0.01	0.93	0.01
Number of books at home				
No books	0.20	0.02	0.31	0.02
1 to 10	0.57	0.02	0.53	0.02
11 or more	0.24	0.02	0.16	0.01
Number of days absent from school	2.76	0.12	1.08	0.07
School head has degree and management training	0.06	0.02	0.08	0.02
School SES	0.09	0.05	-0.02	0.05
School type				
Government	0.98	0.01	0.97	0.01
Private	0.02	0.01	0.03	0.01
Urban schools	0.74	0.04	0.63	0.04
Provinces				
Niassa	0.04	0.01	0.04	0.01
Cabo Delgado	0.04	0.01	0.07	0.01
Nampula	0.13	0.02	0.10	0.02
Zambezia	0.12	0.02	0.15	0.02
Tete	0.07	0.01	0.08	0.01
Manica	0.06	0.01	0.08	0.01
Sofala	0.08	0.02	0.10	0.02
Inhambane	0.10	0.02	0.10	0.02
Gaza	0.10	0.02	0.09	0.02
Maputo Province	0.09	0.02	0.10	0.02
Maputo City	0.18	0.03	0.10	0.02
Reading teacher hours of preparation per week	6.55	0.56	10.15	0.44
Maths teacher hours of preparation per week	6.37	0.42	9.98	0.47
Reading teacher is female	0.28	0.03	0.39	0.03
Maths teacher is female	0.26	0.03	0.30	0.03
Reading school report included comments	0.38	0.04	0.58	0.03
Maths school report included comments	0.40	0.04	0.60	0.03
Class size	53.27	0.84	59.00	0.95
Reading textbook				
No textbook	0.02	0.00	0.04	0.00
Only teacher has	0.05	0.01	0.04	0.01
Shares with 2+	0.21	0.01	0.25	0.01
Shares with 1	0.21	0.01	0.14	0.01
Uses by himself	0.52	0.02	0.52	0.02
Maths textbook				
No textbook	0.02	0.01	0.04	0.01

Only teacher has	0.04	0.01		0.05	0.01
Shares with 2+	0.19	0.01		0.25	0.01
Shares with 1	0.18	0.01		0.14	0.01
Uses by himself	0.57	0.02		0.51	0.02

Appendix F – Oaxaca-Blinder Detailed Decompositions

Table A 8 – Detailed decomposition of reading scores over time

	X_{2000}	X_{2007}	β_{2000}	β_{2007}	**Endowments effect** $[E(X_{2007}) - E(X_{2000})]'\beta_{2000}$	**Coefficients effects** $E(X_{2000})'(\beta_{2007} - \beta_{2000})$	**Interaction effect** $[E(X_{2007}) - E(X_{2000})]'(\beta_{2007} - \beta_{2000})$
Family background					**0.94**	**1.62**	**-1.23**
Pupil SES	0.09	-0.02	5.44	8.65	-0.62	0.29	-0.37
1-10 books	0.57	0.53	2.67	4.03	-0.10	0.77	-0.05
11+ books	0.24	0.16	1.74	10.36	-0.14	2.03	-0.68
Has supper every day	0.85	0.93	24.37	22.64	1.79	-1.47	-0.13
Pupil level					**5.02**	**33.72**	**1.74**
Gets help with school work	0.81	0.83	-1.57	12.68	-0.03	11.52	0.32
Age	14.50	13.95	-4.40	-1.56	2.44	41.12	-1.57
Pupils is a girl	0.41	0.46	-10.91	-12.46	-0.60	-0.63	-0.09
Takes extra tuition	0.28	0.11	-18.91	-7.00	3.19	3.35	-2.01
Repeated grade once	0.36	0.35	-4.36	-5.02	0.03	-0.24	0.01
Repeated grade twice	0.33	0.19	0.12	-12.68	-0.02	-4.23	1.81
Repeated grade 3+	0.09	0.05	-0.56	-28.20	0.02	-2.43	0.95
Speaks Portuguese sometimes	0.41	0.42	33.58	21.39	0.38	-5.03	-0.14
Speaks Portuguese often	0.53	0.50	35.33	23.48	-1.07	-6.27	0.36
Days absent from school	2.76	1.08	-0.41	-1.66	0.69	-3.45	2.10
School level					**-4.36**	**-39.00**	**-0.63**
Head training and degree	0.06	0.08	19.40	5.39	0.29	-0.84	-0.21
School SES	0.09	-0.02	4.59	14.17	-0.53	0.87	-1.10
Private school	0.02	0.03	18.60	-11.49	0.16	-0.62	-0.27
Urban	0.74	0.63	1.27	1.15	-0.14	-0.09	0.01
Cabo Delgado	0.04	0.07	13.05	8.24	0.38	-0.19	-0.14
Nampula	0.13	0.10	76.17	5.08	-2.20	-9.36	2.06
Zambezia	0.12	0.15	58.66	20.61	1.64	-4.62	-1.06
Tete	0.07	0.08	33.41	-16.13	0.28	-3.35	-0.41
Manica	0.06	0.08	44.51	0.14	0.93	-2.47	-0.93
Sofala	0.08	0.10	51.32	-7.81	0.90	-4.65	-1.04
Inhambane	0.10	0.10	52.35	28.35	0.01	-2.31	-0.01
Gaza	0.10	0.09	47.47	18.15	-0.38	-2.79	0.23

Continues next page

Continued from previous page

Maputo Province	0.09	0.10	54.08	23.65	0.62	-2.83	-0.35
Maputo City	0.18	0.10	79.42	47.05	-6.33	-5.75	2.58
Classroom level					**-3.41**	**34.91**	**3.73**
Teacher is female	0.28	0.39	-8.42	-4.24	-0.98	1.16	0.49
Teacher preparation	6.55	10.15	0.47	2.03	1.69	10.25	5.64
Teacher preparation^2	93.76	153.12	-0.01	-0.08	-0.87	-6.11	-3.87
Class size	53.27	59.00	-0.49	-0.28	-2.83	11.52	1.24
Only teacher has textbook	0.05	0.04	-10.07	5.56	0.06	0.77	-0.09
Shares textbook with 2+ pupils	0.21	0.25	9.88	28.27	0.46	3.79	0.85
Shares textbook with 1 pupil	0.21	0.14	16.01	31.71	-1.08	3.29	-1.06
Uses textbook by himself	0.52	0.52	15.07	33.29	0.09	9.41	0.10
Pupil school report	0.38	0.58	0.32	2.50	0.06	0.84	0.43
Constant	**1.00**	**1.00**	**493.29**	**421.69**	**--**	**-71.60**	**--**

Notes: numbers may not add to totals due to rounding

Table A 9 – Detailed decomposition of numeracy scores over time

	X_{2000}	X_{2007}	β_{2000}	β_{2007}	**Endowments effect** $[E(X_{2007}) - E(X_{2000})]'\beta_{2000}$	**Coefficients effects** $E(X_{2000})'(\beta_{2007} - \beta_{2000})$	**Interaction effect** $[E(X_{2007}) - E(X_{2000})]'(\beta_{2007} - \beta_{2000})$
Family background					**1.35**	**-4.79**	**-1.72**
Pupil SES	0.08	-0.02	-1.15	3.65	0.12	0.38	-0.49
1-10 books	0.57	0.53	4.63	6.83	-0.18	1.26	-0.09
11+ books	0.23	0.16	1.80	7.85	-0.14	1.41	-0.46
Has supper every day	0.85	0.93	20.68	11.48	1.55	-7.83	-0.69
Pupil level					**2.70**	**45.85**	**0.13**
Gets help with school work	0.81	0.83	-3.37	11.24	-0.08	11.80	0.34
Age	14.51	13.95	-3.05	- 0.00	1.72	44.26	-1.72
Pupils is a girl	0.40	0.46	-20.88	-12.83	-1.23	3.24	0.47
Takes extra tuition	0.42	0.12	-2.70	2.33	0.82	2.12	-1.53
Repeated grade once	0.36	0.35	-10.17	-5.95	0.07	1.53	-0.03
Repeated grade twice	0.33	0.19	-7.84	-10.45	1.10	-0.86	0.37
Repeated grade 3+	0.09	0.05	-3.35	-17.12	0.12	-1.22	0.48
Speaks Portuguese sometimes	0.41	0.42	27.93	12.91	0.26	-6.22	-0.14
Speaks Portuguese often	0.53	0.50	26.96	15.13	-0.79	-6.26	0.35
Days absent from school	2.76	1.08	-0.42	-1.34	0.71	-2.52	1.53
School level					**-3.42**	**-22.50**	**1.74**
Head training and degree	0.06	0.08	12.59	8.08	0.22	-0.26	-0.08
School SES	0.08	-0.02	5.89	8.81	-0.63	0.24	-0.31
Private school	0.02	0.03	2.81	-7.03	0.03	-0.20	-0.09
Urban	0.74	0.63	1.26	-5.07	-0.14	-4.67	0.69
Cabo Delgado	0.04	0.07	14.18	9.93	0.40	-0.17	-0.12
Nampula	0.13	0.10	50.39	16.15	-1.42	-4.50	0.96
Zambezia	0.12	0.15	32.80	26.43	0.85	-0.78	-0.17
Tete	0.07	0.08	24.93	4.98	0.21	-1.35	-0.16
Manica	0.06	0.08	45.72	17.06	0.89	-1.62	-0.56
Sofala	0.08	0.09	30.88	11.61	0.49	-1.53	-0.30
Inhambane	0.10	0.10	57.78	35.58	-0.02	-2.15	0.01
Gaza	0.10	0.09	37.69	34.54	-0.34	-0.30	0.03
Maputo Province	0.09	0.10	42.82	35.20	0.57	-0.69	-0.10
Maputo City	0.17	0.10	60.62	34.61	-4.51	-4.50	1.94
Classroom level					**-1.13**	**13.60**	**1.06**
Teacher is female	0.25	0.29	-1.48	8.58	-0.06	2.56	0.39
Teacher preparation	6.38	10.00	0.48	2.04	1.73	9.94	5.65

Continues next page

Continued from previous page

Teacher preparation[2]	81.69	151.21	-0.02	-0.06	-1.17	-3.43	-2.92
Class size	53.22	58.97	-0.37	-0.51	-2.13	-7.24	-0.78
Only teacher has textbook	0.04	0.05	-5.75	-3.01	-0.09	0.10	0.04
Shares textbook with 2+ pupils	0.19	0.25	9.53	20.57	0.56	2.11	0.64
Shares textbook with 1 pupil	0.18	0.14	7.85	20.47	-0.29	2.28	-0.47
Uses textbook by himself	0.57	0.51	6.48	21.46	-0.37	8.53	-0.86
Pupil school report	0.40	0.60	3.51	0.33	0.71	-1.25	-0.64
Constant	**1.00**	**1.00**	**513.94**	**435.60**	**--**	**-78.34**	**--**

Notes: numbers may not add to totals due to rounding