Education and Labor Market Outcomes in South Africa: Evidence from the National Income Dynamics Study

By

ESTHER MUMBI KIMANI

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the

School of Economics

UNIVERSITY OF CAPE TOWN

August 2015

Supervisor: Professor Haroon Bhorat

Copyright

Abstract

Existing literature is in agreement on the importance of education in the determination of labor market outcomes. Using data from South Africa's National Income Dynamics Study, this thesis explores this relationship. It does this firstly, by examining the effect of school quality measured by school inputs–pupil-teacher ratio and expenditure per pupil–on educational attainment in South Africa. Following a reduced form production function approach, a partial generalized ordered probit is applied in the analysis. The thesis finds that both pupil-teacher ratio and expenditure per pupil have strong and significant effects on educational attainment of African South Africans. The ratio is more important at lower schooling levels, indicating a 'lagged effect' on educational attainment. A small increase in expenditure has large effects. This suggests declining returns to fiscal investment in education. Alternatively, it could indicate inefficiency in the use of funds, or compensatory funding of poor schools. Secondly, we examine the extent to which wage differences shown among district councils in South Africa can be explained by the magnitude of external returns to education. We use an augmented Mincerian regression to investigate this, considering the effect of district council share of college graduates on workers' wages. The study employs District Council's annual average climate to instrument for the share of district council college graduates. The results show that a 1% increase in a District Council's share of college graduates raises workers' wages by 5-8%. There are also spillovers effects, with college graduates being the beneficiaries. Thirdly, we estimate unemployment duration by gender and by competing risk, that is, exits into employment or economic inactivity. We use the Kaplan-Meier estimator and the Cox proportion hazard model in these estimations. The results from both estimators show factors that influence unemployment differ by exits, and their effects vary by gender. The hazard rates show that transition rate into employment is higher for men than for women with similar characteristics. They show that age and race significantly influence employers' choice between educated men and women. However, this bias is less obvious at higher levels of education.

Contents

List of Figures

List of Tables

Dedication

To my parents Lucia Muthoni and Benjamin Kimani for their commitment to giving me a good education from the start, and their unfailing support through the many years of schooling.

To the memory of uncle Samson Muiruri who opened the doors of university for me, but unfortunately never lived long enough to see me get this far. His words on being a woman of substance still echo in my mind.

Acknowledgements

Completion of this doctoral dissertation is the result of support by dozens of people, who I wish to thank. Firstly, I would like to thank my supervisor Professor Haroon Bhorat for his guidance, understanding, patience, and most importantly his mentorship in research. He has supported me not only by providing a research assistantship over almost six years, but also academically and emotionally through the rough road to finish this thesis. His flexibility in scheduling, gentle encouragement and relaxed demeanor made for a good working relationship and the impetus for me to finish. For this, I cannot thank him enough. I am forever grateful.

At UCT, I have encountered remarkable individuals who I also wish to acknowledge. A big thank you to Dr Reza Daniels, Professor Vimal Ranchhod and Dr Miquel Pellicer for their invaluable insights and time in discussing my work. A special thanks to Data First staff, and in particular Lynn Woolfrey and Alison Siljeur who ensured I had the best service in using the Secure Research Data Centre. To the National Income Dynamics Study team, particularly Michelle Chinhema and Tim Brophy thank you for keeping your doors open for questions and discussions on the NIDS data. Thank you Sarah Marriott and Lynn Woolfrey for the editing. You made all the difference in the text! To Lisl George and Waseema Peterson thank you for your friendship and the warmth you extended in my attachment to DPRU. You made me feel I was part of the DPRU family. I am also grateful to Paula Bassingthwaighte and Yasmina Charles for their support from day one at the School of Economics.

Without consistent funding this thesis would not have been possible. I am thankful to the National Research Foundation: South African Research Chairs Initiative (SARChI) Programme, the Carnegie Corporation, the Development Policy Research Unit (DPRU), and the UCT postgraduate funding office for their generous funding.

My gratitude is also extended to Dr John Okidi whose generosity made it possible to start this journey. John I can never repay you for your unconditional support. I am forever grateful. I am also thankful to Dr Walter Odhiambo who inspired me to pursue economics from my undergraduate years at the University of Nairobi. I am also indebted to Dr Moses Muriithi Kinyanjui and Dr Kamau Gathiaka who saw potential in me, and kept pushing and pestering me to pursue postgraduate studies. You two are my champions! A special thank you to Professor Gemano Mwabu, Professor Damiano Manda Kulundu, and Professor Edwin Muchapondwa for their support over the years.

To my friends Fenah Wanjiru Njuguna, Margaret Maina, Wairimu Mwangi, Wangari Wangombe, Coretha Komba and Simon Wagura thank you for your unwavering friendship over the years. I am lucky to have you for keeps! I am grateful to my friends and colleagues Naomi, Munene, Adeola, David, Otieno, and others at UCT School of Economics who we worked closely and puzzled over many of the same problems.

I am deeply indebted to my family. I am grateful for the sacrifices made by my parents Lucia Muthoni and Benjamin Kimani to give me a good education. Despite all the challenges they faced, they still managed to give me an education. It is the best gift they have given me, and for this I am humbled and forever grateful. To my siblings Sammy Thuo, Evans Kagwima, John Ngugi, and Grace Wangari you all inspired me to aspire higher, thank you! Thank you Grace for being the good listener you are. The many chats and calls always transported me home.

Finally, I am grateful to my best friend, partner, and husband Mare Sarr who lived every minute of this PhD, and without whom I would have probably given up. He constantly propped me up when my morale failed me, and believed in my intellect constantly encouraging me to keep going. Mare has been my greatest supporter, and has unconditionally loved me during my good and bad times. Thank you Mare, you are real gem, and as I always say what could I do without you!

Declaration

I, Esther Mumbi Kimani, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signed by candidate

Signature:..... ..... Date: ...11^{th} August 2015.....

Chapter 1

Introduction

1.1 Overview of the thesis

There is agreement in the economic literature and in the policy arena on the importance of education in determining labor market outcomes. In South Africa apartheid social engineering devised an unequal education system that continues to shape the country's education landscape. Racially segregated schools and Bantu Education created an education system that was biased towards White schools, with a highly skewed distribution of educational resources and opportunities reflected in performance between race groups. By 1994, there was an entrenched bias in allocation of resource such as expenditure per pupil, number of qualified teachers per school, and facilities, among others. White schools were prioritised at the expense of other population groups, that is Africans, Coloureds, and Indians/Asians (commonly referred to as Black). The pupil-teacher ratio averaged at less than 30 pupils per teacher in White schools, while in Black schools it was on average between 40 to 70 pupils per teacher (Case and Deaton, 1999). The per capita expenditure on White pupils was at least 5 times that on African pupils (Moulder, 1992, as cited in Bhorat and Oosthuizen, 2008, p2).

Since the advent of democracy, successive South African governments have attempted to address these issues by implementing education policies that depart radically from the past. Over the past 20 years, there has been much focus on education policies, although with disappointing outcomes (Spreen and Vally, 2006). Quality education is, as one would expect, regarded as key to resolving many of South Africa's economic and social woes such as poverty, inequality, high unemployment and slow growth. An increase in the level of schooling is believed to raise the future earning potential of individuals, together with a rise in productivity levels. This productivity, as reflected in earnings, makes education a major determinant of upward mobility and an important component in the reduction of income inequality. In addition to earning higher wages, individuals with higher levels of schooling are less likely to be unemployed, and have shorter unemployment spells. Aside from education, other factors such as gender, race, age, and non-labor incomes are also important determinants of unemployment duration. In developing countries, women are less

likely to participate in the labour market, and those who participate are less likely to find employment relative to men, often irrespective of their education level. This thesis explores the nexus between education and labour market outcomes in democratic South Africa using data from the National Income Dynamics Survey (NIDS) for 2008-2012. It does so first, by investigating the effects of school quality and resource allocation on educational attainment. Second, it estimates private and social returns to higher education. Finally, it estimates unemployment duration by exit states and gender in South Africa.

Chapter 2 of this thesis is concerned with the effect of the provision of educational resources on educational attainment. This question is of primary importance because resource allocation and quality of education may be closely related. Concerns about the quality of education provided in South Africa, particularly at the general level, have been noted in the literature. These concerns result from the high grade repetition and dropout rate, and continued poor performance in international tests (Murtin, 2013; Van der Berg et al., 2011b). Evidence suggests that 80% of students attend schools of low quality (Van der Berg and Moses, 2011), and that the majority of these students are Africans who mainly attend formerly disadvantaged schools that have remained poorly resourced. To deal with the issue of quality and inequality in provision of education, the South African government introduced a pro-poor school subsidy. This takes the form of school-fee exemption in fee paying schools (mainly formerly White schools) and no-fee schools (predominantly formerly disadvantaged schools). Given that Africans attend schools at the lower strata of quality, and they are the main beneficiary of this subsidy, it is of research and policy interest to investigate whether the subsidy has had any effect on the level of education attained for Africans in South Africa. The subsidy is allocated as an expenditure per pupil, which is a non-capital and non-personnel allocation.This has resulted in a substantial increase in government expenditure in education. Anecdotal evidence suggests that the department of education has used these funds inefficiently (Taylor et al. 2008). Statistical evidence either supporting or refuting these sentiments is largely lacking. Chapter 2 of this thesis tries to fill this gap. Further, the South African government has increased the number of teachers it employs with the objective of equalising the pupil-teacher ratio in all publicly funded schools. How has the reduction in the ratio affected education attainment of Africans?

In chapter 2, the thesis thus explores the supply and demand factors contributing to the production of education. The specific focus is on school inputs, that is, expenditure per pupil and the pupil-teacher ratio. It draws on the theoretical and empirical literature on education production function (Glewwe and Kremer, 2006; Mani et al., 2009; Behrman, 2010; Aturupane et al., 2013). Rather than relying on the structural education production function, which would estimate the welfare effects of school inputs, the thesis employs a reduced form production function approach. There are two reasons behind this choice. First, the estimates from the reduced form give the direct effect of school inputs on educational attainment. Second, the data requirements for the structural function cannot be met by the data available.

Focusing on the sub-sample of African South Africans in the National Income Dynamics Study (NIDS) Wave 1 data collected in 2008, the study estimates the effects of expenditure per pupil and pupil-teacher ratio. Education level attained is considered as an ordered event. The study takes into account the possibility that the effects of these inputs could vary at different levels of education. A partial generalised ordered probit model is employed in the empirical analysis. Marginal probability effects of the pupil-teacher ratio and expenditure per pupil at representative values conditional on the mean of the controls are estimated. In addition, individual, socio-economic, and neighborhood characteristics are controlled for.

The thesis finds that both the pupil-teacher ratio and expenditure per pupil have strong and significant effects on educational attainment of Africans in South Africa. The results show the pupil-teacher ratio is more important at lower levels of education. This suggests the ratio has a 'lagged effect' on schooling and higher education attainment. The results show that higher expenditure per pupil increases the probability of attaining secondary education. However, the effect declines with increase in expenditure. This is an indication that either there are declining returns to fiscal investment in education (a 'thinning effect' of the impact) or there is an inefficiency in the use of allocated funds. The thesis also argues that these results may be an indication of compensatory funding of poor schools. This posits that the positive effects on poor schools are likely to be realised once adequate resources are available, hence the need for continued state funding. However the results show that the secondary-higher education nexus is also driven by other factors.These results are confirmed in the educational attainment index analysis, which accounts for possible bias arising from grade repetition and dropout . Overall, these findings imply that government funding plays an important role in improving educational attainment for Africans in South Africa, and that poor schools still need these funds. The findings show though, that inefficiency in the use of the funds is likely, and calls for appropriate policy measures.

State funding of primary, and in some cases secondary education, has been largely accepted as a norm. However, there has been opposition to increased and continued funding of education, particularly in higher education. This opposition is often based on the view that these funds are not used efficiently, and that their benefits accrue primarily to the individual rather than to society. In chapter 2 the thesis argues that although state funds allocated to primary and secondary education are instrumental in increasing educational attainment, the possibility that they are not being used efficiently cannot be ruled out. Is continued state funding education therefore justifiable? Or are there sufficient benefits to society from such funding which could warrant the use of tax payers' money to pay for individuals' tertiary education in spite of the efficiency concerns? Further, the question of whether there should be full public funding in higher education is also significant for researchers and policy makers alike. Answers to these questions are instrumental in justifying public funding in education. This makes the estimation of private and social/external returns to education necessary. These questions will be investigated in chapter 3. The literature on returns to education in South Africa shows there are positive and convex private returns to education (Keswell

and Poswell, 2004 and Burger and Jafta, 2006). These positive returns increase with the level of education (Mwabu and Schultz, 2000 and Burger and Jafta, 2006). However, the literature is largely silent on the magnitude of external returns and spillover effects of education. The only exception is a study by Michaud and Vencatachellum (2003) which suggests that there are positive external returns. The research of Schultz (2004) argues social returns are smaller than private returns.

Both the theoretical and the empirical literature agree that there are both private and social benefits to education. There is little contention around the private benefits of education, which are estimated to average 10% for each additional year of schooling (Psacharopoulos and Patrinos, 2004). There is, however, a debate on the magnitude of social returns (also referred to as external returns) compared to private returns. These reservations are mainly driven by the lack of direct measures of external returns. Therefore identification of external returns has been at the heart of the research agenda on this debate. Different approaches have been employed to estimate these returns. Findings on average yield different estimates, further fueling a debate on the magnitude of external returns. Estimates from existing literature for different countries show negative returns in Egypt (Nazier, 2013), zero in the US (Acemoglu and Angrist, 2001), and 16% in China (Liu, 2007). The Mincerian regression augmented with a measure of aggregate human capital is the most common approach. This measure identifies external returns by including in a Mincerian regression the average years of schooling. Another approach is to include in this regression the share of graduates with a given level of education in a locality. This latter estimation strategy is used in chapter 3 to estimate the external returns and spillover effects of higher education in South Africa.

Employing an augmented Mincerian regression, the thesis examine the extent to which wage differences in district councils can be explained by the magnitude of external returns to education in South Africa. It does this by considering the effect of a District Council's share of college graduates on workers' wages. Using the unbalanced panel of working adults (aged 15-64) from NIDS waves 1-3, least square, fixed effects and instrumental variable estimates are considered. Employing annual average climate to instrument for the share of college graduates within a District Council, the results show that a 1% increase in a District Council's share of college graduates raises workers' wages by 5-8%. This positive effect is in line with estimates from other countries, albeit significantly larger. This is also in line with expectations, given that South Africa has a high skills premium. The findings also show that there are spillovers, with college graduates being the beneficiaries. A 1% increase in the share of college graduates raises wages of workers with at least some college education by almost 2 percentage points more than the wages of workers with less education than a college diploma. This could be a consequence of the high skills deficit which leads to the supply of workers with higher skills having no negative effect on wages of highly skilled workers. It could also be a consequence of unionisation, resulting in low skill wages not responding to market changes. It could also reflect both these circumstances.

Despite this high skills deficit in South Africa, unemployment levels have been persistently high, and unemployment duration significantly long. Both the level and duration of unemployment

vary by education level attained, gender, race, and age, among other factors. Chapter 4 of the thesis examines factors that influence transition out of employment. In studying unemployment duration, the common approach in literature has been to analyse the transition rate from unemployment into employment. This approach neglects the fact that there are multiple exits that individuals could take. These include part-time employment, engaging in a self-employment activity, or opting out of the labour market. An increasingly popular approach is the analysis of unemployment duration by competing risks. The competing risks approach is instrumental in characterising the unemployed, making it a more appropriate policy tool for guiding policy makers in formulating better targeted policies to deal with unemployment. Exit into either employment or economic inactivity are the most likely in South Africa. This is because most unemployed individuals in South Africa are either actively searching for a job or are discouraged. Empirical evidence suggests that most South Africans are discouraged before they even enter the labour market (Wittenberg, 2002; Kingdon and Knight, 2006). This shows that non participation is an option from the onset, and that unemployment duration varies. This thesis proposes that there are variations in either exit option by gender in South Africa. This is based first, on evidence from other countries which shows different factors influence transition for men and women. Second, evidence seems to show that employers have a bias towards selecting male employees when presented with candidates with equal qualifications and experience. As a remedy for this bias, South Africa's Employment Equity Act No. 55 of 1998 requires employers to prioritise formerly disadvantaged groups, women, and people with disability in their hiring. Third, the transition rate for men and women into either employment or economic inactivity has been argued to be based on traditional gender roles. These roles define men as providers making them more likely to exit unemployment into employment, and women as home makers, making them more likely to exit the labour market (Ollikainen, 2003; Gonzalo and Saarela, 2000). Chapter 4 of this thesis contributes to closing the existing gap in the literature on factors that influence transition from unemployment into either employment or economic inactivity for men and women in South Africa.

A balanced panel from NIDS waves 1-3 is used to study the transition from unemployment into employment or economic inactivity for men and women. The panel is used to calculate the Kaplan-Meier estimator and the Cox proportional hazard model by gender. The findings indicate that men have a higher probability of exiting into employment, while women have a higher probability of exiting into economic inactivity. Results from the Kaplan-Meier estimator by education level attained, race, and age also show unemployment duration varies by exit states, but are less clear-cut. The results from the Cox proportion hazard model indicate that education level attained, age, relationship with household head, and being a recipient of a government social grant, significantly influence transition into both exit options, and their effects vary by gender. The importance of age declines as age increases, and race seems to mainly affect transition into either exit for African and Coloured women only. By contrast, the hazard rate into employment is higher for educated women than for educated men, an indication of higher returns to education. This also reflects and presence of employment equity which places a premium on hiring women . The hazard rates of

an African aged 26-35 with a tertiary education show that men have a higher probability of exiting into employment than women with comparable levels of education. This makes gender significant for this group. However, this bias against women at higher levels of education is less obvious when the average of each characteristics controlled for is considered.

These results confirm that education significantly influences labour market outcomes. This thesis finds that school quality, as measured by the school inputs of pupil-teacher ratio and expenditure per pupil, significantly affect the education level attained. However, the effects of expenditure per pupil indicate declining returns to fiscal expenditure and could be affected by inefficiency. However this funding is still necessary for poor schools which still lag behind with respect to resource allocation. Further, analysis shows that there are substantial social returns to education over and above the private returns to education. There are also large spillover effects from higher education to both less skilled and highly skilled individuals. These spillovers are, however, greater for highly skilled individuals, which is a likely result of the existing skills premium. Finally, the thesis finds that transition out of unemployment into either employment or economic inactivity varies by gender. Men are favored in the labour market in that they are more likely to transit into employment while women are more likely to exit the labour market. This bias is, however, less obvious at higher levels of education.

The rest of the thesis is organised as follows: Chapter 2 analyses the determinants of educational attainment for Africans in South Africa with a specific interest on the effects of expenditure per pupil and pupil-teacher ratio. Chapter 3 investigates the existence of social returns to education over and above private returns. It does so by estimating the social returns of higher education at the District Council level as well as the spillover effects (externalities) from higher education. Chapter 4 investigates unemployment duration by exit states. It specifically studies the transition from unemployment into either employment or economic inactivity, and finds that the rate of transition differs markedly for men and women. A summary of the findings is presented and a conclusion from the three chapters is drawn in chapter 5.

Chapter 2

The Effects of Expenditure per Pupil and Pupil-Teacher Ratio on Educational Outcomes in South Africa

2.1 Introduction

According to UN Millennium Development Goal (MDG) 2 (United Nation 2010, p15) quality education and increasing educational attainment are vital in achieving universal primary education by 2015. There have been substantial gains in achieving universal education, with net enrollment reaching 90% in developing countries, and 77% in Sub-Saharan Africa in 2011, a 7 percentage point increase from 2000 (United Nation 2013, p15). Sadly, this increase in enrollment has not been accompanied by an increase in school quality, which is commonly measured by a variety of different school inputs. Studying the role of education in economic growth in developing countries, Hanushek (2013) finds most students in developing countries who have completed nine years of schooling have skills below international norms and thus cannot compete at a global level with their peers. Provision of a quality education requires both infrastructure and access, and for the type of education offered to be of value it should, at least, impart some basic skills (Hanushek, 2013). Given that public provision of basic education has mainly been a prerogative of governments a continued increase in public expenditure on education is largely inevitable.

In South Africa, the apartheid era government was discriminatory in its provision of educational services and resources. It prioritized White schools at the expense of Black[1] schools. This meant that there were fewer and less qualified teachers, and lower (if any) expenditure allocated to schools meant for Blacks, who made up the majority of the population. At the end of apartheid the teacher-pupil ratio[2] stood at between 1:20 and 1:30 for Whites, and between 1:40 and 1:70 for African

[1] Black in this case refers to Africans, Indians/Asians, and Coloured.

[2] The pupil-teacher ratio indicates the number of educator provided in a school; it includes staff that might not be teaching such as the school management staff. The class size is the average number of learners (learner and pupil are

pupils (Case and Deaton, 1999). Equally, in that era, there was racial divergence in allocation of expenditure per pupil. For example, for every R1 per capita expenditure on a White pupil, only 19 cents was spent on an African pupil (Moulder, 1992, as cited in Bhorat and Oosthuizen, 2008, p2). Resources allocated for Indians and Coloureds fell between those for these two population groups. This divergence in resource allocation had implications for the quality of education offered in the different schools. White schools had the highest quality, while African schools had the lowest quality. To date, these formerly White schools (commonly known as 'Model C' schools) are better resourced and have had the best educational outcomes. The high matriculation[3] pass rate in these schools is a case in point. Although resource provision has changed a great deal, with measures to redress past equity issues, inequality in school quality still persists (Van der Berg, 2007 and Yamauchi, 2011). This persistence has been in spite of efforts by the various education stakeholders to transform the education system.

In accordance with the Constitution and the School Act, South Africans are free to apply for admission to any school regardless of their race or religion. Heterogeneity in student population by race is, however, yet to be achieved in most schools. The population structure in schools is still homogeneous with 96% of African pupils still attending formerly "Black" schools (Van der Berg, 2007). This lack of racial mixing can be associated with two main factors. First, school admission policy gives priority to students within a 5km radius of the school. This spatial rule has meant that, in a racially homogeneous community, still a legacy of aparthied's separate development policies,schools have remained racially homogeneous. The apartheid practice of populations being clustered by race, and spatial boundaries have thus been critical barriers to entry in schools of choice, particularly entry into 'model C' schools. In some cases, particularly in private schools, former White schools have been found to inflate school fees in an effort to discourage Black students from entry (Selod and Zenou, 2003). Second, financial constraints have been partly to blame for this lack of racial mixing in schools. These have been in the form of labour market constraints, and high land and housing prices which make it impossible for low income households to move to particularly formerly White areas (Yamauchi, 2011). Therefore barriers to equal opportunities for entry into a school of choice for some South African students still currently exist, in spite of the efforts by education stakeholders to erode them.

The heterogeneous provision of and access to education in the apartheid era led to disparity in educational attainment by race and type (or level) of skill acquired by the different population groups. Van der Berg and Moses (2011) attribute persistent income inequality to this heterogeneity in education attainment, while Bhorat et al. (2009) find wage income to be the main driver of the income inequality in the country. Income inequality is highly skewed by skill and by race; on the one hand Bhorat et al. (2009) shows that in the decade between 1995 and 2005, wage inequality increased by 14%. They associate this increase with a skills premium. On the other hand, Gradin

used interchangeably throughout the text) in a class (DBE, 2010b).

[3]Matriculation examination is taken at the end of grade 12 and is a prerequisite for university admission.

(2014, p75) shows that in 2005/2006 the median income of Africans in South Africa was 9% that of Whites. The need to close the gap in skills has been a priority in the country. The aim is to move from a position where Whites are highly skilled and Africans mainly have low skills or no skills. The provision of a high quality education to all has been a necessary requirement. Education policies such as allocation of expenditure per pupil by school poverty quintile, and increase in employment of teachers to lower the pupil-teacher ratio in all schools, have been put in place by the government in an effort to improve school quality and offer equal educational opportunities to all South Africans. Provision of opportunity to access equal education to all is perceived to enhance social mobility (Council on Higher Education (CHE), 2004). This is seen as a necessary constituent in post-apartheid South Africa's overall transformation. All these factors make studies of school quality in South Africa imperative.

In recent years government efforts to bring down these barriers, and deal with inequality in provision of education have mainly been through subsidies (Republic of South Africa, 1998b). The School Act of 1996 provided for school-fee exemption for students from poor backgrounds attending fee paying schools (quintile 4 and 5 schools). The National Norms and Standards for School Funding Notice of 2006 provided for no-fees[4] schools . The objectives of these laws have been to try and ease the financial burden of education on poor families, and to promote equitable access to quality education for all. The government implements this through a school quintile[5] system that is used to determine allocation of expenditure per pupil to schools. All ordinary public schools are assigned to a quintile based on the poverty level of the surrounding community[6], and the physical condition, facilities, and population of the school [7](South African School Amendment Act, no 84 of 1996, as amended: 2006, p24-29). Schools in quintile 1 are the poorest schools and receive the highest per pupil fiscal allocation, and those in quintile 5 are the least poor schools and receive the lowest allocation per pupil. This allocation of expenditure per pupil ensures that more funds are allocated to pupils in the largest and neediest schools. The National Norms and Standards for School Funding Notice stipulates that these funds are for recurrent expenditure. This may comprise purchase of textbooks and educational material or equipment for the school; school buildings improvement and maintenance; extra-mural curricula; and provision of services to the school. The funds are not to be used for hiring teachers or personnel and/or for capital expenditure

[4]The no fee status of a school is determined by the school quintile, which is arrived at following the national poverty ranking. It initially applied to 40% of the poorest schools (bottom two quintiles); both primary and secondary schools. These are supposed to receive 60% of the available resources. In 2011, the no-fee policy was extended to 3rd quintile schools (South African School Act of 1996, Amended National Norms and Standards for School Funding Notice 33723, as amended: 2010, p2)

[5]The school quintile determines how much money is allocated per learner in a school. Schools in quintile 1 receive the highest while those in quintile 5 receive the least amount.

[6]In determining the school poverty score, the geographical area of the school and three indicators of poverty: income; dependency ratio (or unemployment rate); and level of education (or literacy rate) of the community, are weighted (South African School Amendment Act, no 84 of 1996, as amended: 2006, p29).

[7]The provincial education department is mandated "to create indices based on the range of physical facilities at the school, learner: classroom ratio, the overall condition and need for repairs, and availability of basic services" (South African School Act of 1996, National Norms and Standards for School Funding Notice 2362 of 1998, p28).

(South African School Act of 1996, National Norms and Standards for School Funding Notice 2362 of 1998, p30).

Assessment of the impact of these funds has been minimal. In a qualitative analysis of the effect of school fees and funding on school quality, Hall and Giese (2008) find that school fee exemption and no-fee schools have undoubtedly increased funding to poor schools. However, they conclude that this increase has not culminated in school quality improvement. According to Hall and Giese, these policies have failed to deal with unequal allocation of teaching capacity, since there is no pro-poor allocation of funding to teachers, an important element in improving school quality. Not many studies have been done on the effect of increased education funding in developing countries. However existing studies on the effect of expenditure per pupil on educational outcome in developing countries have conflicting findings. There is a dearth of studies examining the relationship between expenditure per pupil and educational attainment in South Africa, particularly for Africans. This study aims to fill this gap using a nationally representative data, the National Income Dynamic Study (NIDS). Further, given the existing debate on the effect of the increased funding on educational outcomes, there is a need for empirical evidence to justify this expenditure.

There is consensus among researchers that the quality of education in South Africa is low relative to what is spent on the provision of this service. This view is supported even within government (Department of Education, 2003, p107). Public discontent with the quality of education, particularly in public schools, has been on the increase. In addition to these quality concerns, inefficiency from the education department at the national level to the school level has been a concern (Taylor et al., 2008). Van der Berg and Moses (2011) find that there has hardly been any improvement in the quality of education for 80% of children in South Africa. According to Van der Berg (2007), the increase in resources to poor schools has not reduced the quality differentials among schools. These concerns cannot be overstated particularly with education viewed as an equaliser of income inequality through giving better job prospects, and as an engine for economic growth. In addition to the human capital needs of the economy, the argument for public provision of education is based on the need to distribute incomes, and build social cohesion. Investigating how government policies on schooling affect educational outcomes for students is not just essential to policy makers, but is an important research question in its own right.

Empirical research on the effects of school inputs (commonly referred to as school quality in literature) on educational attainment for South Africa is limited in the the range of outcomes considered. Existing research has mainly focused on test scores, as the outcome variable, and pupil-teacher ratio, as the quality measure, using survey data sets that are not nationally representative. The findings from these studies are conflicting. In a study of school quality in the apartheid era, Case and Deaton (1999) finds that pupil-teacher ratio has a marked effect on three educational outcomes for African children, namely enrollment, education attainment and test score. Two studies in post-apartheid South Africa find conflicting results despite the fact that they use similar data. Using 1999 and 2000 matriculation results Van der Berg (2007) finds that the teacher-pupil ratio

has a negative effect on the matriculation score, except in Black schools. Bhorat and Oosthuizen (2008), on the other hand, find that teacher-pupil ratio had no effect on the 2000 matriculation pass rate. This study extends the existing literature on effects of school inputs on educational attainment by examining the relationship between pupil-teacher ratio and expenditure per pupil, and highest education level attained for African South Africans.

In the literature, the most commonly used measures in educational outcome studies are class size, pupil-teacher ratio, and expenditure per learner. About a third of studies on effects of school quality on educational outcome consider some form of test score as the outcome variable (Wilson, 2001). In this study we consider the highest level of education attained as the outcome measure. Our school quality measures are expenditure per pupil, a non-capital and non-personnel allocation by the South African government, and pupil-teacher[8] ratio. These school quality measures are referred to as school inputs in this study. In addition to data availability, there is motivation for this choice. Firstly, pupil-teacher ratio is the most common quality measure in South African school quality literature, which offers an avenue for comparison. Secondly, there is hardly any consensus on the relationship between pupil-teacher ratio and educational outcomes in South Africa. Thirdly, this specific type of expenditure per pupil, to the best of our knowledge, is yet to be considered in an analysis of household survey data in South Africa. Lastly, most school quality studies in South Africa consider the test score as the outcome variable, while we consider educational attainment. Therefore, an analysis of the effect of school quality on education level attained would be an addition to the literature. The study's hypotheses are therefore that a reduction in pupil-teacher ratio leads to an improvement in classroom instruction quality, resulting from improved interaction between pupils and their teacher. This in turn leads to an increase in educational attainment. Further, an increase in expenditure per learner leads to an increase in the resources available to learners. This results in an improved learning experience, which in turn leads to an increase in educational attainment.

Our analysis focuses on data on educational attainment of Africans provided by Wave 1[9] of the National Income Dynamic Study (NIDS) which was collected in 2008. Using a partial generalised ordered probit we find that an increase in pupil-teacher ratio increases the probability of attainment of lower education levels. An increase in expenditure per pupil increases the probability of attaining more than a primary education, but lowers the probability of attaining a matriculation and higher levels of education. Our estimates indicate that at a pupil-teacher ratio of 21, chances of attaining a primary education are raised by 13.2 percentage points, but chances of attaining matriculation are lowered by 13.4 percentage points. Also, expenditure per pupil of R393 lowers chances of general education attainment by 7.39 percentage points, but raises chances of a matriculation by 6.9 percentage points. Our results show that expenditure per pupil has negative effects at general education, which suggests compensatory funding of poor schools. Also they show that an increase

[8]Teachers in the majority of schools are employed by the government.

[9]We do not consider the NIDS panel because the variation in the education variable is limited.

in expenditure per pupil does not lead to an increase in educational attainment, an indication of declining returns to fiscal investment in education ('thinning effect' of the impact) and a possibility of inefficiency in use of the funds. The results from the pupil-teacher ratio show the ratio is more important at lower levels of education than at higher levels, an indication of a lagged effect on schooling and higher education. Overall, the results indicate that the secondary-higher education nexus is driven by factors outside of the expenditure per pupil by state and pupil-teacher ratio. Our least square analysis supports the negative effect of the ratio and shows a positive effect from expenditure per pupil.

The remainder of the study is organized as follows: Section 2.2 reviews the relevant literature, section 2.3 outlines the approach and method of analysis, and offers a discussion on the data used. Descriptive statistics follow in section 2.4. This is followed by a discussion of empirical results regarding education level achieved and an education level index is provided. A conclusion is given in the last section.

2.2 A review of literature

2.2.1 Determinants of educational attainment

2.2.1.1 Socio-economic factors

Educational attainment is determined by various factors that can be categorized into demand side factors and supply side factors. The demand side factors vary within the three education stages, namely preschool, schooling, and post school (Behrman, 2010 p4887). These factors include parent's educational attainment, household income, household composition, and neighbourhood characteristics. Parent's education is an important determinant of educational outcome. However, its effects may vary by gender of either the parent or the child. These differences could arise from parent's decision on how much to spend on a child education. Parent's education has been shown to positively influence educational attainment and enrollment in Guinea (Glick and Sahn, 2000), in China (Zhao and Glewwe, 2010) and in Turkey (Tansel, 2002), and in South Africa (Gustafsson, 2007; Van der Berg and Moses, 2011). In South Africa, parent's education has a larger effect than socio-economic status (Gustafsson, 2007), and parent's higher education has an influence in higher reading and mathematics scores (Van der Berg and Moses, 2011). The effect of parent's education has been shown to sometimes differ by the parent's gender. Tansel (2002) finds mother's education increased school attainment, and Glewwe and Jacoby (1994) find mother's education significantly increased reading and mathematics scores, but a father's education had no effect. Further, parent's education has been shown to have mixed effect depending on a child's gender. For instance, Glick and Sahn (2000) find mother's education had no effect on boy's educational attainment, while Tansel (2002) finds parental education had a positive and larger effect on school attainment for girls than for boys.

Household spending on education influences children's schooling decision. The level of education attained, particularly post compulsory level (in cases where there is compulsory education) is influenced by a household's income in addition to other factors. In cases where education is free, household income still matters, particularly where households have to provide school inputs and learning materials such as school uniforms and books. Empirical evidence shows that household income has a positive effect. Glick and Sahn (2000) find permanent household income increases educational attainment and enrollment. Similar effects of household income are demonstrated by Zhao and Glewwe (2010), Tansel (2002), and Lee and Barro (2001). Family inputs have also been found to influence school performance Lee and Barro (2001), and family structure could be a factor. Glick and Sahn (2000) find that having siblings under the age of 5 significantly decreases girls' education attainment. However, Bhorat and Oosthuizen (2008) find the number of children in the household has no effect. In South Africa, using early 1990s data Case and Deaton (1999) find household resources have no effect on White learners' educational outcome, but were highly significant for Black learners' outcomes. In the apartheid era, schools for White's had better infrastructure than those designated for Blacks. This meant that the resources Black student's parents provided plays a vital role in these students' educational attainment. Case and Deaton (1999) conclude that high pupil-teacher ratios negatively affect parents' spending on education, which leads to low educational attainment, low test scores, and low probability of enrollment.

Infrastructure and facilities available to learners have mixed effects on educational outcome. In China, Zhao and Glewwe (2010) find that having a science laboratory at school increased schooling attainment, while bad infrastructure such as a leaking classroom decreased schooling attainment. In South Africa, school infrastructure and facilities differ particularly based on former apartheid classification. In a study using data collected at the end of apartheid, Case and Deaton (1999) find the existence of school libraries and laboratories had no effect on test scores at primary school. They found too that the existence of laboratories had no effect at secondary school, but a libraries in secondary schools had significant effects. The apartheid legacy has been shown to have some impact on educational outcomes. For example, Gustafsson (2007) finds historical influences negatively impacted on test scores. In an analysis of determinants of grade 12 pass rates by apartheid population grouping, Bhorat and Oosthuizen (2008) find mixed results for physical resources. They find having computers and a school library are important, while chalkboards, chairs and desks had insignificant influence on matriculation scores. The authors also find race, school type as per apartheid classification, the availability of a telephone, and electricity to be significant but access to water to be insignificant in influencing the scores.

Geographical location has an influence on the type of resources available to students. For example, well qualified teachers may choose to teach in urban areas. Learners in an urban area may not have have a library in their school, but can access a library in the urban centre while those in rural areas may not have that option. These differences resulting from the area of residence are likely to affect students' educational attainment. The effects of area of residence and neighborhood

effects have been studied. Tansel (2002) finds for Turkey urban location increases the probability of school attainment and living in an area with a high proportion of residents working in an industry increased primary education attainment, but it did not increase other levels of education. In South Africa, Van der Berg and Moses (2011) find learners in urban schools had higher test scores, and Bhorat and Oosthuizen (2008) find geographical location influences test scores. Yamauchi (2011) finds parents in areas with high income earners, higher average years of schooling and low unemployment pay higher school fees and hence these areas have higher quality schools. Both geographical area and the environment in which an individuals grows up have some influence on educational attainment. Therefore, resources available to students, influenced by where they live, do matter.

Supply side factors are mainly determined by the institutional structure and include fee payment, type of school, and school inputs such as pupil-teacher ratio and expenditure per pupil. Schools vary by management system or ownership, for instance public versus private schools. In the UK, where schools are either public or private, Dearden et al. (2002) find attending a selective school had a positive and significant effect on educational attainment. In South Africa most schools are public, and are categorised into poverty quintiles. Education attainment has been found to vary by quintile. Using data from the Southern and Eastern Africa Consortium for Monitoring Education Quality (SACMEQ) I and II, Van der Berg and Moses (2011) find reading and mathematics outcomes increase by school quintile. They find learners in school quintile 5 have a $2\frac{1}{2}$ to 4 year learning advantage over those in quintile 4 in the period 2000 to 2007 and the differential is even greater for the lower quintile. Their results indicate an increase in the gap between the average rich learner and the poor learner by three points. They also show that learners with good english proficiency, who own textbooks, attended preschool, and who are from rich schools, performed better. Supply side factors are often in the hands of governments and could be influenced at a country level to drive educational outcomes. In this study we focus on the effects of school inputs, which are supply side factors, and we review existing literature on the inputs in the next section.

This summary of socioeconomic factors indicates that both demand and supply side factors are important determinants of educational outcomes. One can deduce from the summary that firstly, parent's education positively influences educational outcomes, although these effects vary by gender of both parents and learners. Secondly, household income is important even when learning resources are provided at school. Thirdly, effects of resources, as defined by infrastructure or facilities are mixed. On the one hand, having learning facilities such as science laboratories and computers and libraries increases educational outcome, and having infrastructure such as a leaking roof decreased attainment Zhao and Glewwe, 2010. On the other hand having chalkboards, chairs, and desks had no effect on test scores (Bhorat and Oosthuizen, 2008). Fourth, geographical location and neighbourhood characteristics are influential in determining educational attainment. Finally, supply side factors do matter, particularly where institutional structure determines resource availability. Therefore, to identify the effects of school inputs one needs to control for as many so-

cioeconomic factors as possible. In the remainder of this section, we review the existing literature on determinants of educational outcomes in both developed and developing economies, focusing on two school inputs.

2.2.1.2 Expenditure per pupil: Evidence from developed countries

Most of the existing literature on school quality is from developed countries. The majority of the studies are based on the USA, and their findings suggest that school quality has had little to no effect on educational achievement (see a review by Dearden et al., 2002). In both developed and developing countries increasing government expenditure on education is highly contested. Those against increased spending argue that it is inefficient to continue spending more when lower outcomes are realized (Hanushek 1995; Heyneman 2004). In spite of these arguments, expenditure per pupil particularly on non-salary items has increased in most parts of the world (Lee and Barro, 2001; Wossmann, 2003; Heyneman, 2004), and some improvement in educational outcome has been attributed to this increase in expenditure on non-salary inputs (Heyneman, 2004). In developed countries, between 1960 and 1980 this expenditure increased four-fold (Lee and Barro, 2001, p473), and doubled between 1980 and 1994 (Heyneman, 2004, p445). However, school quality and outcomes differ substantially by country (Lee and Barro, 2001). It is, therefore, likely that the increase in expenditure in some developed countries has positively affected educational outcomes.

Evidence from cross-country studies is not clear on the direction of the effect of expenditure. For instance, in a cross-country study using a panel from a cross-section of countries between 1964 to 1991, Lee and Barro (2001) find a negative and insignificant effect of expenditure per pupil on test scores after the inclusion of both the pupil-teacher ratio and teacher salary. The authors argue that allocation of expenditure to other school inputs explains the change from a positive to a negative, albeit an insignificant effect. It is possible that specific regional schooling characteristics could over(under)estimate the effect of school quality. In Lee and Barro (2001) including an East Asia dummy improved the estimates of pupil-teacher ratio and teacher salary, while giving positive and significant effect on test scores. Lee and Barro find variation in effects of school quality on educational outcome among countries, but there are more than unobserved country factors that drive the results. These results are echoed by Hanushek and Woessmann (2011, p471) who find educational spending to be unrelated to educational outcomes in OECD countries. Therefore, the effect of expenditure per learner is likely to be country, region, and use specific.

The impact of increased resource allocations differs by the school systems in a country. Some countries have more productive and efficient systems than others (Hanushek, 1995). Different systems produce different results when either test scores or school progression is considered. For instance, education systems that allow for automatic progression from one class to another regardless of whether or not the students is qualified to progress would have different results if years of schooling was the outcome considered rather than if, for example, an independent test score was considered. School systems differ in their institutional structure, and the institutional structure

effects are influential in determining students' performance (Wossmann, 2003, p119). The institutional structures of an education system can create incentives that produce high-quality education (Hanushek and Woessmann, 2011, p473). In an inefficient system, creation of the right incentives could facilitate policy change (Hanushek, 1995) that would improve educational outcomes. In addition, the focus of the formal curriculum[10] in an education system determines the educational outcome, and this emphasis explains the departures in educational outcome observed in different education systems (Heyneman, 2004).

The type of educational institution matters when looking at the effect of resources on educational outcomes. The effect of expenditure per student is dependent on the type of institutional arrangements of the education system. In a cross-country study, Wossmann (2003) estimates the effect of family background, resources, and institutional arrangements of schooling systems on learners performance. The author finds an increase in expenditure per student decreased students' performance significantly. He also find institutional characteristics such as centralised examinations, curricula, and textbook approval, school autonomy, teacher influence and private competition strongly influenced students' performance. In support of this result and other research in developed countries, Hanushek and Woessmann (2011, p474) formalise three institutional incentive based policies. These are promotion of more competition to produce better educational outcomes at school level, autonomy in local decision-making in schools, and a student performance accountability system. Countries that organize their school system to promote performance and efficient use of resources have better education outcomes (Hanushek, 1995). Further, test scores in subjects that are primarily learnt in schools are better determined by school quality. Thus the emphasis placed by the formal curriculum highly influences educational outcomes (Heyneman, 2004).

Different methods have been used in the analysis of school quality. Wilson (2001) argues that the effects of attending a poor-quality school would be expected to differ when either a production function or a human capital model is applied in an analysis Poor quality may affect the amount of schooling an individual gets in a production function, but would not have an effect on the direct utility of getting an education in the human capital model. In the analysis of institutional effects on educational outcome, hierarchical structure models have been used. Wossmann (2003) uses both a hierarchical and a production function approach to study the effect of student's background, school resources and educational institutions on student's test scores in 39 countries. The results from the hierarchical model indicate that school autonomy favours student's performance. Students in countries where most decisions were taken at a central level performed worse. Funding responsibility and control had a positive effect only at an intermediate level of governance but not at either a local or national level. Results from the two models did not differ substantially. Different methods are likely to yield different effects for different controls.

[10] Whether the curriculum emphasizes on awareness of facts or application of facts to problem solving or on solving problems of unanticipated nature (Heyneman 2004, p447).

2.2.1.3 Expenditure per pupil: Evidence from developing countries

Reviews of the effect of expenditure per learner on educational outcome in developing countries offer little support for increase in this expenditure. The effect in these reviews is mixed at best. An earlier meta-analysis of studies of school quality in developing countries before 1991 by Hanushek (1995) finds the effect of expenditure per pupil to be inconsistent: Of the 12 studies reviewed, six studies had a positive and significant result while the other six had a negative and insignificant result. Hanushek, however, concedes that data quality was a concern and any inferences from the review should be drawn with caution. He relates the negative but insignificant results to inefficiency in use of the resources a view shared by Heyneman (2004). In spite of these results, schools in most countries are still publicly funded and are likely to continue to be publicly funded. The debate on how much public funding is optimal is therefore likely to continue. Hanushek (2013, p211) cautions against provision of resources to a few schools in an effort to improve school quality arguing that it would have a negative impact on access to schooling. In developing countries, school quality is often compromised for school quantity yet high quality schooling influences students to stay in school longer (Hanushek, 1995) and progress quickly from one level to the next. There is, therefore, a need to strike a balance between quality and quantity.

Hanushek (1995) concludes that increased expenditure had no effect on education outcome. However, his approach of aggregating the direction of estimates has been faulted. Hedges *et al.* (1994) and Kremer (1995) (as cited in Lee and Barro, 2001, p469) argue that the studies Hanushek reviews had an underlying high probability of giving insignificant coefficients which biased the results towards insignificance. According to Lee and Barro (2001, p469), an accurate aggregation gives positive and significant effects of school inputs on school outcome. Although more empirical work is needed, a few studies have found a positive effect. Glewwe et al. (2011) reviews studies on the effect of school resources on educational outcome in developing countries between 1990 and 2010. Glewwe and colleagues find only three studies had specifically looked at expenditure per pupil in this period. Their results are mixed. Two of the three studies they review showed a negative and significant effect, while one study found a positive and significant effect. They infer the unexpected negative sign to be an indication of compensatory funding where poor schools get more funding, or it could be as a consequence of inclusion of other school characteristics in the analysis as in the two studies reviewed (Glewwe et al., 2011, p27). A similar argument is presented by Lee and Barro (2001). Although there is not much support for a positive effect in the developed country literature, there is evidence indicating that resources directly used by pupils have a significant effect on education outcome. A number of studies indicate that minimal resources such as textbooks should be provided to achieve better outcomes (Hanushek, 1995). Studies in developing countries indicate textbooks and writing materials play an important role in improving education outcomes (Hanushek, 1995). More recent empirical evidence agrees with this. Most of the studies reviewed by Glewwe et al. (2011) find school infrastructure and pedagogical materials such as textbooks, workbooks, blackboards, libraries, and high quality walls, floors and roofs have

a positive effect on student learning and enrollment. They however, find no support for these results in randomised control trials. An increase in resources, particularly those that affect learning directly, is a prerequisite for improving educational outcome.

Although randomised control trials offer the best impact studies, these studies are not common in studies on school inputs (quality). A study by Das et al. (2011) give an insight into the effects of expenditure on both schooling outcome and household spending on education. In their study, Das and colleagues compare the effect of school and household characteristics on test scores using a survey in Zambia, and an experiment in India. In the Zambian case study, they consider the effect of two school grants meant for non-salary purposes namely an unanticipated discretionary grant that was offered once, and an anticipated rule-based grant introduced in 2001 offered annually. They find that, on the one hand, the anticipated grant reduced household school expenditure significantly and had no effect on test score. On the other hand, the unanticipated grants had virtually no effect on the household's expenditure on schooling and significantly increased the score in English, but not the score in Mathematics. In the Indian case, they consider the effect of the grants by treating the grant in the first year as an unanticipated grant, and the following year's grant as an anticipated grant. The results from the Indian case indicate no adjustment in household expenditure on schooling in year one, but it indicates an adjustment in year two, and test scores were significantly higher in year one but not in year two. These results are similar to those in the Zambian case. They indicate that the effects of school inputs are dependent on "whether such inputs are anticipated or not, and the extent of substitutability between household and inputs in education production function" (Das et al., 2011). To ensure expenditure allocation has the intended impact on educational outcomes, one can argue that expenditure is best spent on school inputs that cannot be easily substituted by the household.

Additional learning resources, other than teachers, are important in improving learning outcomes. In a study of determinants of school choice and student achievement in Ghana, Glewwe and Jacoby (1994) find per capita expenditure has a positive effect on reading score but not on mathematics score. They associate the difference in the score to potential presence of omitted attendance bias. They also find that having a school library significantly increases math test scores, and having textbooks and desks weakly influenced test scores. However, they find that a high proportion of unusable and leaking classrooms reduced grade attainment significantly. The importance of these findings cannot be overstated, particularly in the context of developing countries where classroom conditions are often poor. Various elements in the school environment influence educational outcomes. For instance, in a study of schooling inequality in Vietnam Rolleston and Krutikova (2014) find that class resources are important for individual attainment, teacher absenteeism decreased attainment, having repeated, and the presence of repeaters negatively affects individual performance. High grade repetition and high dropout rates are indicators of poor quality. Students stay longer in good quality schools and students often drop out in poorer schools (Hanushek, 1995).

Historical racial considerations are common in studies of school quality in South Africa. Stud-

ies on educational outcomes indicate that apartheid history is still a driving factor in post-apartheid South Africa. Focusing on historically disadvantaged schools using 2000 SACMEQ data for South Africa, Gustafsson (2007) finds that availability of school buildings and equipment and textbooks increases performance. He finds that historically disadvantaged schools had lower test scores. This is despite measures to redress past inequalities. At the dawn of democracy in South Africa, the issue of fee payment was debated. Although most public schools are no fee paying schools (schools in quintile 1, 2 and 3) by law, voluntary contributions are allowed and some schools do charge fees (schools in quintile 4 and 5). Controlling for racial composition, educational resources and school fees Van der Berg (2007) finds that educational resources and school fees increase matriculation pass rates, and Bhorat and Oosthuizen (2008) find that a user fee was a significant determinant of pass rate. Although pupil-teacher ratio is considered by the government in its allocation of expenditure per learner, hiring of teachers using these funds is not permitted. However, schools can use funds raised in the school to hire teachers. Bhorat and Oosthuizen (2008) also find Whites are advantaged in both learning and non-learning resources, and in household characteristics, relative to other population groups. In one of the few studies that examine the effect of government subsidies (proxied by local resources) on school quality, Yamauchi (2011) finds both government subsidies and school fees improve school quality significantly, and the effect of the subsidy is greater. The author finds that an increase in the subsidy led to an increase in school quality. Most studies in South Africa use a production function approach in their analysis. Gustafsson (2007) uses a hierarchical linear model and a least square model and finds the hierarchical model understates the school, teacher, and socio-economic status effects on test scores, compared to the least square model.

The review of literature on the effects of expenditure per pupil on educational outcomes indicate varied effects in both developed and developing countries. There are a number of conclusions that can be drawn from this. First, more studies have been undertaken in developed countries than in developing countries and reviews of these studies indicate mixed effects. Second, different methods of analysis have been employed to compare results when different approaches are employed such as hierarchical linear models and the production function approach (Wossmann, 2003; Gustafsson, 2007), and production function and human capital model (Wilson, 2001). Results from the different methods are mixed. Wilson (2001) and Gustafsson (2007) find the effects differ, but Wossmann (2003) finds no substantial difference in effects. Third, effects differ by country and region and are use specific. Fourth, expenditure on resources that are directly used for learning increases education attainment (Bhorat and Oosthuizen, 2008), and effect of expenditure per pupil is best realised if spending is on school inputs not easily substitutable by households Das et al. (2011). Fifth, school system and institutional structure are important determinants of outcomes . Last, there is limited evidence on expenditure per pupil in South Africa yet there have been increased expenditure per pupil allocations.

2.2.1.4 Pupil per teacher ratio: Evidence from developed countries

Pupil-teacher ratio and class size are used in literature interchangeably, however, they are different. A teacher-pupil ratio often includes both teaching and non-teaching staff and as a result this ratio may be smaller than the class size indicates. Therefore, the results from the two could be different. In literature, the results are generally mixed. Wossmann (2003) finds class size had a positive and significant effect on mathematics and science test scores, while the teacher-pupil ratio has a positive but insignificant effect. He finds the class size result holds even when the average grade-level class size in the school is used as an instrument. In a cross-country analysis using country level data, Lee and Barro (2001) find lower pupil-teacher ratios increase students performance in test scores in mathematics, science, and reading. They also find an increase in the pupil-teacher ratio led to an increase in repetition and dropout rates. A small teacher-pupil ratio reduces repetition and dropout when parental income and education, and other school inputs were accounted for (Lee and Barro, 2001). The effect of teacher-pupil ratio by gender has also been examined. Using the British National Child Development Survey (NCDS) Dearden et al. (2002) find the effect of teacher-pupil ratio for educational attainment of both both men and women was insignificant and nonlinear. They, however, find lower ability women benefited from a low teacher-pupil ratio.

Class size is a more common measure of school quality. The issue of whether class size matters in learner achievement has been debated. Various factors influence the effect of class size, such as subject, race, and student ability (Akerhielm, 1995). In an analysis of class size, Akerhielm (1995) finds an increase in class size leads to an increase in student performance and the result was robust with family background included. However, its effect varied by race and subject. Akerhielm's findings also indicate that the effects of class size depend on the level of class size and had nonlinear effects. The use of students' ability in class allocation is common, and is a likely source of bias in studying the effect of class size. In the US, public schools that allocate lower ability students to smaller classes had higher effects of class size on test scores(Akerhielm, 1995).

A natural experiment is one of the most appropriate way to deal with endogeneity on teacher-pupil ratio or class size which is a common problem in any study of school quality. Angrist and Lavy (1999) uses the Maimonides rule as an instrument for class size to study the impact of class size on test scores. They find a positive relationship between class size and test scores, and their interaction of class size with a variable for percentage of disadvantaged students shows smaller classes largely benefited schools with the highest proportion of poor background pupils. They, however, find mainly insignificant results for grade 3 which they associate with the cumulative nature of education meaning that benefits of smaller classes take time to be realized.

2.2.1.5 Pupil per teacher ratio: Evidence from developing countries

In developing countries, where resources are scarce, the teacher-pupil ratio is substantially higher since schools are characterized by overcrowding of students, and under-staffing of teachers. One

would therefore expect that a lower pupil-teacher ratio would have a positive effect on any educational outcome. In a detailed review of school resources and educational outcome studies, Glewwe et al. (2011) find most of the studies they review found teacher-pupil ratio had a negative and significant effect. However, there were a few studies with a positive and significant effect. A positive coefficient could be as a result of high quality schools having large classes when they enroll high numbers of students. Randomised control studies in the review find a negative and significant effect. According to Glewwe et al. (2011), using a randomized control trial is the only possible way to know the effect of this ratio. Most of the studies reviewed by Hanushek (1995) indicate an ambiguous relationship between teacher-pupil ratio and educational outcome despite the inclusion of family background and other educational inputs. Overall Hanushek's review showed no support for the advantages of a low teacher-pupil ratio.

In South Africa, the literature on the effects of school quality is limited but growing. A pioneering study by Case and Deaton (1999) investigates the relationship between pupil-teacher ratio and three educational outcomes namely years of schooling, enrollment, and numeracy and literacy test scores. They find that the pupil-teacher ratio had a negative and insignificant effect for Whites, but it had a negative and significant effect on years of schooling of Black pupils. However, they find no evidence of nonlinear effects of the pupil-teacher ratio. Their interaction of age with pupil-teacher ratio, household head education, and household per capita expenditure show resources affect learners' progression through the education system from enrollment to completion by influencing the rate of progression through school. A number of studies in the post-apartheid era have found an ambiguous effect. Gustafsson (2007) finds the effect of class size on test scores did not stand out in historically disadvantaged school. Similarly, using national matriculation performance data Van der Berg (2007) finds teacher-pupil ratio was insignificant in Black schools, and Bhorat and Oosthuizen (2008) in a quantile regression, find the ratio was insignificant except above the 95th percentile of the school pass rate. Moloi and Chetty (2011) find that education quality had improved from 2000 to 2007 since the mean teacher-pupil ratio decreased from 1:37 to 1:34, however, the class size had increased from 42 to 44. This is a higher class size than the government set level of 40 for primary and 35 for secondary school (Department of Basic Education 2010b p50). These few studies indicate an inconclusive effect of the pupil-teacher ratio which warrants more research.

We can draw a few conclusions from this review of the literature on pupil-teacher ratio. One, the effects of pupil-teacher ratio on educational outcomes in both developed and developing countries is not clear. The international literature is equally conflicted, some find a positive effect (Angrist and Lavy, 1999), some find no effect (Dearden et al., 2002), while others find the effect differs by gender, race, subject, and student ability (Akerhielm, 1995), and by socio-economic status (Angrist and Lavy, 1999). Card and Krueger (1992) find school quality to have a positive effect on earnings for men, and Dearden et al. (2002) find it has no effect on wages of men. The results are equally mixed for South Africa. Case and Deaton (1999) find negative and significant effects for Blacks, but no effect for whites. Van der Berg (2007) find no effect for Black students, and (Bhorat and

Oosthuizen, 2008) no effect but for the 95th percentile. Second, the effects of school inputs differ between developed and developing countries. Third, educational outcome, in addition to school inputs, is determined by education system in place, individual characteristics, family background and neighborhood characteristics. Fourth, South African literature has mainly focused on pupil-teacher ratio as a measure of school inputs (school quality) and specialised test score and matriculation as the main education outcomes. Fifth, the literature on the effects of pupil-teacher ratio in South Africa is limited both in amount and educational outcomes considered. In this regard, it is prudent to include more measures of school inputs in an analysis of South African school quality, an effort that this study makes. An attempt to quantify school effects on educational attainment is the focus of this chapter. We consider the effect of the pupil-teacher ratio and expenditure per pupil, once we control for personal, household, school and neighborhood characteristics. In the next section, we present the conceptual framework and method of analysis employed to estimate the effects of these school inputs on educational attainment.

2.3 Approach and Method

2.3.1 Conceptual and analytical framework

In this section, we outline a framework that guides us in the estimation and interpretation of the impact of pupil-teacher ratio, expenditure per pupil, and socioeconomic factors on two related educational outcomes. These are actual education attainment and a relative educational attainment index. We assume that in South Africa parents (or guardians) are the main decision makers on investment in post compulsory education, that is, after grade 9 or age 15. Households (in particular parents or primary care givers) are assumed to maximizes (life-cycle) utility (U) at different points in time subject to budget constraints (Y) and a schooling production function (S) (Aturupane et al., 2013; Behrman, 2010; Glewwe and Kremer, 2006). We present this as follows:

$$U = u(C, L, S) \tag{2.1}$$

$$P^c C + P^m M = Y \tag{2.2}$$

$$S = f_{pf}\,(F; C; M, Q; H\,P^m) \tag{2.3}$$

The utility function depends on household consumption of goods and services (C), leisure (L), and the child's schooling (S). The child's schooling is assumed to enter parents' preference function directly (say because parents are altruistic). It is, therefore, considered purely as a consumption good from which the household directly derives utility. This also implies that the schooling outcomes of siblings are not independent. P^c are prices of consumption goods and services, P^m are prices of school inputs (M), and Y is the household income. Prices are assumed to be exogenously determined. Household income affects learning indirectly through purchase of inputs and

is likely to be endogenous when learners work to contribute towards household income and as a consequence allocate less time to schooling.

In formulating the education production function specified in equation 2.3, we follow the education production function by Glewwe and Kremer (2006), Mani et al. (2009), Behrman (2010), and Aturupane et al. (2013). It is a structural relationship between educational outcome S; *'pf'* denotes a production function and a vector of household characteristics F, a vector of child $i's$ characteristics C, a vector of school and teacher characteristics Q, a vector of neighborhood characteristics H and a vector of schooling inputs M. S can be considered as one output for simplicity of analysis although several outputs are possible (Glewwe and Kremer, 2006). These include outcome measures such as enrollment, grade attainment, relative grade attainment, and test scores. Family characteristics include household socio-economic variables such as income and home location, and parental schooling. Some family characteristics are often unobservable. This includes parental taste for schooling which would affect the type of school in which they choose to enroll their children, and their overall involvement in their children's education. Child's characteristics C includes observable characteristics such as age, gender, and race, and unobservable characteristics such as ability and motivation which are endogenous educational inputs M are under the control of parents making them endogenous in the production of education. They include school attendance, time spent studying at home, uniforms, books and other school supplies purchased by parents. The household is assumed to derive utility from inputs M indirectly through schooling S. School and teacher characteristics Q include pupil-teacher ratio, expenditure per pupil, type of school per apartheid classification, and distance to school. Neighbourhood characteristics H include access to piped water, availability of electricity and street lighting, availability of landline telephone, collection of refuse by municipality, availability of toilet facility, average household size and house size, and community unemployment rate.

We assume that parents choose which school to enroll their child from more than one school, and they can influence some characteristics of the school. For instance, in South Africa, parents are allowed to give voluntary contributions towards the running of the school, and in some schools they have to pay fees. Parents are included in the School Governing Body (SGB) and by this inclusion they are involved in making decisions concerning how school funds are used. They could, therefore, influence decisions on for example hiring more teachers or buying instructional materials which could affect school quality. In addition, by law parents have a say in whether, and what fees can be charged in the school their child attends. This means that the school and teacher characteristics Q and prices P^m are endogenous. Such policies on how school funds are used could interact with neighbourhood characteristics to influence school quality and prices.

Households maximize their utility with respect to each schooling choice and choose the school that gives the highest utility. Conditional on choosing that school, they choose the school inputs M.

We consider a reduced form equation which is expressed as follows:

$$S^* = g_{rf}(Y, F, C, E, H) \tag{2.4}$$

where subscript rf stands for reduced form. This reduced form equation 2.4 is a causal relationship rather than a structural production function because it reflects household preferences in schooling and it includes prices of school inputs. We estimate the impact of school inputs and socio-economic characteristics on educational outcome of African South Africans from equation 2.4 mainly based on the data available to us in the National Income Dynamics Study (NIDS). This estimimation allows for the changes in school inputs M due to changes in E to be considered (total derivative) hence allowing us to show what really happens to educational outcome when school quality changes although it underestimates the effect of E (for a detailed discussion see Glewwe and Kremer (2006) and Aturupane et al. (2013)).

2.3.2 Empirical specification

The specification adopted in this study is similar to that applied in Dearden et al. (2002) and Mani et al. (2009). We analyse the effect of school inputs (pupil-teacher ratio and expenditure per learner) and socio-economic characteristics on educational outcomes, specifically, the highest education level attained S_i for individual i. Following 2.4 we define our estimation model as follows:

$$S_i = \alpha_0 + \alpha_1 M_i^I + \alpha_2 M_i^O + \alpha_3 C_i + \alpha_4 F_i + \alpha_5 H_i + \varepsilon_i \quad \forall i = 1...n, \tag{2.5}$$

where for individual i : M_i^I, represents school inputs such as pupil-teacher ratio and expenditure per learner; M_i^O represents other school characteristics such as school type and distance to school; C_i represents individual characteristics such as age, race, and gender; F_i represents family characteristics such as monthly household per capita income, parents' education level, and number of siblings; and H_i represents neighbourhood characteristics such as access to piped water, availability of electricity and street lighting, availability of landline telephone, collection of refuse by municipality, availability of toilet facility, average household size and house size, and community unemployment rate. We let $A_i = \left[F_i, C_i, H_i, M_i^O\right]$ represent the observable characteristics except for school inputs. The measure of the effect of school quality on education outcome is given by α_1. To estimate this effect, we assume that any selection in school inputs is based on observable characteristics, and therefore conditioning the estimation of this effect on observable characteristics is sufficient to deal with endogeneity of school input variables. We also assume that, on average, individuals similar in observable characteristics but who attend schools with different school input variables do not differ in the unobservable characteristics; ε_i , and that these unobservable characteristics are randomly distributed. That is to say that: $E(\varepsilon_i|A_i, M_i^I) = E(\varepsilon_i|A_i)$.

Following Dearden et al. (2002) we extend this model to allow for school quality heterogeneity in the population *i.e.* $\alpha_{1i} = \alpha_1 + \varepsilon_i$ where the variance of the error terms is greater than zero *i.e*

$var(\varepsilon_i) > 0$. In any population, M_i^I is likely to be heterogeneous particularly because schools, parents, and local governments are often in a position to influence the school inputs available in a given school. However, although the effect of heterogeneous school inputs may be heterogeneous in a population, we assume that the person making the choice of the school inputs only knows the average population values of α_1. This implies that parents do not know the exact returns of M_i^I that would accrue to their child. Therefore, $E(\varepsilon_i|A_i, M_i^I)M_i^I = E(\varepsilon_i|A_i)M_i^I$ and the average school quality effect on education outcome α_1 is modeled as follows:

$$S_i = \alpha_0 + \alpha_1 M_i^I + \alpha_{2t} A_i + \alpha_{3t} \left(A_i \otimes M_i^I\right) + \varepsilon_i \tag{2.6}$$

where $E(\varepsilon_i|A_i, M_i^I) = 0$ and ε_i represents both a random measurement error and unobserved effects of school quality on education level attained. Heterogeneity in the effect of school input M_i^I is given by coefficient α_3.

2.3.2.1 Econometric estimation approach

Our outcome variable is the individual's highest education level attained. In the sample, this variable is discrete and non-normally distributed, that is, it is spiked at no schooling and at matriculation level. Some of the observations on highest level of education attained are right censored for respondents who were still enrolled in 2008. Further, given our variables of interest are pupil-teacher ratio and expenditure per pupil, we do not consider individuals with no schooling which makes education level attained left censored. In the presence of non-normality, discreteness, and censoring, using an ordinary least square estimation would potentially yield biased and inconsistent estimates (Tansel, 1997; Glick and Sahn, 2000; Mani et al., 2009, p12). Glick and Sahn (2000, p69) argue that the censoring problem can be overcome by restricting the sample to individuals old enough to have completed, at a risk of considering older individuals whose schooling determinants could have changed at the time of the survey. We therefore restrict our sample to younger respondents (15 to 30 year olds). To address the problem of censoring, we restrict respondents who have completed education to one sample and use an ordered probit specification to obtain unbiased and consistent estimates. A similar approach has been adopted in (Tansel, 2002; Mani et al., 2009).

To avoid sub-sampling, one could use a censored ordered probit as in (King and Lillard,1983[11] and 1987, and Zhao and Glewwe, 2010) where respondents who are still enrolled and those who are not enrolled enter the likelihood function separately. The limitation of this approach, however, is its assumption that the respondents who indicate as not enrolled would not be re-entering school, a possibility in South Africa and particularly in our sample of 15 to 30 year olds. This age group comprises students who have at least completed grade 9, and are therefore likely to re-enter school. Re-enrollment is an identifiable feature in school-to-work transition for the youth in South Africa. Using the Cape Area Panel Study of youth aged 14 to 22, Pugatch's(2012) panel analysis shows

[11] This was the first study to analyze cumulative schooling.

21% re-enroll before completing high school while a third re-enroll at some point during their schooling life, and Levinsohn and Pugatch (2010) show that 6% of the youth returned to school on a part-time basis before obtaining their first job.

We employ an ordered probit to estimate the effect of school inputs on educational attainment for Africans not enrolled at any institution of learning who we refer as having achieved a given level of education. Education level attained has a natural order relation to it. However, the difference between having no schooling and having a primary(general) education is not the same as the difference between having a matriculation and having higher education. The benefits of schooling are therefore unlikely to be the same, thus ruling out the use of an ordinary least square that assumes a linear expected value locus. An ordered probit gives us the advantage of not treating the dependent variable as normally distributed, since education level attained is discrete and bimodal (In the sample, it is spiked at no schooling and at grade 12 (matriculation) level). We assume that we can explain all educational outcomes using a single index model given by the right-hand equations 2.5 and 2.6. The underlying latent variable model in consideration is:

$$l_i^* = x_i\alpha + \varepsilon_i \quad with \quad \varepsilon \sim N(0,\sigma^2) \tag{2.7}$$

where l_i^* linearly dependent on x_i and $x_i = [m_i, f_i, c_i, h_i]$ represents all the observable variables. We assume the disturbance term ε_i is independently and normally distributed with a mean 0 and variance σ^2. We also assume a standard normal distribution hence $\varepsilon \sim N(0,1)$. Although l_i^* is unobserved, we observe where an individual's index lies given categories $j = 1,...,J$. The index is defined by its unknown lower bound μ_{j-1} and upper bound μ_J. We assume that $\mu_{j-1} = -\infty$ such that $F(-\infty) = 0$, and $\mu_J = \infty$ such that $F(\infty) = 1$. Education level attained is divided into four ordinal categories namely: general level (Grade 1 to grade 9), some secondary education (Grade 10, 11), matriculation (grade 12), and higher level (Grade 9, 10 or 11 plus any further education and training certificate, and grade 12 plus any additional higher education level attained including certificate, diploma, or degree).

In the ordered probit model, the cumulative probability of the discrete outcome is related to the single index as follows:

$$Pr[l \leq j|x] = F(\mu_j - x'\alpha) \qquad j = 1,...,J \tag{2.8}$$

where μ_j and $\beta_{(k\times 1)}$denote unknown parameters in the model, and $F(\cdot)$ is a standard normal distribution function. For well defined probabilities we restrict μ_j>μ_{j-1} ,$\forall j$, and the bounds of $F(\cdot)$ are as defined above. To ensure identification of these parameters we assume that $\alpha_0 = 0$, that is, x does not contain a constant, and we normalize the variance of the distribution of function F.

A common practice is to employ a standard ordered probit in this kind of analysis, but in analyzing marginal product effects, the standard ordered probit is not without limitations which mainly stems from its assumptions. These assumptions are the single crossing property, the constant threshold, and the distributional assumption that does not allow for the realization of hetero-

geneity between individuals (Boes and Winkelmann, 2006). We first run a standard ordered probit and test for the assumption of parallel lines. A likelihood ratio test (using omodel in Stata) on the assumption of the parallel lines assumption in the standard ordered probit gives us a $\chi^2_{32} = 131.72$ with a $p - value = 0$ an indication that this assumption is violated at 1%-level. Due to this violations and to deal with the limitations of the standard ordered probit, we follow a generalization of this model as proposed by Boes and Winkelmann (2006). This model relaxes the three assumptions of the standard ordered probit, and a partial generalization of this model by Williams (2006) gives more flexibility. This partial generalized ordered probit allows for heterogeneous controls that vary in each category to vary, and those that are constant in each category to remain constant. A test of the parallel lines for each variable in the partial generalized model indicates that the quadratic logarithmic pupil-teacher ratio, both the logarithmic expenditure per pupil and its quadratic, quadratic age, gender, attending independent homeland or a self-governing territory school, and distance to school violate the parallel line assumption and are therefore allowed to vary. A Wald test of the partial generalized model gives us a $\chi^2_{26} = 30.64$ with a $p - value = 0.242$ indicating that this model does not violate the parallel lines regression. This means that the effects of some of our controls vary with the education level attained while others remain constant at each level. This model generalizes the threshold parameters by making them dependent on covariates as follows:

$$\mu_j = \tilde{\mu}_j + x'\omega_j \tag{2.9}$$

where ω_j is a $k \times 1$-dimensional vector of response specific parameters. We replace μ_j in equation 2.8 to obtain the cumulative probabilities in the generalized threshold model as:

$$Pr[l \leq j|x] = F(\tilde{\mu}_j + x'\omega_j - x'\alpha) = F(\mu_j - x'\alpha_j) \qquad j = 1, ..., J$$

where $\mu_j = -\infty$ and $\mu_j = \infty$ are as before, and $\alpha_j \equiv \omega_j - \alpha$ since ω_j and α cannot be separately identified using the same x entering the index and generalized threshold models. This allows us to estimate the different parameters of x for each category.

We also consider an alternative measure of educational attainment by creating a relative measure of educational attainment. Following Chernichovsky (1977) we create a relative educational attainment index. This measure exploits the argument that assuming household characteristics are constant and neighbourhood effects are exogenously determined, a child's level of schooling is determined by their individual characteristics. In our case we have age and gender as the individual characteristics. Therefore, we standardise the individual's level of schooling by age and gender. We do this by estimating an ordinary least square estimation of year of schooling as determined by age, gender, and their interaction as follows:

$$S_i = \beta_0 + \beta_1 age_i + \beta_2 gender_i + \beta_3 age_i * gender_i + \nu_i \tag{2.10}$$

where ν_i is a residual that accounts for variation in individual's schooling which could be at-

tributed to other observed household characteristics and unobserved characteristics. We get the predicted level of schooling $\hat{S}_i$ which is fixed for all children of the same age and gender from equation 2.10, and calculate the relative educational attainment index as the ratio between the observed years of schooling S_i and predicted years of schooling $\hat{S}_i$, that is,

$$R_i = \frac{S_i}{\hat{S}_i}$$

where R_i is the relative educational attainment index. For example, using a predicted average educational attainment of 8 years for a 15 year old individual i , a 15 year old individual i with educational attainment of 6 years would have an index equal to 0.75 while another 15 year old individual i with 8 years of schooling would have an index equal to 1. As outlined in Chernichovsky (1977, p7) this index is a positive and linear function of the years of schooling, and for older children it assigns a lower value to a given level of schooling and to a given increase in level of schooling. According to Chernichovsky this index performs best where *(i)* school entry age is uniform which is consistent in measuring schooling in absolute term, in this case by age group and gender, and *(ii)* the population of school-age children is made up of children who have never attended school, dropouts, repeaters, and children who normally attend school regularly. By taking dropout and repetition into consideration children with the same level of education but of different ages are assigned different levels with the younger child assigned a higher level. However, the index fails to distinguish late entrants from dropouts and may assign a high value to dropouts. This is not a concern for us since education is compulsory from age 7 in South Africa which ameliorates possibilities of late entrance. A major limitation of this measure is that it does not take into consideration whether the individual is enrolled or not, but treats both groups identically which implies that our estimates are likely to be biased. The measure also has a distortionary feature since a lag in attainment at younger age distorts the index more heavily than a lag at an older age (King and Lillard, 1983).

Other similar measures include a measure of relative grade attained which is defined as the actual grade attained divided by mean educational attainment for the relevant age group, by Birdsall (1982 as cited in Mani et al., 2009, p12). Another is as proposed by Mani et al. (2009, p8). They define a relative grade attained as actual grade attained divided by the potential grade. Mani and colleagues define potential grade as the highest grade the individual would have completed had they completed one grade at age 7 and continued to accumulate an additional grade each year. They point out that estimates from this measure would, however, be biased downwards for individuals who enter school late and accumulate more than no grade at all. These measures of relative educational attainment take into account delays in both enrollment and level attained. In addition, they create a continuous variable of highest education level attained from which consistent least square estimates can be estimated. The estimates are consistent since the relative educational attainment measure, by considering individual characteristics, takes care of censoring bias that could arise from the sample of individuals still enrolled. We undertake an ordinary least square estimation of the determinants

of the relative educational attainment index using equation 2.4 with relative educational attainment as our dependent variable.

If all factors that determine education level attained are accurately measured and exogenous such that they are uncorrelated with the random component ε, then our estimates from the ordered probit and OLS would be unbiased. However, this is not the case in the data available to us. In the following section we discuss the variables used in the analysis, and the likely identification concerns related to the National Income Dynamics Survey Wave 1 (2008) data used in this study.

2.3.2.2 Data issues and identification concerns

In the estimation of school effects, parental choice of school is likely to be a source of endogeneity in school inputs. Parents who place a higher value on education are likely to choose a school with a small pupil-teacher ratio or smaller classes. They are also likely to be more involved in their children's performance and progression in an effort to ensure their children perform well. Such parents may also relocate to an area with a school they consider better, and they may be willing to spend more on housing to access a better school. They are are also more likely to engage in political actions that would lead to improved local school quality and funding (Case and Deaton, 1999). This bias from parent's choice of a school is also likely when a parent with a high ability child selects a high or low resource school, while a similar parent with a low ability student chooses differently (Graddy and Stevens, 2003). Graddy and Stevens argue that it is possible for high ability students to do well anyway. They also state that when high ability students study with other high ability students their performance is likely to be high, and returns to school resources would be differentiated by students ability. In South Africa, learners residing within a 5km radius of a school are given priority in admission to the school. Parents who place a high value on education are likely to relocate to a neighbourhood with a higher quintile school. In South Africa these would mainly be the 'model C' schools which were formerly Whites-only schools. They have a reputation of better matriculation[12] performance, have better learning facilities, and on average have a lower pupil-teacher ratio and class sizes, but receive a lower expenditure per pupil allocation. Therefore, parental choice of the school is likely to be influenced by the type of school available, as formerly defined along the racial lines.

School fees can also be an influencing factor in school enrollment. In South Africa schools are either fee paying or non-fee paying. Parents who do not want or cannot afford to spend on the education of their children may opt to enroll their children in lower quintile schools that do not have fees (schools in quintile 1 through 3). Pupils who enroll in these schools are exempt from paying any school fees. This likely self-selection into schools may lead to an upward bias of the estimates of school quality measures. However, a downward bias is likely if parents with pupils in good schools invest less time and resources in their children's education (Dearden et al., 2002).

[12]Matriculation is a national examination administered at the end of grade 12 and marks the end of secondary schooling.

To avoid this bias we do not estimate the structural equation 2.3 **??** and **??** but the reduced form equation 2.4. In the analysis we control for observable student, family, neighbourhood, and school characteristics. This includes a measure for distance to school to ameliorate any likely bias due to school choice selection (Wilson, 2001, p522), and we include parent's education as an indicator of parental taste for education.

Another source of choice is in the sample of respondents. In developing countries, schooling is characterised by complete non-attendance of some children, grade repetition, school dropout, and attendance at different schools at different points of schooling (Glewwe, 2002). In South Africa, education is compulsory up to grade 9 or age 15, whichever comes first. Non-attendance until grade 9 turns out not to be a problem, as enrolment rates are high. Statistics do confirm this. For instance, in 2009 the participation rate for learners of compulsory school-going age (7-15) was 99% (Department of Basic Education, 2011, p9). However, the education system in South Africa is not immune to grade repetition and high drop-out rates. The average repetition rate at the basic level in 2009 was 9% (Department of Basic Education, 2011, p10), but the rate is higher at tertiary level. According to the Council on Higher Education(2013) only 29% of those enrolled in 2006 for a 3 year degree graduated in 2008 and 5% dropped out in 2008[13] –their final year. School dropout at different levels of schooling is a serious issue for education policy makers in South Africa: In 2007/2008 the drop-out rate averaged 1.6% (grades 1-8), 6.5% (grade 9), 11.5% (grade 10) and 11.8% (grade11) (DBE, 2011c p3). Attrition is higher in post compulsory education, for instance, in 2009 only 39% of students complete secondary education (grade 12) (DBE, 2011c p3). Given these issues, selection is a likely source of bias in an analysis of education attained in South Africa. Although NIDS includes data on grade repetition, controlling for it would bias our variables of interest (school inputs) since they are likely to be correlated, and grade repetition is a likely outcome variable in school quality (for a discussion see Angrist and Pischke, 2008). However, inclusion of an analysis of our relative educational attainment index addresses some of these problems.

Unobserved learner, family, neighbourhood and school characteristics are a likely source of omitted variable bias. These include learner's ability,[14] parents belief in the value of schooling, schools criteria in allocation of teachers, and class size. These unobservables are in the error terms, and are likely to be correlated with the observed characteristics. In household surveys such as NIDS it is difficult to observe variables which have no formal parameters, such as teacher qualifications, student class allocation criterion, and school admission criterion. Glewwe (2002) argues that omission of such variables leads to an upward bias of the school quality variable estimates. An inclusion of as many school and teacher related quality variables as possible would mitigate this

[13]This is excluding those enrolled at the University of South Africa which mainly offer courses by distance learning.

[14]A numeracy test score was administered in Wave 1, this score cannot be used in this study as a proxy for measure of cognitive ability since the test scores are likely to be correlated with our variables of interest–pupil-teacher ratio and expenditure per pupil–since the respondents had already started schooling when the test was administered. Also the score could be considered as an outcome variable in a school quality analysis hence it would be a bad control if included (Angrist and Pischke 2008).

bias. However, a small sample and high correlation between these variables could lead to insignificant estimates (Glewwe, 2002). Further, schools are likely to select students conditional on their ability. The selected students are more likely to have better performance and attain higher level of education. Therefore, this selection is likely to bias school quality variables upwards. A student with higher ability is likely to attain a higher level of education even with lower school quality variables. The lack of a measure of ability in our study is likely to lead to a higher estimate of school quality for students with higher ability, and vice versa.

Measurement error in variables is another source of endogeneity. There is a possibility that the highest education attained is measured with error. People tend to over- or under-report their level of education (Cahuc and Zylberberg, 2004, p91), also, reported education attained is often rounded-off, resulting in measurement error. The measurement error in education attained is not a concern here since endogeneity is usually not a problem in cases where the mismeasured variable is a dependent variable because the zero conditional mean assumption is not violated. However, when the mismeasured variable is the independent variable then endogeneity is a problem, since the mismeasurement could lead to higher standard errors. It is likely that pupil-teacher ratio and expenditure per pupil are mismeasured. Although we use expenditure per pupil as gazetted by the government, most provinces do the final allocation based on a provincial quintile rule which could differ in some provinces. Similarly our pupil-teacher ratio is as reported by schools to the Department of Basic Education and one cannot rule out misreporting. Glewwe (2002) argues that such errors are most probably random and the mismeasurement could lead to underestimation of the school quality effects on educational attainment. In our sample of Africans in the NIDS 2008 data, we use household and neighbourhood characteristics to analyze their effect on educational attainment. While this data is an accurate match for respondents who were still in school in 2008, it may be inaccurate for respondents who are no longer in school but completed years earlier. However, we assume there has been limited change in these characteristics, and also focus our sample on a younger population, 15 to 30 year olds. If this assumption does not hold, the presence of measurement error could result in biased estimates of the effects of these characteristics on educational attainment. If the measurement errors in the school quality variables are random, we are likely to underestimate the effect of school input variables on education attained. It would thus be hard to tell the direction of the bias due to nonrandom measurement errors.

The final source of endogeneity we consider is education financing. The Government of South Africa is the major funder of public education. In addition to employing the majority of teachers, it allocates funds to schools to finance non-personnel and non-capital expenditure. These include textbooks, stationary, non-educational items (such as cleaning materials), and services relating to maintenance and repair. These funds are allocated in the form of expenditure per pupil and are determined by the quintile to which the school is assigned. Schools are not allowed to use these funds for hiring new teachers, building classrooms or other physical infrastructure (South African School Amendment Act, no 84 of 1996, as amended: 2006 p28). However, fee paying schools

(schools in quintile 4 and 5) are allowed to raise funds internally. Parents are involved in the decision of how much they should pay as school fees, and by being part of the school governing body, make decisions on how these funds are used. These funds can be used to hire more staff (both teaching and non-teaching), or build new classes , or provide other learning resources needed in the school. This is likely to introduce bias in our pupil-teacher ratio estimate for schools in quintile 4 and 5. These schools are also likely to have a smaller pupil-teacher ratio than the no-fee schools, in quintile 1, 2, and 3. It is also possible that the expenditure per pupil may be biased downward by the effect of fees, particularly if the total amount raised is higher than the government allocated funds. Fee based schools receive at most a 15% allocation of the resources (quintile 4 schools) and at least 5% (quintile 5). Since schools in quintile 4 and 5 are the formerly White schools, we control for school type based on apartheid school classification. However, we are not concerned with endogeneity in their pupil-teacher ratio due to per pupil expenditure since this allocation cannot be used to either hire teachers or build classes. The school governing board has the mandate, by law, to determine how the government allocated funds are spent, as long as they adhere to the the requirements of the South African School Amendment Act, no 84 of 1996, as amended: 2006 p28.

2.3.2.3 Data

NIDS is a nationally representative household and individual level panel that commenced in 2008, and is undertaken every two years. It is undertaken every two years with the most recent Wave 3[15] surveyed in 2012 (SALDRU, 2013). NIDS collects data on household composition and structure, education, labour market participation and economic activity, health, and agriculture, among other issues (Leibbrandt et al., 2009). These issues, are covered in four instruments, namely household, individual adult, individual proxy, and child questionnaires. In arriving at the households' sample, NIDS employed a stratified two-stage cluster sample design. The Wave 1 response rate was 69%, that is, 7,305 households and 28,225 individuals responded across South Africa (Leibbrandt et al., 2009). To take care of sample non-response and representativity, NIDS has both design weights and post-stratification weights (wittenberg, 2009), however, a panel weight for the three waves is currently unavailable. Wave 2 was conducted between 2010 and 2011. The attrition rate was 19% from Wave 1, excluding those who died or moved out of scope (Brown et al., 2012). Wave 3 was administered in 2012, and had a 16% non-response rate from Wave 2. In Wave 2 the sample was increased, and individuals failed to be interviewed in Wave 1 were traced leading to an overall increase in the sample by 6800. We use Wave 1, 2008, to study the effects of pupil-teacher ratio and expenditure per pupil on education attainment. To ensure that we can make inferences from the African sample to the African population in South Africa we use post-stratification weights in the analysis.

In post-apartheid South Africa, there remain differences between the different population groups.

[15] Data collection for Wave 4 is currently underway.

Our sample of interest in NIDS is African[16] respondents. Given the legacy of apartheid that differentiated education provision by race, a pooled analysis of all population groups is likely to mask the effects of school inputs on African's educational attainment. We therefore focus on the sample of African in the NIDS 2008 wave 1. Most studies on school quality consider data on children of school going age, which ensures that the socioeconomic and neighbourhood characteristics are a perfect match for this group of individuals. In this study, we consider 15-30 year olds who were still living with their parents in 2008. We assume that these individuals grew up and schooled in these neighbourhoods, and that the socio-economic characteristics at the time of the survey are valid for this group. In South Africa, schooling is compulsory until age 15 or grade 9, whichever comes first. The South African National Youth Policy defines the youth as individuals aged between 14 to 35 years. Thus our sample focuses on individual who are past the legal age of minimum schooling, but are still considered youth. The NIDS adult questionnaire focuses on the education of 15 to 30 year olds. Thus we truncate our selection at 30 years. Further it is more probable that the family background and neighbourhood characteristics of this group in NIDS are accurate. This is because this group is more likely to be residing at their parents home. Ideally, a study of determinants of educational attainment ought to account for final level of education attained and relate it to characteristics of the environment . As a result, we need to ensure that in our sample are respondents who are most likely to have been raised in a given household and are dependents of the household head. We therefore further restrict our sample to respondents who are dependent children or guardians of the household head. It is likely that some of the respondents changed neighbourhoods, which could introduce bias in our estimates. Our estimation sample is therefore made up of 2992 young African individuals who are either not enrolled (1337) or enrolled in institutions of learning (1624).

The education outcome variable we consider is the highest level of education of the individual i. Respondents in NIDS were asked to give their highest level of education. We use this as a measure of actual educational attainment, and use it to calculate a relative educational attainment index. In the analysis, we also include a least square estimation of the years of schooling. We define a year of schooling to be equivalent to one grade or one level of education attained. This is likely to be lower than the actual years of schooling since we do not take into account grade repetition or studying on a part-time basis. To deal with some of these issues, we use a relative education attained index. Other variables necessary for this study, such as individual, household and neighbourhood characteristics, are available in NIDS. Variables on school characteristics are mainly from the NIDS school administrative data that has been created for NIDS from government records on schooling. This was through a rigorous process of matching respondents schools with the schools in the Department of Education (DoE) register (see SALDRU, 2012). The administration data has information on pupil-teacher ratio, school quintile, type of school per apartheid classification, and distance to school. Table 12 in the appendix gives a description of the variables used.

[16] As of census 2011, Africans were 79.2% of the South African population. In NIDS our sample of individuals who are 15-30 year old and are dependents, Africans are over represented at 90%.

We derive the expenditure per learner variable by matching respondents' school quintile variable in the NIDS administration data to the amount of expenditure per learner[17] allocated by the government. Currently, the government allocates expenditure per learner using a school quintile rule. In the past, allocation differed by area, province, and race. Before 1994, schools were categorised by race and region (largely the Republic of SA and the 10 "homelands"). The geographical and racial differentiation in education funding lasted until the end of apartheid. Since the School Act of 1996 funding has been allocated using the school quintile rule. The school type variable in the data indicates the schools as categorised in apartheid education departments. The school quintile variable indicates school quintile in Wave 1 for 2005, 2007, and 2008. The school quintile variable is limited in use in this study because some of the respondents in our sample of 15 to 30 year old's went to school before 2005 when schools were not categorised by school quintile. It is, however, feasible to relate the school quintile to school type per apartheid classification. Most formerly Black schools are in quintile 1, 2, or 3 while most formerly White schools are in either quintile 4 or 5. In our consideration of the effect of expenditure per pupil, we expect the effect of an increase in allocation per pupil to have a different effect in formerly Black schools than in formerly White schools. Most formerly Black schools are still catching up in putting in place resources such as textbooks, libraries, and classrooms. Expenditure per pupil is as gazetted by the government at the national level. The provincial governments are allowed to change these allocations marginally, using certain rules to suit their school's exact needs. This is likely to be the case in some provinces, which means the allocations we used may differ from the actual allocation. However, we opt to use the national figure since we could not identify the rules considered by each province. We expect expenditure per pupil to have a positive effect on education attainment, and to have nonlinear effects. We use the log expenditure per pupil because we would expect the effect of more expenditure to initially increase at an increasing rate then level off, and also for scaling purposes.

Pupil-teacher ratio is available in the NIDS school administration data. This ratio, for both enrolled and achieved sample, is the total number of pupils in a given school in 2005, 2007, and 2008 divided by the total number of teachers, as provided by the Department of Basic Education. A major limitation is that we are assuming that the ratio was the same in these schools in the years between 1998 and 2005 which is unlikely (our sample's oldest respondents were last in school in 1998). This is likely to understate the ratio for those earlier years when the ratio was likely to be higher. The pupil-teacher ratio includes teachers who teach and those who are involved in administration of the school, so the ratio is likely to be understated. We do not control for the quality of teachers, and it is possible that if the overall teacher quality in the country was low(high) the positive coefficient would be reduced (increased). For instance, Bhorat and Oosthuizen (2008) controls for teacher quality using a proxy and finds pupil-teacher ratio to be insignificant. Although we control for other school characteristics, the possibility of bias in our estimate of pupil-teacher ratio cannot be ruled out. We expect pupil-teacher ratio to have a negative effect on educational

[17] Figures extracted from the South African Schools Act, Amended national norms and standards for school funding of 2006 p34 and Hall and Giese, 2008 p39

attainment. In the regressions, we opt to use a log of the pupil-teacher ratio rather than the ratio itself. This is mainly because the logged ratio gives invariant results whether we use teacher-pupil ratio or pupil-teacher ratio. We also control for distance to school.

We include gender and age as individual characteristics, and also the squared term of age to capture nonlinear effect of age on educational attainment. Family characteristics include parent's[18] highest education attained which we consider as a proxy for both the household well-being and parental taste for education. Other family characteristics included are home language and location, number of siblings, and household monthly per capita income.

Neighbourhood characteristics are derived by grouping responses by cluster, calculating the average in each cluster[19], and then assigning each individual their corresponding cluster average value. In total, we consider 10 neighbourhood characteristics: Percentage of households with electricity, percentage of households whose refuse or rubbish is removed at least once a week by the local authority, percentage of households with street lighting, percentage of households with access to piped water, average household size in the neighbourhood, percentage of individuals employed in the neighbourhood, percentage of households with ownership of their dwelling, average number of rooms (excluding bathrooms and toilets) in the neighbourhood, percentage of households that share a toilet, and percentage of households with a land-line telephone connection in the neighbourhood. We use principal component analysis[20] to reduce these neighbourhood characteristics to three characteristics, as indicated in the descriptive statistics section discussed next.

2.4 Descriptive statistics

2.4.1 Characteristics of the sample

We present the summary statistics of education attainment, the outcome variable, in Table 2.1. The table shows the frequency of Africans who: *(i)* are not enrolled in an educational institution. Respondents not enrolled have finished their education (general, some secondary, completed secondary (matriculated), or higher education, that is a further education and training certificate or any other certificate, diploma, or degree). Respondents who are enrolled in an educational institution are studying for a general education (which includes primary level), at grade 10 or 11, or at grade 12, or at an FET college or a higher education institution. The majority (73.23%) are enrolled or have attained either a general or some secondary education. More individuals have a matriculation or an FET certificate or a higher level of education than are currently enrolled for these levels.

[18]Most respondents gave either mother's or father's education, we combined this responses to get a response for parent's education level. Using either mother's or father's education level in the analysis gave similar results.

[19]In NIDS a cluster(also referred to as a Primary Sampling Unit (PSU)) consist of one enumeration area(EA) and is the smallest portion of land demarcated for census in South Africa (Leibrandt *et al*, 2009 p9).

[20]We got similar results when we included factors from factor analysis rather than components from PCA. However, one of the two factors was insignificant.

However, the number enrolled for primary or some secondary education is higher than those who have attained this level of education. From this we could infer that more are likely to complete at least secondary education (matriculate) in the future.

Table 2.1: Education attainment by enrollment

Highest education level attained	Frequency(Percentage)					
	Enrolled		Not enrolled/Achieved		Total	
General	798	(48.14)	308	(23.04)	1110	(37.10)
Some secondary	616	(38.93)	445	(33.28)	1081	(36.13)
Matriculation	182	(11.21)	472	(35.30)	663	(22.16)
FET & Higher	28	(1.72)	112	(8.38)	138	(4.61)
Sample Size	1624	(100)	1337	(100)	2992	(100)

Source: Own calculation from NIDS Wave 1 conducted in 2008. Note: Higher education includes those with a certificate, a diploma or a degree.

Table 2.2 shows the mean of individual, household, school, and neighbourhood characteristics for our estimation sample. In the table, we present the distributions of these statistics by highest education attained, and the full sample. The statistics for individuals enrolled at a given level are presented in Table 2.9 in the appendix. As shown in Table 2.2, the pupil-teacher ratio averages between 30 and 32 pupils per teacher. Notably, the general level has the highest average at 32, while tertiary (a FET or higher) level has the lowest average at 30. This is lower than the average of 40 in the early 1990s (Case and Deaton, 1999). This average reflects the national average which has been less than 35 in the last 10 years (Department of Basic Education, 2010b p50). The ratio in the sample is much more dispersed for those who have attained a primary education than for those who have attained a secondary education. A clearer picture of the ratio's distribution by level of education attained for those who are not enrolled is presented in figure 2.1. The average expenditure per pupil[21] for those enrolled is almost twice the average for those already completed. The difference is evident in the densities presented in figure 2.2 (see appendix Figure 2.5 and 2.6 for distributions of enrolled individuals).

As shown in Table 2.2, the sample is of Africans in their late teens to mid-20s, on average. The average age difference between those enrolled and those who have completed is 5 years. Those enrolled at the general level and at grade 12 are older for their age by 2 and 1 year respectively (This is based on a grade 1 enrollment age of 7.). The sample is predominantly female with the exception of those enrolled, and who have attained a general level of education at 44%. The majority of the parents have on average no formal education while a relatively small proportion have at least some secondary education. Notably, 10.9% and 18.7% of individuals who have attained a matriculation and higher education respectively have parents with at least some secondary education. One can

[21] We use CPI to deflate the values and take 2008 as the base year.

infer from this that parental education does matter in their children's educational attainment. On average respondents have 3 siblings.

Table 2.2: Mean characteristics of the sample by education outcome

Variable name	Not enrolled /Achieved				Both enrolled and achieved			
	General	Some secondary	Matriculation	FET & Higher	General	Some secondary	Matriculation	FET & Higher
School inputs								
Log(pupil-teacher ratio)	32.07	31.20	30.89	30.23	32.21	31.55	30.86	30.15
	(0.558)	(0.494)	(0.361)	(0.874)	(0.444)	(0.356)	(0.290)	(0.718)
Log(expenditure per pupil)	275.6	283.9	297.6	150.6	450.0	402.5	342.0	164.9
	(16.38)	(14.20)	(12.65)	(19.02)	(13.30)	(11.14)	(11.68)	(18.61)
Age	21.63	23.37	22.55	24.39	18.24	20.63	21.58	23.96
	(0.299)	(0.190)	(0.214)	(0.417)	(0.160)	(0.146)	(0.168)	(0.391)
Gender (Female=1)	0.439	0.685	0.584	0.654	0.440	0.603	0.559	0.644
	(0.040)	(0.032)	(0.031)	(0.068)	(0.021)	(0.021)	(0.025)	(0.057)
Parents education (%): No schooling	0.661	0.552	0.501	0.513	0.577	0.556	0.497	0.528
	(0.033)	(0.041)	(0.033)	(0.069)	(0.022)	(0.023)	(0.028)	(0.058)
General	0.318	0.380	0.390	0.300	0.353	0.354	0.363	0.318
	(0.033)	(0.045)	(0.029)	(0.055)	(0.021)	(0.025)	(0.024)	(0.049)
At least some secondary	0.0201	0.0689	0.109	0.187	0.0699	0.0899	0.140	0.154
	(0.009)	(0.020)	(0.019)	(0.059)	(0.010)	(0.014)	(0.021)	(0.045)
Number of siblings	2.781	2.595	2.694	2.448	2.768	2.662	2.504	2.543
	(0.178)	(0.188)	(0.174)	(0.234)	(0.112)	(0.124)	(0.152)	(0.209)
Household income per capita(monthly)	486.8	645.2	767.9	1134.8	591.4	646.9	954.7	1273.1
	(76.10)	(93.87)	(80.58)	(167.9)	(46.50)	(53.89)	(109.51)	(158.45)
Home location (%)n: Rural/Traditional	0.572	0.496	0.492	0.372	0.604	0.537	0.451	0.363
	(0.057)	(0.057)	(0.053)	(0.084)	(0.047)	(0.045)	(0.047)	(0.075)
Urban	0.371	0.399	0.452	0.605	0.350	0.386	0.504	0.608
	(0.057)	(0.054)	(0.052)	(0.084)	(0.047)	(0.043)	(0.047)	(0.075)
Farm	0.0562	0.106	0.0558	0.0231	0.0459	0.0766	0.0444	0.0296
	(0.018)	(0.036)	(0.023)	(0.016)	(0.015)	(0.022)	(0.018)	(0.015)
School Type (%):								
Independent homelands	0.270	0.146	0.168	0.107	0.249	0.130	0.173	0.111
	(0.044)	(0.026)	(0.030)	(0.045)	(0.031)	(0.019)	(0.028)	(0.043)
Self-governing territory	0.347	0.393	0.383	0.403	0.379	0.395	0.336	0.343
	(0.048)	(0.039)	(0.043)	(0.080)	(0.034)	(0.030)	(0.037)	(0.065)
Department of education	0.303	0.351	0.330	0.367	0.271	0.332	0.358	0.373
	(0.044)	(0.039)	(0.042)	(0.073)	(0.032)	(0.034)	(0.039)	(0.062)
House of assembly, representative and delegates	0.0234	0.0330	0.0404	0.0488	0.0391	0.0485	0.0543	0.0972
	(0.010)	(0.011)	(0.013)	(0.022)	(0.009)	(0.010)	(0.012)	(0.028)
New schools	0.0577	0.0765	0.0786	0.0743	0.0617	0.0935	0.0787	0.0759
	(0.017)	(0.020)	(0.019)	(0.029)	(0.012)	(0.019)	(0.017)	(0.028)
Distance to school(km)	82.87	40.73	64.51	58.15	38.91	48.01	54.91	54.32
	(19.84)	(6.82)	(22.94)	(21.02)	(7.22)	(18.04)	(17.26)	(16.94)
Neighbourhood characteristics: Number of rooms per person	4.240	4.567	4.625	4.587	4.494	4.657	4.532	4.613
	(0.164)	(0.167)	(0.162)	(0.222)	(0.117)	(0.132)	(0.139)	(0.196)
% with electricity	0.725	0.798	0.817	0.854	0.727	0.801	0.839	0.876
	(0.045)	(0.029)	(0.029)	(0.030)	(0.035)	(0.024)	(0.023)	(0.024)
% shared toilet	0.157	0.214	0.199	0.255	0.150	0.192	0.203	0.254
	(0.026)	(0.032)	(0.024)	(0.043)	(0.023)	(0.025)	(0.023)	(0.039)
% refuse collected	0.313	0.395	0.419	0.551	0.306	0.369	0.472	0.566
	(0.050)	(0.052)	(0.046)	(0.068)	(0.043)	(0.042)	(0.043)	(0.062)
% street light	0.310	0.353	0.400	0.502	0.286	0.339	0.439	0.519
	(0.046)	(0.045)	(0.046)	(0.067)	(0.036)	(0.038)	(0.042)	(0.061)
Average household members	6.262	6.113	6.290	5.882	6.175	6.044	6.030	5.877
	(0.274)	(0.322)	(0.286)	(0.450)	(0.179)	(0.246)	(0.239)	(0.363)
% employed	0.174	0.204	0.192	0.244	0.148	0.182	0.195	0.244
	(0.012)	(0.017)	(0.011)	(0.022)	(0.009)	(0.011)	(0.011)	(0.019)
% with piped water	0.848	0.869	0.860	0.915	0.806	0.844	0.871	0.922
	(0.032)	(0.023)	(0.025)	(0.025)	(0.031)	(0.025)	(0.021)	(0.021)
% member own house	0.871	0.857	0.874	0.848	0.892	0.872	0.864	0.834
	(0.016)	(0.021)	(0.014)	(0.030)	(0.010)	(0.016)	(0.013)	(0.027)
% have landline telephone	0.0847	0.0865	0.132	0.168	0.0820	0.103	0.148	0.179
	(0.013)	(0.012)	(0.021)	(0.028)	(0.009)	(0.013)	(0.021)	(0.025)
Sample size	**1337**	**1337**	**1337**	**1337**	**2992**	**2992**	**2992**	**2992**

Note: *(1)* Standard errors in parenthesis. *(2)*Pupil-teacher ratio is as provided by the Department of Basic Education in the NIDS administration data. *(3)* Expenditure per is learner derived from matching respondent's school quintile variable in the NIDS administration data to the amount of expenditure allocated by the government to each student using school quintile every year. *(4)*For all the neighbourhood characteristics the variables are calculated by (a) calculating the average value in each cluster and (b) assigning each individual a value corresponding to his/her cluster. *(5)* The sample is of African respondents aged 15-30 years. *(6)* See Appendix for definitions of all variables used. *(7)* Design weights are used.

Source: Author's calculation based on NIDS Wave 1 (2008).

The average household monthly income per capita in the sample increases with the education level enrolled in or completed. The average income is lowest for individuals who are enrolled in /have completed a general level of education, and is highest for individuals enrolled in /completed a higher education qualification. There is almost a threefold difference in the average household income between the two groups. We could infer from this that household income plays an important role in educational attainment. The majority of individuals with primary and secondary education live in rural areas, while those with higher education live in urban areas. Most individuals attend self-governing schools, and those with primary education live furthest from home. The average individual in the sample lives in a neighbourhood where the average number of rooms per person[22] is 4. The average household size is 6, and almost all individuals live in a house owned by a household member. The majority live in households with access to piped water, and electricity[23], but most have no street lighting. Fewer individuals live in areas where refuse is collected, share a toilet, and have a land line telephone connection. Most respondents live in areas with high unemployment.

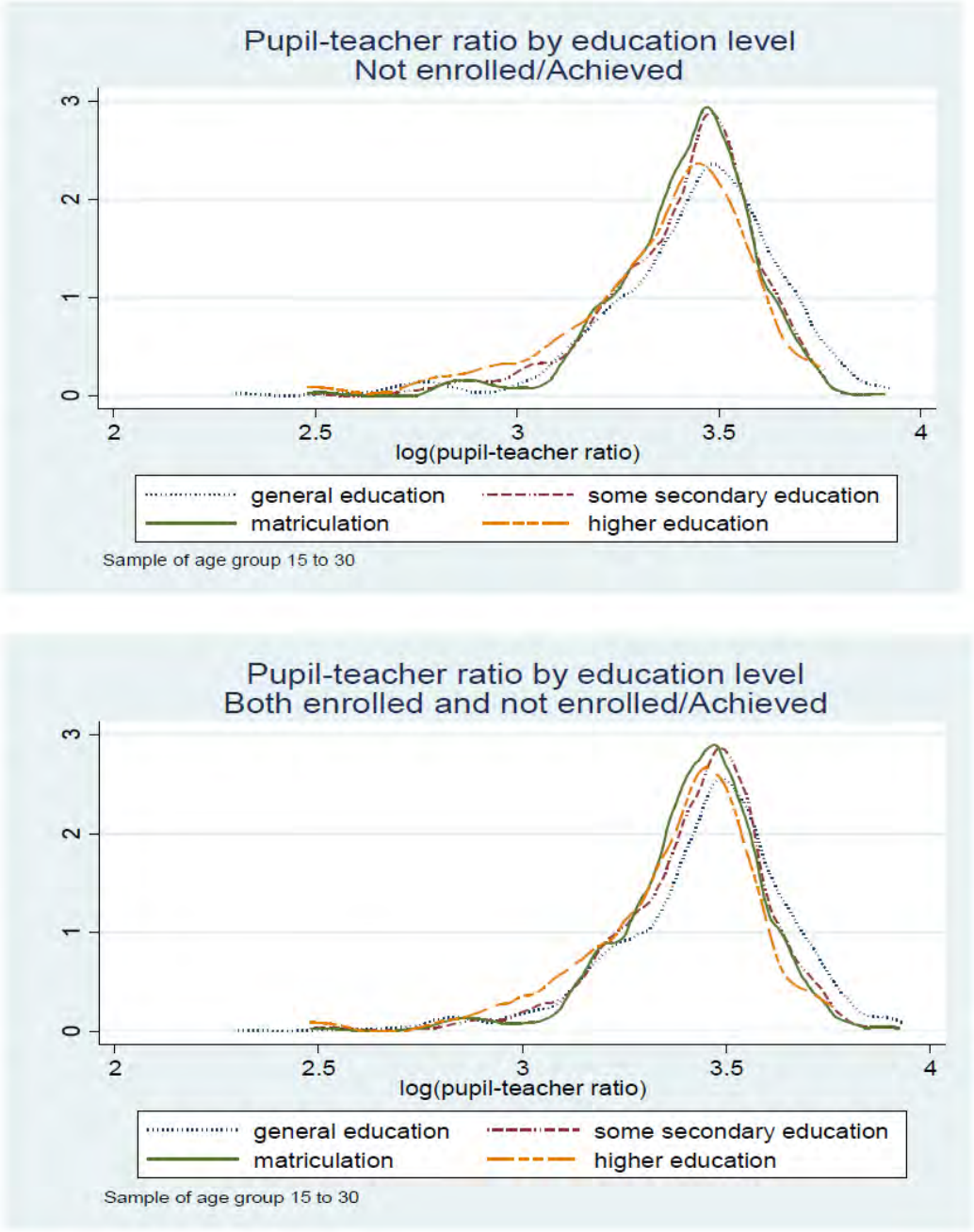
2.4.2 Distributions of school inputs

We present the kernel densities of the two school inputs by education level attained. In each figure, the distributions of the achieved and the pulled sample (both achieved and enrolled) are presented, in that order. First, we present the densities of pupil-teacher ratio followed by the densities of expenditure per pupil. Figure 2.1 shows that primary education is more skewed to the right, has lower peaks, and has longer tails than the other levels. On the other hand, higher education is skewed to the left, and has the shortest tail. An interpretation of these distributions is that lower pupil-teacher ratios are associated with being enrolled for or attaining a higher education qualification. Higher ratios are associated with being enrolled for or attaining a primary education. We conduct a Kolmogorov-Smirnov test to determine whether individuals who are enrolled at institutions of learning, and those who have achieved a given level of education are from a population with identical distribution functions. The results for pupil-teacher ratio show that the individuals who are enrolled and individuals who have achieved primary, some secondary, matriculation, higher education are from populations with identical distribution functions. This means that at all levels of education there is no significant difference between the ratio for those enrolled and for those who have achieved as is evident in the pooled sample densities.

[22] This number excludes bathrooms and toilets.

[23] This includes households with disconnected electricity.

Figure 2.1: Distribution of the logarithmic pupil-teacher ratio for Africans



Source: Author's calculation based on NIDS Wave 1 (2008).

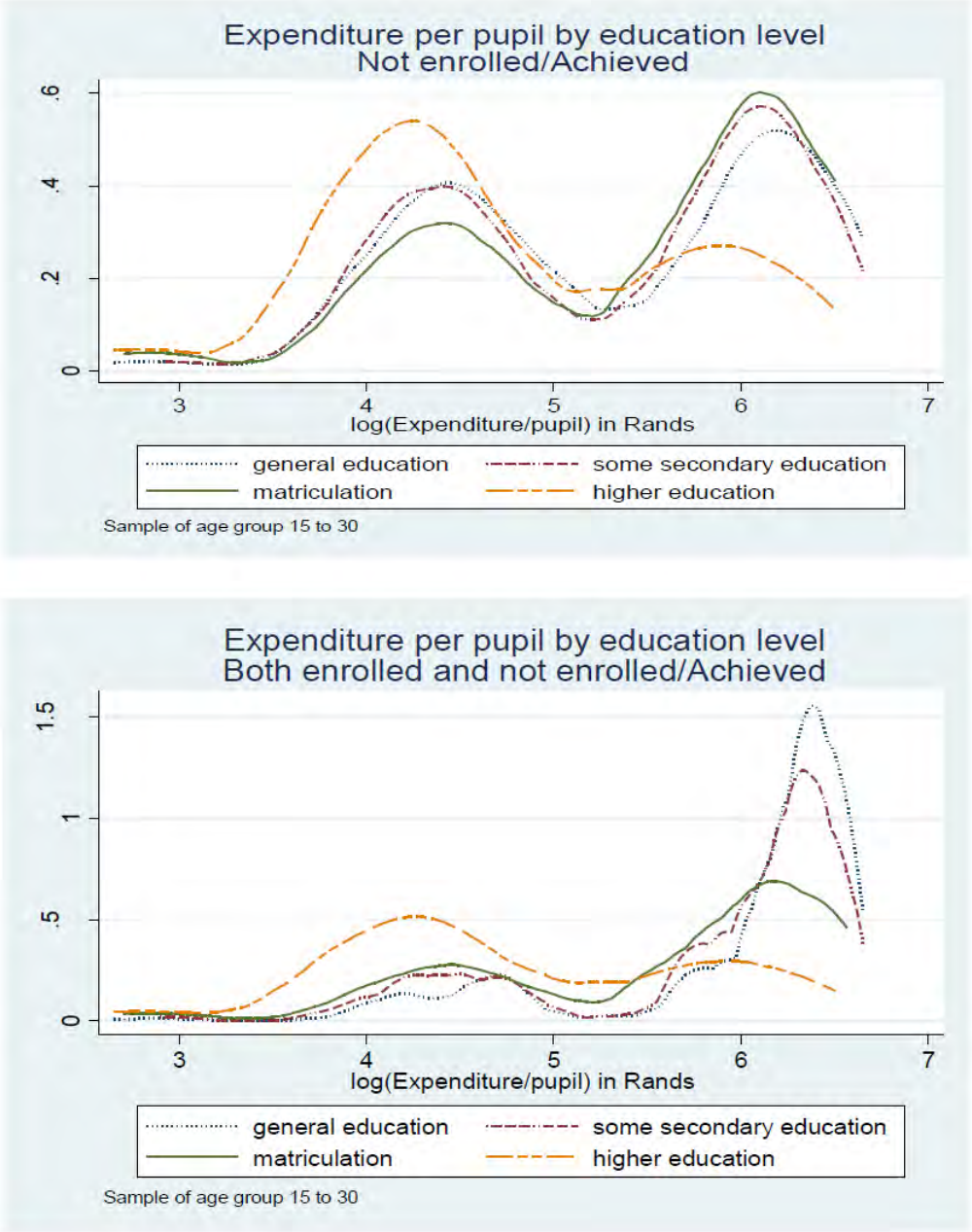
Figure 2.1 shows the distributions of expenditure per pupil for those enrolled at all levels of education are bimodal. However, the distribution for those enrolled in FET colleges or in higher education institutions is rather flat, which could be an indication of insensitivity to, or independence of, expenditure per pupil in an individual's higher education enrollment. A notable feature of the distributions for those who have achieved a given level is the change of peaks between the different

education levels as expenditure per pupil increases. At lower values of expenditure per pupil, higher education has the highest peak, while secondary education has the lowest peak. At higher values of expenditure per pupil, more individuals have achieved a secondary education, while fewer have attained a higher education. A simple interpretation of this is that increased expenditure per pupil has improved the attainment of general education through to matriculation but has not improved higher education attainment.

The full sample statistics, from Table 2.2 and Figure 2.2, indicate all distributions are bimodal, but both primary and secondary are more skewed to the right while higher education is skewed to the left. The skewness is, however, not surprising because the average age of those who have attained a given level of education is higher by about 3 years relative to those who are still enrolled. The proportion of those who have attained a higher education qualification are 6 percentage points more than those still enrolled in higher education institutions. These differences in the distributions between enrollment and attainment are therefore more likely driven by time. The bimodal distribution is likely to be a result of the allocation pattern where quintile 1, 2, and 3 schools have progressively been allocated higher expenditure per pupil than schools in quintile 4 and 5. Schools in quintile 1 and 2 were categorised as non-fee paying schools in 2007/08 and in 2012 schools in quintile 3 also became non-paying schools. A significant proportion of Africans attend schools in quintile 1, 2, and 3. Schools in quintile 4 and 5 are mainly 'model C' schools. These receive substantially less allocation per pupil at 15% their non-personnel and non-capital costs, but have had a higher success rate at matriculation and enrollment into higher education. A simple comparison of these distributions tells us that, firstly there is no significant difference in expenditure per pupil between those who are enrolled or have attained a primary or some secondary education. Secondly, respondents who have attained a higher education qualification show a lower expenditure per pupil than those who are currently enrolled. Thirdly, individuals enrolled in grade 12 or who have completed grade 12 attended schools with higher expenditure per pupil. This could be an indication that increasing expenditure per pupil improves educational attainment.

A Kolmogorov-Smirnov test was undertaken to show the distributions of expenditure per pupil. Individuals who are enrolled for /achieved primary/some secondary/matriculation are at 1%-level, and for higher education at 10% from a population with different distribution functions. This means that there is a significant difference between the expenditure per pupil for those enrolled and those who have achieved across the different levels of education. At all levels of education those who have attained are consistently from a lower expenditure per pupil distribution than those who are still enrolled in institutions of learning. This could be explained by firstly by increased annual allocations per pupil over the years. That is, the amount allocated to each quintile in 1998 was substantially lower than the amount allocated in 2008. Secondly, most students who proceed to higher education levels are those who attend quintiles 4 and 5, and these two quintiles receive the lowest allocations per learner.

Figure 2.2: Distribution of the logarithmic expenditure per pupil for Africans



Source: Author's calculation based on NIDS Wave 1 (2008).

2.4.3 Principle component analysis of the neighbourhood characteristics

Most of the neighbourhood characteristics examined are related. We therefore consider it optimal to consolidate these characteristics into fewer categories for the regression analysis. We opt to use principal component analysis to identify and compute composite scores for these neighbourhood characteristics. Our analysis narrows these down to three components (see appendix Table 2.10 for PCA results). Component (i) is Community economic status. This comprises average house-

hold size, proportion of households member-owned dwellings, neighbourhood employment level, and proportion of households that share a toilet. Component (ii) is Serviced by local municipality. This comprises proportion of households with electricity, street-lighting, piped water, and weekly municipal refuse collection. Component (iii) is Population density and housing quality. This comprises average number of rooms per household, and proportion of households that share a toilet, and proportion of households with electricity. In table 2.3 we present the mean characteristics of these components.

Table 2.3: Mean components

Variable name	Not enrolled/Achieved				Both enrolled and achieved			
	General	Some secondary	Matriculation	FET & Higher	General	Some secondary	Matriculation	FET & Higher
Serviced by municipality	0.753	0.492	-0.208	-0.402	0.685	0.410	0.188	-0.223
	(0.114)	(0.138)	(0.172)	(0.349)	(0.112)	(0.151)	(0.138)	(0.248)
Community economic status	-0.523	-0.222	0.562	0.748	-0.472	-0.181	0.163	0.488
	(0.161)	(0.134)	(0.154)	(0.255)	(0.155)	(0.124)	(0.130)	(0.169)
Population density and housing quality	-0.277	-0.127	0.00411	0.244	-0.322	-0.204	-0.0615	0.0713
	(0.058)	(0.085)	(0.192)	(0.273)	(0.053)	(0.080)	(0.106)	(0.143)
Sample size	**1337**	**1337**	**1337**	**1337**	**2992**	**2992**	**2992**	**2992**

Note: *(1)* Standard errors in parenthesis. *(2)* The sample is of African respondents aged 15-30 years. *(3)* See Appendix for definitions of all neighbourhood characteristics used to derive the components. *(4)* Design weights are used.

Source: Author's calculation based on NIDS Wave 1 (2008).

From Table 2.3 we see that fewer individuals enrolled for/have attained a primary or grade 10 or 11 education live in neighbourhoods serviced by the municipality. Most individuals enrolled/attained a higher education qualification live in neighbourhoods serviced by a municipality. The majority of individuals who enrolled/achieved primary education live in communities with better economic status. More individuals who enrolled/achieved a higher education qualification live in neighbourhoods of high density and with lower quality housing.

2.5 Empirical results

In this section, using the reduced form equation 2.6, we estimate the relationship between the school inputs (pupil-teacher ratio and expenditure per pupil) and education level attained. We consider this relationship when other determinants of educational attainment are included. We present the estimated parameters under several model specifications. In the analysis of educational attainment, we use a partial generalized ordered probit to analyse actual education level attained from a sample of Africans who have achieved a given level of education. That is, those who are not enrolled in any educational institution. We also use an ordinary least square (OLS) approach to analyse a relative educational attainment index on a combined sample of Africans who are either enrolled for or have achieved a given level of education in 2008. The findings from these analyses are presented in the following sub-sections.

2.5.1 Educational attainment

2.5.1.1 School inputs effects

Table 4 displays the estimated coefficients on logarithmic pupil-teacher ratio, logarithmic expenditure per pupil, and their quadratics from the partial generalised ordered probit model. In the model, three different parameter vectors θ_1, θ_2, and θ_3 are estimated. The vectors give a series of estimated probit analyses as follows: θ_1 which presents a probit analysis of having a primary education versus having either some secondary, matriculation, or higher education; θ_2 which presents a probit analysis of having a general or some secondary education versus having matriculation or a tertiary education qualification; and θ_3 which presents a probit analysis of having either a general or some secondary education, or matriculation versus having a tertiary education qualification. The results are from our full specification which, in addition to the school inputs, controls for age, age squared, gender, parent's education, number of siblings, real household monthly per capita income, home location, classification of school attended, distance to school, and neighbourhood characteristics. Results for the rest of the controls are presented in table 2.11 in the appendix.

Table 2.4: Coefficients of school inputs from the partial generalized ordered probit

	Goprobit		
Dependent variable: Education level attained	**θ_1**	**θ_2**	**θ_3**
Log(pupil-teacher ratio)	-0.622**	-0.622**	-0.622**
	(0.298)	(0.298)	(0.298)
Log(pupil-teacher ratio^2)	-1.724***	-1.212**	-0.286
	(0.663)	(0.586)	(0.599)
Log(expenditure per pupil)	0.367***	0.104	-0.291***
	(0.0565)	(0.0666)	(0.109)
Log(expenditure per pupil^2)	0.132**	0.0730	-0.0766
	(0.0647)	(0.0568)	(0.0700)
Sample size	1337	1337	1337

Note: (1) Standard errors in parenthesis. (2) The * indicates level of significance: * p<0.1, ** p<0.05, and *** p<0.01. (3) control variables in the full specification include age, age squared, gender, parents education dummies, number of siblings, log of household monthly income per capita, school type per apartheid classification dummies, distance to school, home location, and neighbourhood characteristics components.(4) The sample is of African respondents aged 15-30 years. (5) Both expenditure per pupil and household income are in real Rands, deflated by CPI with 2008 as base year. (6) See Appendix for definitions of all variables used. (7) Post-stratification weights are used.

Source: Author's calculation based on NIDS Wave 1 (2008).

The coefficients on the logarithmic pupil-teacher ratio indicate that an increase in the ratio decreases the likelihood of attaining a general education or higher, holding all other characteristics constant. These coefficients are identical in magnitude and level of significance at all the levels of education considered. However, the coefficients of quadratic ratio vary in magnitude and significance, an indication that the ratio, as included in our model, has heterogeneous effects on attainment of each level of education. That is, it violates the parallel lines regression assumption. These results suggest that a small ratio increases the probability of educational attainment. The coefficients on logarithmic expenditure per pupil have mixed signs. This is an indication that the variable has a

Table 2.5: Marginal probabilities of pupil-teacher ratio at representative values

Dependent variable: Education level attained	General	Some secondary	Matriculation	FET & Higher
Pupil-teacher ratio by sample distribution				
Predicted probability at means	0.185***	0.365***	0.381***	0.0691***
	(0.0143)	(0.0197)	(0.0187)	(0.0102)
Log(21) (5 percentile)	0.132***	0.115*	-0.134***	-0.113
	(0.0485)	(0.0683)	(0.0493)	(0.0686)
Log(28) (25 percentile)	0.156**	0.0914*	-0.157**	-0.0907*
	(0.0714)	(0.0489)	(0.0737)	(0.0470)
Log(31) (50 percentile)	0.168**	0.0775**	-0.165**	-0.0808**
	(0.0827)	(0.0360)	(0.0813)	(0.0376)
Log(38) (75 percentile)	0.185*	0.0555***	-0.173**	-0.0677***
	(0.0982)	(0.0161)	(0.0869)	(0.0258)
Log(43) (95 percentile)	0.196*	0.0384***	-0.175**	-0.0591***
	(0.108)	(0.0105)	(0.0872)	(0.0187)

Note: *(1)* Standard errors in parenthesis. *(2)* The * indicates level of significance: * p<0.1, ** p<0.05, and *** p<0.01. *(3)* The effect of pupil-teacher ratio is nonlinear at primary and matriculation education levels. *(4)* Control variables in the full specification include age, age squared, gender, parent's education dummies, number of siblings, log of household monthly income per capita, school type per apartheid classification dummies, distance to school, home location, and neighbourhood characteristics components. *(5)* The sample is of African respondents aged 15-30 years. *(6)* Both expenditure per pupil and household income are in real Rands, deflated by CPI with 2008 as base year. (7) See Appendix for definitions of all variables used. *(8)* Post-stratification weights are used.

Source: Author's calculation based on NIDS Wave 1 (2008).

heterogeneous effect on educational attainment. More expenditure per pupil increases the probability of attaining a general education, as indicated by the positive and significant θ_1, has no effect on the probability of attaining a general or some secondary education, as indicated by an insignificant θ_2, and significantly decreases the probability of attaining a matriculation or general level of education as indicated by the coefficient of θ_3. Expenditure per pupil has non-linear effects only on the likelihood of attaining a primary education or lower.

In a bid to gain a better interpretation of the effects of logarithmic pupil-teacher ratio and logarithmic expenditure per pupil on education level attained, we calculate marginal probability effects at representative values conditional on the mean of the controls. The nonlinear relation of logarithmic pupil-teacher ratio with itself, and similarly the nonlinear relation of logarithmic expenditure per pupil with itself motivate our choice to present marginal probability effects at representative values. The values at different points of both variables' distributions enable us to show how at these values these variables affect the probability of attainment of the different education levels. In table 2.5, we display the marginal probability effects of logarithmic pupil-teacher ratio at different percentile values in the sample's distribution.

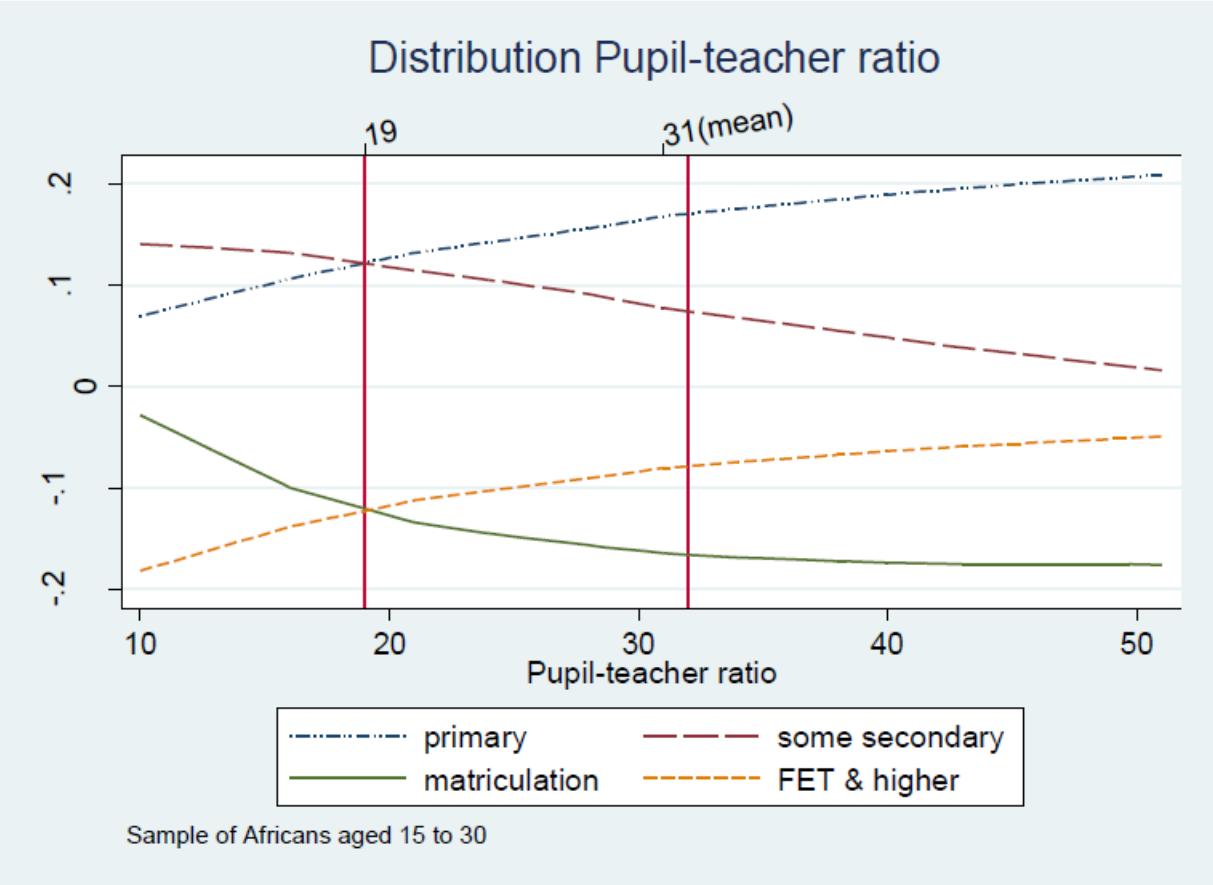
In table 2.5, we also present the predicted probabilities of attaining the four levels of education conditioned at the means of all covariates. The results show that the pupil-teacher ratio significantly raises the probability of attaining lower levels of education, but significantly lowers the probability of attaining higher levels of education. For example, for individuals who attend a school with a pupil-teacher ratio of 31 the probability of attaining a general education increases by 16.8 percent-

age points and by 7.8 percentage points for some secondary education, holding all other factors constant. The effect on general education is fairly small compared to the average predicted probability of attaining a primary education of 19%, but the effect on some secondary education is fairly large compared to the average predicted probability of attaining a matriculation of 38%. This is an indication that the ratio's role in secondary school attainment is substantial. However, the effect of the ratio turns negative for probability of attaining a tertiary education, which is an indication of declining importance of the ratio in educational attainment as education level increases.

The positive effects on general and some secondary education are consistent with empirical research that the ratio positively affects educational attainment (or test scores) at these levels of education (Angrist and Lavy, 1999; Wossmann, 2003). This could imply that the ratio is more important at lower levels of education than at higher levels. This is an indication that younger learners are more reliant on teachers and direct contact with a teacher improves the probability of achieving at these levels. The negative effects on matriculation and tertiary education could be an indication that the significance of the ratio in attainment declines with education level as other factors become more important. This is likely to be in support of the argument that in these years peer relations influence education aspiration and achievement and attitude towards schooling (Johnson, 1981). These negative effects indicate a 'lagged effect' of the large pupil-teacher ratio on school and higher education performance. We, however, do not find cumulative effects of the ratio, which is against the view that the benefits of small class size occur later (Krueger, 2003, p56). To give a picture of these results, in Figure 2.3 we present the marginal probability effects at more points of the ratio's distribution for the different levels of education .

Figure 2.3 displays the marginal probability effects at different points of the entire sample distribution of the logarithmic pupil-teacher ratio. The ratio has an increasing effect on the probability of attaining a general or some secondary education, but the magnitude of its effects on the probability of attaining a general education increases as the ratio increases, while the magnitude of its effect decreases for some secondary education as the ratio increases. On the other hand, the ratio has a decreasing effect on the probability of attaining grade 12 or tertiary education, but the magnitude of its effects on the probability of completing secondary education increases as the ratio increases, while the magnitude of its effects decreases for tertiary education as the ratio increases. Overall, the effect of the ratio declines as the level of education increases. This confirms the results depicted in Table 2.5 that the ratio is more important at lower levels of education, and has no cumulative effects.

Figure 2.3: Distributions of marginal probability of pupil-teacher ratio effects at representative values



Source: Author's calculation based on NIDS Wave 1 (2008).

In Table 2.6, we present the marginal probability effects of the logarithmic expenditure per pupil at selected values. The marginal probability effects presented are at the average values of each quintile in the sample, followed by the marginal probabilities at amounts allocated by school quintiles in 2008. The results from the two sets of quintiles indicate a similar pattern in effects, and we therefore discuss the results from the sample quintile (top panel). The results show that expenditure per pupil significantly raises the probability of attaining some secondary education or a matriculation, but lowers the probability of attaining a primary education or a tertiary education qualification. For example, expenditure per pupil of R255 raises the probability of attaining some secondary school or matriculation by 4.6 and 7.5 percentage points respectively, but reduces the probability of attaining a primary or tertiary education by 8.7 and 3.3 percentage points respectively, holding all other factors constant. The effects on both some secondary or complete secondary education are substantial compared to the predicted probability of attaining some secondary or completing secondary education (matriculation) of 37% and 38% respectively. On the other hand, the effects on either attaining a general or tertiary education are small compared to the predicted probability of 19% and 6% respectively. This indicates that the importance of the role of expenditure per pupil

in secondary education is substantial. The results indicate the effect of expenditure per pupil at all levels of education decreases as the amount of expenditure per pupil increases.

Across the levels of education, expenditure per pupil has first a decreasing effect on attainment of general education then an increasing effect on secondary education attainment, and then a decreasing effect on attainment of tertiary education. The negative effect at the general level of education, although unexpected, is similar to findings of studies by Nannyonjo (2007) and Du and Hu (2008) as cited in Glewwe et al. (2011). Glewwe and colleagues suggest that this could be an indication of compensatory funding, where too much money is spent on poor schools, or it could be an artifact of inclusion of other schooling characteristics. Given that more money is allocated to lower quintile schools, which are poorer, and we control for several schooling characteristics, these two arguments are instructive to our findings. These results suggest that there is a decreasing return on fiscal investment in education post secondary, a 'thinning out' of the impact and possible inefficiency in use of the funds. They also indicate that the secondary-tertiary education nexus is driven by factors other than expenditure on pupils by the state. In figure 2.4 we display the marginal probability effects at more points of the distribution of expenditure per pupil for the different levels of education.

Table 2.6: Marginal probabilities of expenditure per pupil at representative values

Dependent variable: Education level attained	General	Some secondary	Matriculation	FET & Higher
Expenditure per pupil by sample quintile average				
Predicted probability at means	0.185***	0.365***	0.381***	0.0691***
	(0.0143)	(0.0197)	(0.0187)	(0.0102)
Log(393) (Quintile 1)	-0.0739***	0.0323	0.0690**	-0.0274***
	(0.00907)	(0.0223)	(0.0275)	(0.00744)
Log(338) (Quintile 2)	-0.0785***	0.0369*	0.0709**	-0.0294***
	(0.0103)	(0.0218)	(0.0277)	(0.00871)
Log(258) (Quintile 3)	-0.0874***	0.0460**	0.0749***	-0.0334***
	(0.0128)	(0.0210)	(0.0285)	(0.0115)
Log(165) (Quintile 4)	-0.102***	0.0605***	0.0819***	-0.0406**
	(0.0170)	(0.0204)	(0.0304)	(0.0167)
Log(55) (Quintile 5)	-0.133***	0.0923***	0.102**	-0.0615*
	(0.0248)	(0.0214)	(0.0401)	(0.0329)
Expenditure per pupil by 2008 quintile allocation				
Log(775) (Quintile 1)	-0.0533***	0.0117	0.0608**	-0.0191***
	(0.00459)	(0.0249)	(0.0267)	(0.00319)
Log(711) (Quintile 2)	-0.0557***	0.0141	0.0617**	-0.0201***
	(0.00498)	(0.0246)	(0.0268)	(0.00354)
Log(581) (Quintile 3)	-0.0616***	0.0200	0.0640**	-0.0224***
	(0.00612)	(0.0238)	(0.0270)	(0.00459)
Log(388) (Quintile 4)	-0.0741***	0.0325	0.0691**	-0.0275***
	(0.00913)	(0.0223)	(0.0275)	(0.00751)
Log(129) (Quintile 5)	-0.110***	0.0684***	0.0860***	-0.0450**
	(0.0192)	(0.0204)	(0.0320)	(0.0200)

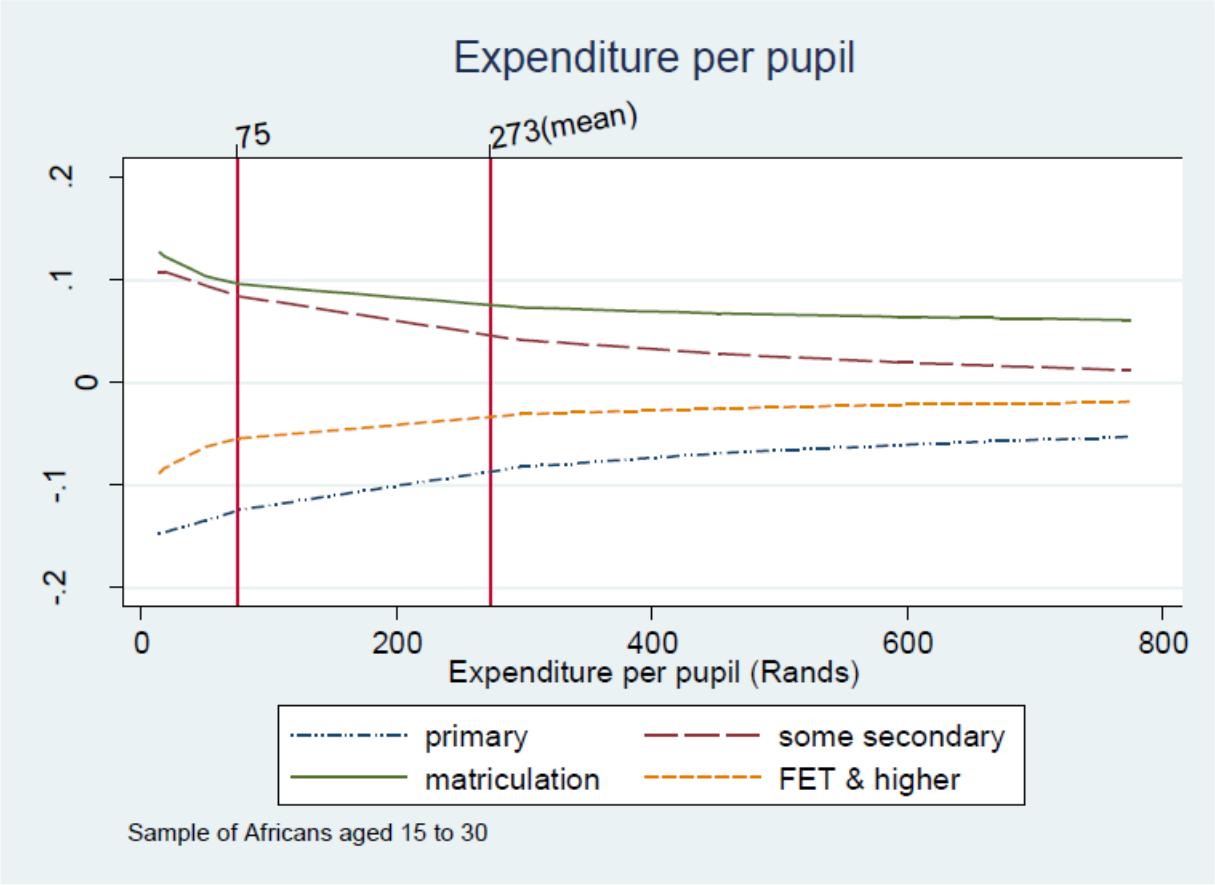
Note: *(1)* Standard errors in parenthesis. *(2)* The * indicates level of significance: * p<0.1, ** p<0.05, and *** p<0.01. *(3)* The effect of expenditure per pupil is nonlinear at primary and matriculation education levels. *(4)* Control variables in the full specification include age, age squared, gender, parent's education dummies, number of siblings, log of household monthly income per capita, school type per apartheid classification dummies, distance to school, home location, and neighbourhood characteristics components. *(5)* The sample is of African respondents aged 15-30 years. *(6)* Both expenditure per pupil and household income are in real Rands, deflated by CPI with 2008 as base year. (7) See Appendix for definitions of all variables used. *(8)* Post-stratification weights are used.

Source: Author's calculation based on NIDS Wave 1 (2008)

Figure 2.4 displays the marginal probabilities at different points of the entire sample distribution of the the logarithmic expenditure per pupil plotted. Expenditure per pupil has an increasing effect on the probability of attaining some secondary or complete secondary education, and the magnitude of its effects on the probability of attaining both levels of education decreases as the expenditure increases. On the other hand, expenditure per pupil has a decreasing effect on the probability of attaining a general or tertiary education. The magnitude of its effects on the probability of attaining both levels of education also decreases as the expenditure increases. This shows that educational attainment does not increase with expenditure per pupil, and is a likely indication of inefficiency in the funding allocations. This supports arguments that resources allocated to schools are 'spent in unproductive ways' (Hanushek, 1995), and that educational efficiency is not strongly correlated to expenditure (Heyneman, 2004). These results are in line with existing evidence on non-capital

non-personnel education expenditure that shows expenditure per pupil has positive and significant effects on attainment when used for items such as textbooks, workbooks, and school equipment (Hanushek, 1995; Glewwe et al., 2011; Gustafsson, 2007). These are essential items particularly in preparation for matriculation examination. Our results are in agreement with Van der Berg (2007) and Bhorat and Oosthuizen (2008) who find resources positively affect matriculation pass rate, and Yamauchi (2011) who finds increase in expenditure per pupil improves school quality in South Africa.

Figure 2.4: Distributions of marginal probability of expenditure per pupil effects at representative values



Source: Author's calculation based on NIDS Wave 1 (2008)

The results for both logarithmic pupil-teacher ratio and expenditure per pupil indicate that both school inputs significantly affect educational attainment. The pupil-teacher ratio is more important at lower levels of education, while expenditure per pupil increases the probability of attaining secondary education. These inputs are implemented at the general and secondary level of education, and therefore policies that ensure low pupil-teacher ratios and spending at these levels positively improve educational attainment at these levels, but do not necessarily have cumulative effects. The policy on expenditure per pupil plays a significant role. Most schools are still poor and still need

the funds to catch up on resources, but policies that would ensure efficient use of monies allocated per pupil are needed. For both pupil-teacher ratio and expenditure per pupil factors that are likely to be driving the differences in the effects at different levels of education attained include unobserved parental taste for education as indicated by the parental education, household income, and the type of school attended as defined by historical (apartheid) classification. The significance of the pupil-teacher ratio and expenditure per pupil in education level attained is expected to decline as parental education and household income increases because more educated parents are likely to help their children with homework, and ensure that their children attain a higher level of education. Similarly, households with a higher income are more likely to enroll their children in fee paying schools where the pupil-teacher ratio is likely to be lower since the schools could higher more teachers, and expenditure per pupil is generally less important for these schools since they already have substantial resources. Attending a historically White school is likely to reduce the significantly reduce the significance of government provided school inputs. This is mainly because these schools are on average more endowed in resources relative to historically Black schools which heavily rely on government employed teachers and funds. Students who attend historically White schools have been shown to have a higher grade 12 (matriculation) pass rate, and are therefore more likely to attain a higher education level relative to students who attend historically Black schools. The effects of these factors on educational attainment are discussed in detail in the next sub-section.

2.5.1.2 Effects of individual, family and neighborhood characteristics

In table 2.7, we present the marginal effects of all controls at their mean. This shows the partial generalised ordered probit in the first 3 columns, and two least square estimations of the relative education attainment index and years of schooling in the last 2 columns. In all models, in addition to school inputs, we control for individual, family and neighbourhood characteristics which are deemed important determinants of educational attainment. In this section, we discuss results for the partial generalised ordered probit analysis. The results show that age is an insignificant determinant of education level attained, however, it has the 'wrong' sign and has significant nonlinear effects. This is likely to be due to the small variation in age, a likely result of the sub-sampling by age of 15-30 year olds. We find that girls and women are less likely to attain a general education, but more likely to attain some secondary education, holding all other controls constant. However, we cannot reject the hypothesis that the probability of attaining a matriculation or a higher level of education is the same for both women and men.

Family characteristics are important determinants of educational attainment. We consider parental education both as an indicator of parental taste for education and as a control in its own right. If one parent has at least some secondary education, the probability of attaining a higher education qualification is raised by 13.7 percentage points. This is relative to parents with no formal education, holding all other controls constant. This effect is matched by a decrease of 13.8 percentage points in the likelihood of attaining a general education level. Having parents with a general educa-

tion has no effect on educational attainment. We can, therefore, infer from these results that more educated parents are likely to offer more educational opportunities to their children, offering them more social mobility. This is in agreement with Gustafsson's(2007) findings for South Africa. We also find that being in a household with a high household income per capita increases the likelihood of attaining a higher education qualification by 2.71 percentage points. This positive correlation between household income and education attainment implies that education is a normal good. The increase in the effect on the likelihood of higher education is matched by a decrease of 2.63 percentage points in the likelihood of attaining some secondary education. However, the number of siblings and location of residence seem to have no effect.

The type of school attended as defined by historical (apartheid) classification is an important factor. Attending a formerly self-governing school significantly increases the likelihood of attaining a higher education qualification by 3.63 percentage points. This is relative to attending a formerly independent homeland school, holding all other controls constant. This effect is matched by a decrease in the likelihood of attaining some secondary education of 3.56 percentage points.

Table 2.7: Determinants of education attainment for Africans

Dependent variable	Generalised ordered probit				Ordinary least square	
Highest education level attained	General	Some secondary	Matriculation	FET & Higher	Relative education index	Years of schooling
School Inputs: log(pupil-teacher ratio)	0.166**	0.0803**	-0.164**	-0.0827**	-0.122***	-1.150***
	(0.081)	(0.039)	(0.080)	(0.039)	(0.030)	(0.290)
log(pupil-teacher ratio^2)	0.460***	0.0199	-0.442**	-0.038	-0.285***	-2.663***
	(0.178)	(0.171)	(0.200)	(0.080)	(0.084)	(0.810)
Log(expenditure per pupil)	-0.098***	0.0567***	0.0799***	-0.0386**	0.0370***	0.380***
	(0.016)	(0.021)	(0.030)	(0.015)	(0.010)	(0.103)
Log(expenditure per pupil^2)	-0.0351**	0.0062	0.0391*	-0.0102	0.0104*	0.113*
	(0.017)	(0.024)	(0.021)	(0.009)	(0.006)	(0.059)
Individual characteristics: Age	0.0110	0.0053	-0.0108	-0.0055	0.0111***	-0.107***
	(0.008)	(0.004)	(0.008)	(0.004)	(0.003)	(0.034)
Age^2	0.0055***	-0.0019***	-0.0018	-0.0018***	0.0005**	-0.0297***
	(0.001)	(0.0006)	(0.0012)	(0.0006)	(0.0002)	(0.002)
Gender (female=1)	-0.158***	0.128***	0.00419	0.0262	-0.0034	0.528***
	(0.032)	(0.036)	(0.036)	(0.021)	(0.008)	(0.081)
Parent's Education (No schooling=0):						
General	-0.0267	-0.0135	0.0265	0.0138	0.0082	0.0949
	(0.021)	(0.012)	(0.021)	(0.011)	(0.009)	(0.087)
At least some secondary	-0.138***	-0.130***	0.131***	0.137***	0.0640***	0.639***
	(0.023)	(0.034)	(0.020)	(0.042)	(0.013)	(0.116)
Number of siblings	-0.0092	-0.0044	0.0091	0.0046	0.0019	0.0190
	(0.006)	(0.003)	(0.006)	(0.003)	(0.003)	(0.025)
Household income per capita(monthly Rands)	-0.0543***	-0.0263***	0.0535***	0.0271***	0.0236***	0.235***
	(0.016)	(0.009)	(0.016)	(0.009)	(0.006)	(0.056)
Ex-department classification(Independent homeland=1)						
Self-governing territory	-0.0676**	-0.0356**	0.0669**	0.0363*	0.0433***	0.417***
	(0.033)	(0.018)	(0.033)	(0.019)	(0.0132)	(0.130)
Department of education	-0.0570	0.0589*	-0.0011	-0.0008	0.0305**	0.295**
	(0.038)	(0.034)	(0.048)	(0.025)	(0.014)	(0.135)
House of assembly, representative and delegates	-0.0354	-0.0208	0.0356	0.0207	0.0447**	0.453**
	(0.047)	(0.032)	(0.048)	(0.031)	(0.018)	(0.178)
New schools	-0.0804**	-0.0579	0.0809**	0.0574	0.0652***	0.624***
	(0.039)	(0.037)	(0.038)	(0.038)	(0.0214)	(0.212)
Distance to school(kms)	0.00001	0.000002	-0.00001	-0.000002	0.0433***	0.417***
	(0.00004)	(0.00002)	(0.00004)	(0.00002)	(0.0132)	(0.130)
Home location(Rural=1):						
Urban	-0.0069	-0.0034	0.0068	0.0035	-0.0230	-0.192
	(0.040)	(0.020)	(0.040)	(0.020)	(0.016)	(0.155)
Farm	0.0577	0.0206	-0.0549	-0.0235	-0.0255	-0.240
	(0.065)	(0.016)	(0.059)	(0.022)	(0.0225)	(0.226)
Neighbourhood characteristics: Serviced by municipality	0.0213*	0.0103*	-0.0210*	-0.0106*	-0.012***	-0.115***
	(0.013)	(0.006)	(0.013)	(0.007)	(0.00425)	(0.0429)
Community economic status	0.0023	0.0011	-0.0023	-0.0011	0.0128**	0.122**
	(0.013)	(0.006)	(0.013)	(0.006)	(0.005)	(0.054)
Population density and housing quality	-0.0373**	-0.0181**	0.0368**	0.0186**	0.0161***	0.160***
	(0.015)	(0.008)	(0.015)	(0.008)	(0.005)	(0.049)
Constant					0.903***	9.401***
					(0.036)	(0.369)
F-statistic					10.49	51.31
Sample size	**1337**	**1337**	**1337**	**1337**	2993	2993

Note: (1) Standard errors in parenthesis. (2) The * indicates level of significance: * p<0.1, ** p<0.05, and *** p<0.01. (3) control variables in the full specification include age, age squared, gender, parents education dummies, number of siblings, log of household monthly income per capita, school type per apartheid classification dummies, distance to school, home location, and neighbourhood characteristics components.(4) The sample is of African respondents aged 15-30 years. (5) Both expenditure per pupil and household income are in real Rands, deflated by CPI with 2008 as base year. (6) See Appendix for definitions of all variables used. (7) Post-stratification weights are used

Source: Author's calculation based on NIDS Wave 1 (2008)

Attending a new school increases the likelihood matriculating by 8.09 percentage points rela-

tive to attending a formerly independent homeland school. Similarly, this effect is matched by a decrease of 8.04 percentage points in the likelihood of attaining primary education. The importance of classification of the school attended is in line with the findings by Bhorat and Oosthuizen (2008) that the type of school attended, as defined by apartheid classification, positively affects matriculation pass. Distance to school has no effect on the likelihood of educational attainment.

Neighbourhood characteristics are important in determining educational attainment, due to associated externalities. Our results indicate that different neighbourhood characteristics have opposing effects. For instance, living in a neighbourhood that is serviced by the municipality raises the likelihood of attaining a primary education by 2.13 percentage points, holding all other controls constant. This effect is matched by lowering the likelihood of attaining a matriculation by a 2.1 percentage points. On the other hand, living in a densely populated area with shared housing raises the likelihood of attaining a higher education qualification by 1.85 percentage points. This effect is matched by a decrease of 1.81 percentage points in the likelihood of attaining some secondary education. Living in a neighborhood of high economic status has no effect on the likelihood of attaining any level of education. Interestingly, the effects of most of these characteristics depict a symmetrical pattern similar to that depicted by the school inputs. Although we do not discuss the coefficients of these characteristics from the partial generalised probit, table 2.11 in the appendix presents them.

2.5.1.3 Robustness checks

In our robustness checks, first we compare the effects of school inputs on the probability of education attainment from specification with and without other control variables. We find that including other controls substantially reduces the estimated marginal effects of the inputs. We interpret this as an indication that our controls are influential in school inputs provided. As an additional robustness check, we redefine new categories for the dependent variable by reducing them to three, that is, general, secondary (combines some secondary and matriculation), and higher/tertiary (which includes FET as above) education. The estimation of this dependent variable leads us to similar results. That is, negative effects of the pupil-teacher ratio, and positive effects of expenditure per pupil on the probability of, attaining more than a general level of education, but negative effects on attaining secondary and higher education[24]. We further run a least square estimation of the entire sample of individuals who are enrolled and those who have completed, to compare the estimates above. These results are presented in column 6 of Table 2.7.

2.5.2 Relative educational attainment index

In this section, we present results from our full sample. These are individuals who have achieved a given level of education, and individuals who are still studying or enrolled in any institution of

[24]The results from these specifications, although not provided here are available on request.

learning. A commonly used outcome measure is years of school. However, an ordinary least square estimation of this measure is likely to be biased. This is mainly because observations on years of schooling for individuals still enrolled would be right censored, some of the older respondents are likely to have started school late, and some of the individuals are likely to have repeated a grade or more.Thus, our observed years of schooling measured by highest grade attained, is likely to understate the number of years the individual has been in school. To ameliorate some of this bias, we use a relative education attainment index which accounts for any late enrollment and grade repetition.

In the last 2 columns of table 2.7, we present the least square estimation of the relative educational attainment index and years of schooling, controlling for the same variables as in the partial generalized ordered probit analysis. The estimates from the relative educational attainment index model are lower than those from the years of schooling model by a magnitude of about 10. The smaller effect shown in the relative education level index are not surprising because by construction the years of schooling estimates are expected to be larger. Overall, the results from the attainment index are comparable to those from the partial generalised ordered probit. From the results, a smaller pupil-teacher ratio raises the educational attainment index, as indicated, by a –0.122, and this is nonlinear. This means that reducing the pupil-teacher ratio by 10 from the average 31 to 21 would increase educational attainment by 0.05, while reducing the ratio by 20 would increase the attainment by 0.13, a more than twofold increase. This is, however, not a large effect and is significantly smaller than the 0.52 year effect found by Case and Deaton (1999) at the end of apartheid. A possible explanation for the decline in the effect is that there is less variation in the ratio.

We find the effect of this ratio to be nonlinear, a result not shared in (Case and Deaton, 1999). Although similar negative effects are found in other developing countries as reviewed in Glewwe et al. (2011). Our results differ from those of Van der Berg (2007) and Bhorat and Oosthuizen (2008) who using South African data find the ratio to be insignificant. We find expenditure per pupil is a significant (at 1%-level) determinant[25], and has nonlinear effects. An increase in expenditure per pupil by R100, for example, from the mean of R273 to R373 increases attainment by 0.01 while a R200 increase from the mean would yield a 0.02 effect, a twofold increase.

An increase in age by 1 year increases educational attainment by 0.01, and has nonlinear effects. Gender is not an important determinant of attainment. This effect is a contradiction to the years of schooling result and the ordered probit results (the achieved sample), where it has positive effects. This is likely to be due to the use of gender in the generation of the relative education index. Parent's education has a significant impact in determining relative educational attainment. Having a parent with at least some secondary education significantly increases educational attainment by 0.06, relative to having a parent with no formal education. This effect is significantly lower than the effect from the partial generalised ordered probit. However, having a parent with primary education

[25]Expenditure per pupil has no impact when we do not include any other controls but school inputs, and when we include individual characteristics.

has no effect.

Household income per capita has significant effects on attainment. A 100% increase in household income increases educational attainment by 0.024. Unlike in our ordered probit analysis of the achieved sample, school attended is an important determinant in this model. The results indicate that attending a formerly self-governing school or a formerly administered by the Department of Education, or House of Assembly, Representatives or Delegates, or attending a new school, significantly increases educational attainment. This is relative to attending a formerly independent "homeland" school. New school attendance has the largest effect, at 0.07. The number of siblings, home location, and distance to school have no effect on relative educational attainment.

All three neighbourhood characteristics considered have significant effects on educational attainment. The results indicate that living in a neighborhood that is serviced by a municipality decreases educational attainment by 0.012. On the other hand living in a densely populated area with shared housing, and living in a neighbourhood of higher economic status increases attainment by 0.013 and 0.016, respectively. This is in line with our expectation that urban areas that are serviced by a municipality and high density areas are likely to have more schools, which would increase access and hence attainment.

2.6 Summary and conclusion

This study investigates the determinants of educational attainment for Africans in South Africa with a specific interest in school inputs –pupil-teacher ratio, and expenditure per pupil. We considered both actual education attained and a relative educational attainment index using the National Income Dynamics Survey Wave 1 (2008), a household survey that has been matched with national school data to identify the impacts of school quality (pupil-teacher ratio and expenditure per pupil) and socioeconomic characteristics such as parental education, home language, and home location on educational attainment. We also consider neighborhood characteristics such as access to piped water, availability of electricity and street lighting, availability landline telephone, collection of refuse by municipality, availability of toilet facility, average household size and house size, and community unemployment rate. We focus on a sample Africans aged 15 to 30 years who are not household heads or in a position of having dependents.

We find that the school inputs and some of socio-economic factors considered are important determinants of educational attainment. Our findings from the partial generalised ordered probit analysis indicate heterogenous effects in both inputs at the different levels of education considered. The pupil-teacher ratio is more important in influencing the likelihood of educational attainment at lower levels of education than at higher levels, an indication that the ratio has a negative lagged effect on attainment. The effects of expenditure per pupil is first negative at the general level, positive at grade 12, and then a negative at higher level of education. These effects indicate declining returns to fiscal investment in education. The results indicate that other factors such as individual

and socio-economic characteristics are also important, and the results indicate that the secondary-higher education nexus is driven by factors outside of the expenditure per pupil by state, and the pupil-teacher ratio.

The results from socioeconomic characteristics indicate that being female raises the likelihood of attaining some secondary education than being male, but the effects are indifferent on the likelihood of attaining a matriculation or higher education. They show parental education has a positive relationship with education level attained, which suggest more social mobility for Africans with educated parents. Household income is shown to positively influences educational attainment. This is an indication that households consider education as a normal good. The home environment is equally shown to be important; living in an area serviced by municipality negatively affect chances of a higher educational attainment, while living in a densely populated neighborhood that is likely to have had more schools raises chances of a higher educational attainment.

Overall, these findings suggest that in South Africa resources do matter in educational attainment. The results give support to the current government policy of reducing the pupil-teacher ratio and of allocating expenditure per pupil by school quintile. The significant effects of the pupil-teacher ratio are, however, against the general findings and sentiments in South African literature where there is largely no support for a lower pupil-teacher ratio (or class size). Gustafsson (2007), Van der Berg (2007) and Bhorat and Oosthuizen (2008) find insignificant effects and argue against policies that would lower the pupil-teacher ratio, but Case and Deaton (1999) in a study at the end of apartheid find positive and significant effect. There is support in South African literature on spending. Gustafsson (2007) shows support for learning infrastructure such as textbooks, while Bhorat and Oosthuizen (2008) find mixed effects from learning infrastructure, but they find positive effects on other infrastructure such as library. The declining effects with increase in expenditure per pupil is supportive of the general sentiment in literature that any increase in resources is inefficient, and policies increasing resources should be discouraged. The expenditure per pupil we considered is a recurrent expenditure that is specifically for non-capital and non-personnel expenditure and is therefore used for purchase of textbooks, and educational material or equipment for the school; school buildings improvement and maintenance; extra-mural curriculum; and provision of services to the school. The use of these expenditure is therefore at the discretion of school management which makes it vulnerable to misappropriation and inefficient use. Policies to ensure efficient and appropriate use of these resources would therefore be beneficial. On the method of analysis, using the generalized ordered probit allowed us to identify the different patterns of school input effects on educational attainment, a pattern we would not have otherwise identified with a standard ordered probit or with the least square estimates.

Appendix 1

Table 2.8: Definition of variables

Variable	Definition
Education level attained	The highest education level a respondent has attained. The categories are defined as: primary/general (grade 1-9), some secondary (grade 10 or/and 11), matriculation (grade 12), and higher (grade 9 and further education and training certificate, or and grade 12 plus certificate or diploma or degree).
Pupil-teacher ratio	Pupil-teacher ratio in each school as per the Department of Basic Education records matched to NIDS school administration data
Expenditure per learner	Expenditure per is learner derived from matching respondent's school quintile variable in the NIDS administration data to the amount of expenditure allocated by the government using school quintile every years in Rands. We use CPI with 2008 as the base year to calculate the real values.
Age	Respondents age in years as calculated using interview date less their date of birth.
Gender	Respondents gender, female=1 and male=0
Number of siblings	Respondents number of sibling
Respondents/Mothers/fathers education	No schooling=0, general education(grade 1-9)=1, and at least some secondary(grade 10-12, further education and training certificate, and bachelor's degree and other qualifications)=2
Household monthly income	Total household monthly income per capita in Rands. We use CPI with 2008 as the base year to calculate the real values.
Geographic area	Respondents geographical home area; Traditional/rural area=1, , Urban=2, and farm=3
Ex-department classification	Independent homelands=1, Self-governing territory=2, Department of education=3, House of Assembly, representative, delegates=4, and New schools=5
Neighbourhood characteristics	The neighbourhood characteristics the variables are calculated by (a) calculating the average value in each cluster and (b) assigning each individual a value corresponding to his/her cluster. They are: • average rooms per person per PSU, • average number of household members per PSU % of respondents with electricity per PSU, • respondents who share toilet with other households per PSU=1 • respondents who live in a neighbourhood refuse collected by the municipality per PSU=1 • respondents with electricity connected per PSU=1 • Respondents with street lights in their neighbourhood per PSU=1 • people employed per PSU=1 • respondent with pipe water per PSU=1

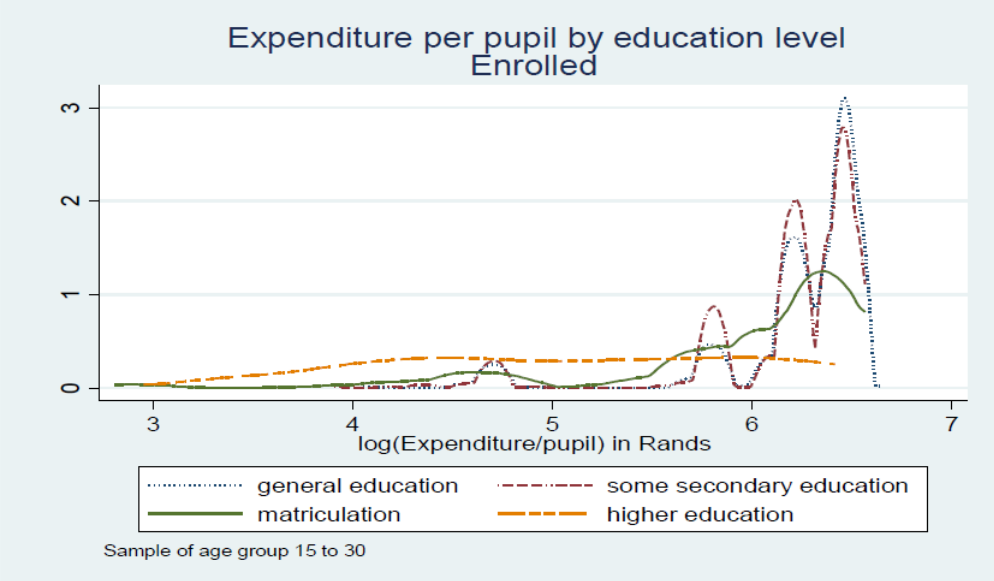
Table 2.9: Mean characteristics of the sample of enrolled by education outcome

Variable name	Enrolled			
	General	Some secondary	Matriculation	FET & Higher
School inputs: Log(pupil-teacher ratio)	32.27	31.80	30.75	29.84
	(0.480)	(0.410)	(0.477)	(0.993)
Log(expenditure per pupil)	524.7	497.8	448.0	220.3
	(13.92)	(12.19)	(18.00)	(31.06)
Age	16.78	18.38	19.28	22.24
	(0.103)	(0.124)	(0.246)	(0.554)
Gender (Female=1)	0.439	0.538	0.505	0.597
	(0.022)	(0.025)	(0.045)	(0.107)
Parents education (%): No schooling	0.539	0.551	0.485	0.597
	(0.024)	(0.026)	(0.059)	(0.107)
General	0.369	0.341	0.298	0.370
	(0.025)	(0.025)	(0.041)	(0.106)
At least some secondary	0.0916	0.108	0.217	0.0329
	(0.014)	(0.018)	(0.047)	(0.023)
Number of siblings	2.754	2.734	2.078	2.950
	(0.113)	(0.122)	(0.224)	(0.466)
Household income per capita(monthly)	637.2	653.2	1399.8	1800.6
	(57.16)	(57.24)	(300.15)	(427.9)
Home location (%)n: Rural/Traditional	0.618	0.565	0.352	0.336
	(0.049)	(0.045)	(0.055)	(0.120)
Urban	0.341	0.381	0.628	0.610
	(0.049)	(0.044)	(0.056)	(0.125)
Farm	0.0416	0.0544	0.0196	0.0544
	(0.017)	(0.018)	(0.016)	(0.040)
School Type (%):				
Independent homelands	0.238	0.119	0.177	0.127
	(0.034)	(0.021)	(0.046)	(0.115)
Self-governing territory	0.395	0.398	0.233	0.122
	(0.036)	(0.036)	(0.041)	(0.084)
Department of education	0.258	0.314	0.424	0.386
	(0.032)	(0.038)	(0.061)	(0.111)
House of assembly, representative and delegates	0.0460	0.0617	0.0851	0.281
	(0.011)	(0.015)	(0.028)	(0.100)
New schools	0.0637	0.107	0.0817	0.0835
	(0.014)	(0.024)	(0.027)	(0.052)
Distance to school(km)	20.09	22.44	34.45	40.98
	(4.43)	(5.81)	(9.059)	(12.59)
Neighbourhood characteristics:				
Number of rooms per person	4.604	4.727	4.335	4.722
	(0.117)	(0.133)	(0.167)	(0.323)
% with electricity	0.729	0.804	0.886	0.953
	(0.035)	(0.026)	(0.022)	(0.017)
% shared toilet	0.146	0.177	0.216	0.251
	(0.025)	(0.024)	(0.031)	(0.058)
% refuse collected	0.303	0.352	0.598	0.613
	(0.044)	(0.043)	(0.055)	(0.113)
% street light	0.276	0.331	0.534	0.578
	(0.035)	(0.039)	(0.050)	(0.093)
Average household members	6.139	5.995	5.424	5.882
	(0.157)	(0.225)	(0.191)	(0.472)
% employed	0.137	0.167	0.203	0.242
	(0.009)	(0.010)	(0.016)	(0.030)
% with piped water	0.788	0.826	0.895	0.947
	(0.034)	(0.028)	(0.026)	(0.029)
% member own house	0.902	0.883	0.842	0.780
	(0.010)	(0.014)	(0.022)	(0.054)
% have landline telephone	0.0810	0.116	0.189	0.215
	(0.010)	(0.016)	(0.036)	(0.047)
Sample size	**1624**	**1624**	**1624**	**1624**

Note: *(1)* Standard errors in parenthesis. *(2)*Pupil-teacher ratio is as provided by the Department of Basic Education in the NIDS administration data. *(3)* Expenditure per is learner derived from matching respondent's school quintile variable in the NIDS administration data to the amount of expenditure allocated by the government to each student using school quintile every year. *(4)*For all the neighbourhood characteristics the variables are calculated by (a) calculating the average value in each cluster and (b) assigning each individual a value corresponding to his/her cluster. *(5)* The sample is of African respondents aged 15-30 years. *(6)* See Appendix for definitions of all variables used. *(7)* Design weights are used

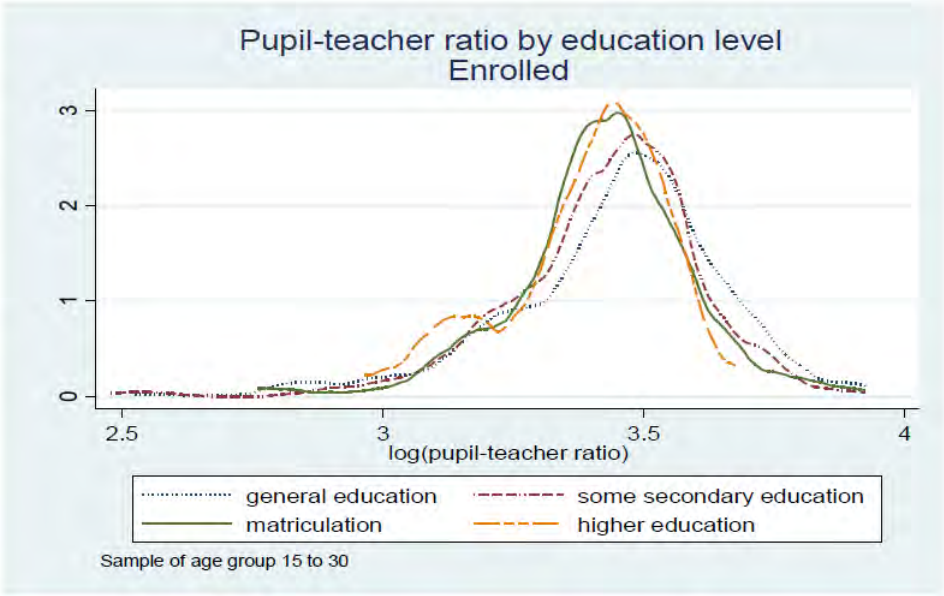
Source: Author's calculation based on NIDS Wave 1 (2008)

Figure 2.5: Distribution of the logarithmic expenditure per pupil for those enrolled



Source: Author's calculation based on NIDS Wave 1 (2008)

Figure 2.6: Distribution of the logarithmic pupil-teacher ratio for those enrolled



Source: Author's calculation based on NIDS Wave 1 (2008)

Table 2.10: Components loadings and communality based on principal component analysis with promax rotation for 10 neighborhood characteristics

Neighbourhood characteristics	Community economic status	Serviced by local municipality	Population density and housing quality	Communality
Average household members	0.3927			0.5187
% employed	-0.4949			0.301
% member own house	0.5817			0.3126
% shared toilet	-0.3596		-0.3158	0.5939
Number of rooms per person			0.6771	0.272
% have landline telephone			0.5879	0.1976
% with electricity		0.59		0.3723
% refuse collected		0.3905		0.2412
% street light		0.3833		0.2431
% with piped water		0.564		0.3996

Note: Factor loading of <0.3 are suppressed

Source: Author's calculation based on NIDS Wave 1 (2008)

Table 2.11: Marginal probability effects of full model

Dependent variable: Highest education level attained	θ_1	θ_2	θ_3
School Inputs: Log(pupil-teacher ratio)	-0.622**	-0.622**	-0.622**
	(0.298)	(0.298)	(0.298)
Log(pupil-teacher ratio^2)	-1.724***	-1.212**	-0.286
	(0.663)	(0.586)	(0.599)
Log(expenditure per pupil)	0.367***	0.104	-0.291***
	(0.057)	(0.067)	(0.109)
Log(expenditure per pupil^2)	0.132**	0.0730	-0.0766
	(0.065)	(0.057)	(0.070)
Individual characteristics: Age	-0.0412	-0.0412	-0.0412
	(0.030)	(0.030)	(0.030)
Age^2	-0.0204***	-0.009**	-0.0135***
	(0.004)	(0.004)	(0.005)
Gender (female=1)	0.568***	0.0768	0.202
	(0.111)	(0.107)	(0.165)
Parent's Education(No schooling=0):			
General	0.102	0.102	0.102
	(0.081)	(0.081)	(0.081)
At least some secondary	0.696***	0.696***	0.696***
	(0.155)	(0.155)	(0.155)
Number of siblings	0.0344	0.0344	0.0344
	(0.024)	(0.024)	(0.024)
Household income per capita(monthly Rands)	0.204***	0.204***	0.204***
	(0.061)	(0.061)	(0.061)
Home location(Rural=1):			
Urban	0.0260	0.0260	0.0260
	(0.152)	(0.152)	(0.152)
Farm	-0.201	-0.201	-0.201
	(0.212)	(0.212)	(0.212)
Ex-department classification(independent homelands=1)			
Self-governing territory	0.261**	0.261**	0.261**
	(0.128)	(0.128)	(0.128)
Department of education	0.221	-0.0048	-0.0059
	(0.151)	(0.150)	(0.191)
House of assembly, representative and delegates	0.141	0.141	0.141
	(0.198)	(0.198)	(0.198)
New schools	0.349*	0.349*	0.349*
	(0.192)	(0.192)	(0.192)
Distance to school (kms)	-0.00002	-0.00002	-0.00002
	(0.0001)	(0.0001)	(0.0001)
Neighbourhood characteristics: Serviced by municipality	-0.0799*	-0.0799*	-0.0799*
	(0.048)	(0.048)	(0.048)
Community economic status	-0.0086	-0.0086	-0.0086
	(0.048)	(0.048)	(0.048)
Population density and housing quality	0.140**	0.140**	0.140**
	(0.057)	(0.057)	(0.057)
Constant	-0.574	-1.568***	-2.794***
	(0.420)	(0.419)	(0.462)
Sample size	**1337**	**1337**	**1337**

Note: *(1)* Standard errors in parenthesis. *(2)* The * indicates level of significance: * p<0.1, ** p<0.05, and *** p<0.01. *(3)* Pupil-teacher ratio is as provided by the Department of Basic Education in the NIDS administration data. *(4)* Expenditure per is learner derived from matching respondent's school quintile variable in the NIDS administration data to the amount of expenditure allocated by the government to each student using school quintile every year. *(5)* For all the neighbourhood characteristics the variables are calculated by *(a)* calculating the average value in each cluster and *(b)* assigning each individual a value corresponding to his/her cluster. *(6)* The sample is of African respondents aged 15-30 years. *(7)* Both expenditure per pupil and household income are in real Rands, deflated by CPI with 2008 as base year. *(8)* See Appendix for definitions of all variables used. *(9)* Post-stratification weights are used.

Source: Author's calculation based on NIDS Wave 1 (2008)

Chapter 3

External Returns of College Education in South Africa: Evidence from Longitudinal Data

3.1 Introduction

An individual's education level can be considered either as a factor of production, or as a signal of innate ability by employers. As a factor of production, it is perceived to have both private and social benefits. The magnitude of private benefits has been unequivocally established in empirical work at a global average of 10%, and higher in sub-Saharan Africa (Psacharopoulos and Patrinos, 2004). The existence of social benefits–also known as external returns–is identifiable in increased productivity as manifested in wages, and non-market effects, such as increased participation in leisure activities. However, in cases where an individual's education is a signal of innate ability, external returns would be zero, since the effect of increase in the aggregate education level on aggregate earnings would be zero (Muravyev, 2008 and Moretti, 2004a), or they would be less than private returns (Moretti, 2004a and Acemoglu and Angrist, 2001).

A debate on the magnitude of external returns–whether they are larger or smaller than private returns–has been on-going in the literature. The source of contention emanates from issues of identification of external returns. External returns are thought to accrue through formal and informal worker interactions where they share knowledge and information (Rauch, 1993). Following interactions, spillover effects in worker productivity could be evident in improved productivity of a lower skilled worker, that is, through technological spillovers (Moretti, 2004a); pecuniary spillovers (expanded technological possibilities that could be observed in for instance, a firm's discovery, adaptation and use of new knowledge) (Moretti, 2004a; and Acemoglu and Angrist, 2001); and community non-market effects that could be observed in say low crime rates, and improved public participation, among others (Mingat and Tan, 1996; and Lange and Topel, 2006).

It is, however, difficult to separate the external effect of education from the effects of other factors, and a clean measure of external returns is unavailable. The question of the magnitude of social benefits of education is therefore, still a valid empirical question. Since the early 1990s researchers such as, Rauch (1993); Acemoglu and Angrist (2001); Moretti (2004a); and Ciccone and Peri (2006) have made efforts to formulate approaches for the estimation of human capital externalities. According to Rauch, the average level of human capital has an indirect influence on total factor productivity through both formal and informal interactions, in which knowledge and skills are shared for technological improvement. Researchers have exploited this view by using average education or share of highly educated workers in a city, state, or region, to estimate external returns for different countries. In a similar manner, we examine the extent to which district council wage differences can be explained by the magnitude of external returns to education in South Africa, by considering the effect of the district council share of college graduates on workers' wages.

The estimation of the magnitude of external returns is important for at least two reasons. The first is the assessment of efficiency of public investment in education. Acemoglu and Angrist (2001) argue that an external return to education of 1-3% is sufficient to justify significant public subsidies for education. A second reason is that for local government, when faced with several policy options that could influence the supply of educated workers in their locality, knowing the magnitude of external returns is critical in guiding their selection of an optimal development policy (Moretti 2004a). In most studies, the question of efficiency of public investment in higher education has been at the crux of the efforts to estimate external returns (such as, Rauch, 1993; Acemoglu and Angrist, 2001 and Moretti, 2004a). The responsibility of financing higher education (whether private or public) is contentious. The contention arises from two opposing views: One, that higher education accrues private returns to the individual, and two, that higher education yields positive externalities to society.

In South Africa, higher education has been argued to lead to more private returns than social returns (Schultz, 2004). Provision of free higher education in South African has been contested on the basis that it would disproportionately benefit the rich, and that it would be too costly for the state (Wangenge-Ouma and Cloete, 2008). The South African government views higher education as having both enormous social and public value, which justifies the use of public resources and energies (Council for Higher Education, 2004). Abuja and Schultz (1996) advocate for an increase in government funding of secondary and tertiary education to address existing inequality, and as a means to benefit previously disadvantaged population groups. Filling in the education gaps and eliminating the racial, class and gender education inequalities created in the apartheid era, has been a priority for the government. In addition, to South African policy makers, education is a means to increasing employment levels, and an important element in fostering political and social stability in the country (Council for Higher Education, 2004). In the case of South Africa, where education is considered to have some characteristics of a public good, the existing evidence of high and convex

private returns may be an underestimation of the full benefits of education to society, and might be leading to an inefficient level of higher education provision. The estimation of external returns of higher education for South Africa is therefore of chief importance, particularly given its current funding status.

Currently, funding of higher education in South Africa is based on a cost sharing system. In the last decade, the South African government expenditure on education as a percentage of GDP averaged 4% (DBE, 2011), which is lower than the 5.5% spent by Brazil (OECD, 2012), but is much higher than the average 3% spent in Sub-Saharan Africa (DBE, 2011). The government's expenditure on education as a percentage of the total government expenditure has, however, declined from 20.5% in 2000/01 to 17.7% in 2009/10 (DBE, 2011), while that of Brazil has increased by 6.3 percentage points in the same period (OECD, 2012). This decline in government's expenditure on education has affected the funding allocated to higher education. For instance, government expenditure allocation to higher education institutions as a percentage of GDP was 3.02% in 2000, it declined to 2.38% in 2008, but marginally increased by 0.09 percentage points in 2011(Financial and Fiscal Commission, 2012). This decrease in funding could potentially compromise future human capital in a country already dealing with a skills deficit situation. This calls for empirical evidence on the external benefits of higher education over and above private returns, that would inform policy on higher education funding.

The South African government recognizes and acknowledges the social role played by higher education. The National Plan for Higher Education (NPHE) of 2001, a government policy plan, recognized that highly educated individuals would work to contribute to political social stability, and that having an education would enhance the capacity of citizens to exercise and enforce their democratic rights, and participate effectively in decision making. This view is reiterated in the National Development Plan, 2030[1]. Given the ongoing debate in developed countries, and the importance of the contribution of higher education to South Africa's social development, it is rather surprising that only one study (Michaud and Vencatachellum, 2003) has considered human capital externalities in South Africa. This study is therefore an important contribution to the scant literature on this topic.

Drawing on the first nationally representative longitudinal data for South Africa–the National Income Dynamics Survey–we estimate the magnitude of external returns over and above private returns. Narrowing down on an unbalanced sample of workers aged 15 to 64 years, who are employed in full-time or part-time jobs, we investigate the interaction of district council share of college graduates and the wages of these workers. We estimate a Mincerian regression augmented by district councils' share of college educated individuals. To deal with the endogeneity of the district council share of college graduates, in addition to controlling for individual and district council fixed effects, we use the average annual district council maximum temperature, minimum temperature, and rainfall, as instruments. Defining favorable climatic conditions as amenities, skilled workers are more

[1] http://www.gov.za/issues/national-development-plan-2030

likely to migrate to district councils with a favorable climate However, the favorable climate is unlikely to affect their productivity, particularly in the long run. We find that a 1% increase in the district council share of college graduates raises workers wages by 5-8%. This positive effect is in line with estimates from other countries, albeit significantly larger. This is also in line with our expectation given that South Africa has a high skills premium. We also find that there are spillover effects with college graduates being the winners. Our results show that a 1% increase in the share of college graduates raises wages of workers with at least some college education by almost 2 percentage points more than the wages of workers with less than a college diploma. This is contrary to the theoretical model and finding by Moretti (2004b). These results, however, are similar to those for Germany (Heuermann, 2011) and considering the prevailing labor market conditions of high skills premium and unionization of low skilled workers in South Africa, they do make sense.

The remainder of this chapter is organized as follows: Section 2 reviews the literature on external returns to human capital. Section 3 outlines the approach and method of analysis. Section 4 gives the background of education and employment in South Africa, and the data used. Section 5 discusses the empirical results of external returns and spillover effects, and conducts robustness checks of the results on external returns. A conclusion is given in Section 6.

3.2 Evidence of external returns to human capital

The question of the magnitude of external returns and whether they exceed private returns is yet to be settled. There is limited evidence to support or dispute the debate, and theory does not give clear guidance (Acemoglu and Angrist, 2001). In one of the earliest estimations of external returns, Rauch (1993) estimates that social returns exceed private returns by a factor of 1.7. He argues that positive externalities from formal education explain this divergence between private returns and social returns. Since Rauch's work, evidence of the magnitude of external returns over and above private returns has been growing. Most of this evidence, which is mainly from developed countries, indicates that there are external returns over and above private returns. Using college share in a city as a measure of external returns, Moretti (2004b) finds a return of 0.6-1.2% above and beyond private returns. However, an earlier study by Acemoglu and Angrist (2001) found no external returns, which they infer to be a reflection of inadequate evidence in support of external returns.

The divergence in estimates of these two studies as discussed in Sand (2013), can be explained by three factors: First, as Moretti (2004b) points out, using cities gives larger effects than using states; Acemoglu and Angrist's (2001) analysis is at state level which could explain the zero effects. Second, their identification of spillovers differ; Acemoglu and Angrist (2001) use compulsory schooling which is at a lower schooling level as compared with Moretti (2004b), who exploits variation in the share of college education. Finally, the two studies analyze data from different time periods; Acemoglu and Angrist (2001) analyze data from 1960-1980, while Moretti (2004b)

analyzes data from the 1980s. Using data from 1980, 1990, and the 2000 US Census, Sand (2013) finds social returns to be unstable, and they declined between the two periods: 1980-1990 and 1990-2000. Considering social returns at a metropolitan level and using three different instruments–city lagged demographic structure, immigration flows, and city climate–to identify the fraction of workers with college education in a city, Sand (2013) finds social returns of 1.84-2.16% for 1980-1990, which is consistent with estimates by Moretti (2004b). However, for 1990-2000, Sand finds zero to negative and significant external returns, results similar to those by Acemoglu and Angrist (2001).

Different methods of analysis have been developed in the estimation of human capital externalities. The pioneering study by Rauch (1993) used the Mincerian approach. This was further adopted by Acemoglu and Angrist (2001). Moretti (2004a) introduces an alternative to the Mincerian approach by augmenting the Mincerian estimation with a measure of aggregate human capital–in his case the measure being the share of college graduates. This approach has been the most adopted in studying externalities in different countries, and using different data sets. Studies that have employed these two approaches differ in three main ways. First, the geographic level of analysis. Some studies use state (Acemoglu and Angrist, 2001) others cities (Liu, 2007 and Moretti, 2004b), while others use region-level data (Heuermann, 2011 and Kirby and Riley, 2008). Given that externalities are localized, the size and significance of estimates may be substantially influenced by the geographical level of analysis considered. Moretti (2004a) and Sand (2013) argue that the reason for low estimates in Acemoglu and Angrist (2001) is due to the use of state rather than city.

The second difference is in the identification of human capital externalities. Identification of external returns has been at the crux of empirical work. Rauch (1993) uses two site characteristics; climatic mildness and coastal location, to deal with unobserved characteristics that could influence wage and rent, and two measures of wage dispersion to deal with selection. He, however, fails to consider the possibility of reverse causality—an issue well addressed in Acemoglu and Angrist (2001), Moretti (2004b), and Sand (2013), among others. Some of the instruments used include: compulsory schooling laws and child labor laws (Acemoglu and Angrist, 2001), the lagged city demographic structure and presence of land grant colleges (Moretti, 2004b), immigration flows and city climate (Sand, 2013), and regional number of schools and of students (Heuermann, 2011).

Third, is the measure of aggregate human capital considered. To avoid finding instruments for multiple measures, most studies assume that the supply of human capital can be measured using a single measure. In the existing literature, the measures mainly used are the average years of schooling and the share of workers with at least a given level of education; for example, the share of workers with college education. The average level is likely to yield a lower effect when compared with the effect of the share of, for instance, college graduates, as shown in Moretti (2004b).

A more recent approach is one developed by Ciccone and Peri (2006) which the authors refer to as the constant-composition approach. The approach offers two alternatives for identifying aggregate human capital externalities; either as an earnings-weighted average percentage-change in wages, or as a log-change in the average wage, holding skill-composition constant. Ciccone and

Peri argue that unlike the constant-composition approach, the Mincerian approach by weighting wage changes by employment rather than earnings, yields positive externalities even when there are none, and its requirement of the estimation of individual returns to human capital calls for instrumentation of individual schooling. These different approaches lead to different estimates of external returns. Studies using the Mincerian approach and the Mincerian augmented approach by Moretti (2004a) have mixed findings on the existence of human capital externalities–the estimate ranges from zero in the U.S. (Acemoglu and Angrist, 2001) to as high as 16% in China (Liu, 2007). Using their constant-composition approach, Ciccone and Peri (2006) find no evidence of externalities from average-schooling when more educated and less educated workers are considered imperfect substitutes, but find positive and significant externalities when these workers are considered as perfect substitutes. The question of the existence and magnitude of human capital externalities is therefore still a valid empirical question.

A spline function of education level attained is commonly used in empirical analysis. Its attraction potentially is that it makes possible the identification of spillovers, that is, the effects of educated workers in a geographical area on wages of workers with different education levels. In looking at the spillovers from an increased productivity angle, low skilled workers are expected to benefit. However, whether highly skilled workers benefit from the increase is not given, but is dependent on whether spillovers exist, and if they do, how large they are. According to Moretti (2004a) and Moretti (2004b), a positive effect on wages for low skilled workers is expected whether there are externalities or not, but any effect on wages of skilled workers is dependent on whether the downward effect of the increase in supply of skilled workers is smaller than the spillover. Using census data for 1980 and 1990, Moretti (2004b) finds an increase in the share of college workers in a city raised wages of college graduates by 0.4%. This is lower by more than one percentage point when compared to 1.9%, 1.6%, and 1.4% for high school dropouts, high school graduates, and workers with some college education, respectively. Similarly, using US census data for 1980 and 1990, Sand (2013) finds an increase in the share of college graduates raised wages of high school graduates by a slightly higher percentage than that for college graduates. On the contrary, in analysis of social returns in Germany for the period 1995 to 2001, Heuermann (2011) finds an increase in the share of highly skilled workers led to a three-fold increase in wages for highly skilled workers relative to low skilled workers. The author associates this effect with the collective wage agreements which sets wages for the majority of low skilled workers. Unions are therefore likely to significantly influence the effect of human capital externalities in a country.

Perceived external returns to education have been used as justification for public funding of education. Public funding of basic education–at primary, and in some countries, secondary level–has largely been an accepted norm. However, public funding of tertiary education is highly debated since it is viewed to confer benefits to the individual, and evidence on social benefits of education has been limited. Estimates of external returns in one of the earliest studies (Rauch, 1993) indicate an additional year of average education in a city raised wages by 2.8%; a considerably high effect.

Acemoglu and Angrist (2001) argue that social returns of 1-3% are sufficient to justify government funding of education. A similar argument for government funding is made by Moretti (2004b), who finds a external return of 0.6-1.2% from a one percentage point increase in college share in a U.S. city. These results are confirmed by Sand (2013). Focusing on a sample of men aged 30-49 in the U.K., Kirby and Riley (2008) find an increase in the industry average education level raised individual wages by 2.6-3.9% over and above private returns.

Empirical evidence of human capital externalities for emerging and/or developing economies is scarce. Emerging economies are also often in transition–either in structural or policy reforms– and are characterized by growth in both physical and human capital, albeit not necessarily in equal proportion. Evidence from emerging economies seems to support the existence of human capital externalities when the measure considered is either average education of workers, or share of workers with higher education. For instance, on the one hand, using average years of schooling in a city, Liu (2007) finds a one year increase in a city's average education led to external returns of 11-13% in China over and above private returns, leading to estimated social returns of as high as 16%. Similarly, using regional average education as a measure of human capital externalities in Turkey, Filiztekin (2011) finds social returns of 3-4% over and above private returns. Their results using the share of university graduates indicate higher education has a much higher social return. On the other hand, using the share of workers with higher education in 39 Russian cities, Muravyev (2008) finds that a 1% increase in the share of workers with higher education in a city led to a 1.33% increase in social returns.

Most of the emerging economies are also characterized by a large government sector. A large government sector has been shown to lead to lower social returns. In India, Schündeln and Playforth (2014) found the substantial government employment of educated workers led to lower social returns than would have been realized if the skilled workers were employed in the private sector. Schundeln and Playforth associate these low returns with unproductive government employment positions that are commensurate with higher wages relative to their productivity (direct effect), and government employees exerting negative externalities on the productivity of the private sector through activities such as licensing, regulations, and rent-seeking. Psacharopoulos and Patrinos (2004) argue against the inclusion of government employed workers in estimation of returns to education, since their wages are not based on the labor market. The limited evidence of external returns from Africa is mixed. Nazier (2013) finds higher education in Egypt has negative externalities for female workers and none for male workers. Nazier attributes this to the excess supply of higher education graduates in the labor market, and a possibility that higher education is perceived as a signal of ability rather than as a factor of production.

In South Africa, private returns to education have been shown to be positive and convex (Keswell and Poswell, 2004 and Burger and Jafta, 2006). They, however, vary by age, gender, race, level of education and region. Contrary to the theory of human capital, Mwabu and Schultz (2000) show that returns in South Africa increase for the young and inexperienced. They found the returns in-

crease by 8% and 4% higher for 16 to 29, and 30 to 44 year olds, respectively. Returns to education in South Africa increase with the level of education (Mwabu and Schultz, 2000 and Burger and Jafta, 2006). Returns to education appear to discriminate by gender and race. Mwabu and Schultz (2000), using South Africa Project for Statistics on Living Standards and Development (PSLSD) 1993 data, find returns to tertiary education are higher for women than men amongst all non-whites, and in the rural areas they are higher for Africans. Mwabu and Schultz (2000) also found higher education institution type had no special returns.

Although the literature on private returns to education in South Africa is broad, the literature on social returns is limited. One exceptional study is that by Michaud and Vencatachellum (2003), who exploit the 1993 South Africa PSLSD data for existence of human capital externalities at provincial level by race using a Mincerian regression. The authors find inter-race human capital externalities differ by race; an increase in the average human capital of Blacks increases wages for Whites. However, they find an increase in the human capital of Whites reduces the wages of Blacks. They infer this to imply that as more higher educated White workers become available, employers replace Black employees with White employees. They also find positive human capital externalities within race—an implication that the effect of an increase in supply of more highly educated workers does not override the demand effect within race. Although this study considers the possibility that selection in employment could bias the estimates, it neglects the possibility that the estimates of human capital externalities could be biased by both individual and regional unobservables, and supply and demand factors. Further, the data used in the study is outdated, in time, and structure of the South African institutions.

The above review can be summarized in five key points: One, although a few studies have found no external returns (Acemoglu and Angrist, 2001 and Ciccone and Peri, 2006), the majority show there are positive and significant external returns to education over and above private returns (for example, Rauch, 1993; Moretti, 2004b; Liu, 2007; Kirby and Riley, 2008; Heuermann, 2011; and Filiztekin, 2011). However, the magnitude of the effects vary by the geographic area considered, the measure of aggregate human capital considered, and the identification approach taken. On the one hand, the variation by geographic area indicates that external returns are localized, and the size of the area matters (Sand 2013). On the other hand, the variation by the measure of aggregate human capital indicates that smaller estimates are to be expected when the average years of education is used, while larger estimates are to be expected when the share of higher education is used. Also, the instruments chosen, particularly if related to education, could influence the estimates. For instance, Acemoglu and Angrist's (2001) choice of compulsory schooling laws that affect lower education levels, and Moretti's(2004b) choice of land grant colleges which affect higher education.

Two, different methods of analysis have been employed in the estimation of external returns, and they include: Mincerian regression (Rauch, 1993; Acemoglu and Angrist, 2001 and Michaud and Vencatachellum, 2003), Mincerian regression augmented with a measure of aggregate human capital (Moretti, 2004b; Liu, 2007; Kirby and Riley, 2008; Heuermann, 2011; Filiztekin, 2011 and Nazier, 2013), and the constant-composition approach Ciccone and Peri (2006). The estimates from these methods are conflicting. Three, spillover effects of aggregate human capital can be estimated by considering different education groups. Although the theory as proposed in Moretti (2004a) shows that low skilled workers' wages benefit more from an increase in highly skilled workers than the skilled workers themselves do (Moretti, 2004b), there is evidence that prevailing market conditions could lead to a different outcome, as in the case of Germany (Heuermann, 2011). Four, the evidence from transition and developing countries is largely supportive of the existence of positive external returns (Liu, 2007; Filiztekin, 2011 and Muravyev, 2008), however, zero or negative effects cannot be ruled out as yet, given the evidence from Egypt (Nazier, 2013). Finally, the literature on external returns in South Africa is limited. Only one study–by Michaud and Vencatachellum (2003) exists; at least to our knowledge. This study therefore contributes towards the sizable research gap that exists.

3.3 Approach and Method

3.3.1 Conceptual and analytical framework

The theoretical and analytical framework of this study is grounded on the theory of human capital pioneered by Schultz (1961), Becker (1964), and Mincer (1974). In the theory, education is considered an investment which enhances an individual's productivity, and this productivity is rewarded with higher wages. To estimate the external returns to education, we follow the theoretical framework by Moretti (2004a) and (2004b). In the following outline, we first, briefly outline the framework as defined in Moretti (2004a) and Moretti (2004b), and contextualize it to our variable of interest–the district council share of college graduates–and the South African labor market institutions that are likely to affect the estimates from this framework.

As in Moretti, we assume a Cobb-Douglas production function that uses human capital–a composition of educated (college graduates) E_1 and uneducated (non-college graduates) E_2 labor–and physical capital K in the production of a composite good Y in district council d and is consumed nationally:

$$Y_d = (\theta_{1d}E_{1d})^{\alpha_0}(\theta_{2d}E_{2d})^{\alpha_1}K^{1-\alpha_0-\alpha_1} \tag{3.1}$$

where θ_{jd} is the productivity of the worker type $j = 1,2$, which allows for spillovers, and is assumed to be a function of the group specific productivity enhancing effects ϕ_{jd} with $\phi_{2d} > \phi_{1d}$, and of the effects of the share of college graduate workers in the district council d:

$$log(\theta_{jd}) = \phi_{jd} + \gamma\left(\frac{E_{2d}}{E_{1d}+E_{2d}}\right) \tag{3.2}$$

In the absence of human capital spillovers, that is, $\gamma = 0$, individual productivity solely depends on own human capital. In this case, the model collapses into a Mincerian model where we only estimate private returns. However, there is some empirical evidence that human capital spillovers exist, hence $\gamma > 0$. Let $S_d = E_{2d}/(E_{1d}+E_{2d}) < 1$ define the share of college educated workers in a district council. Assuming that the spillover is external to the individual firm, but internal to the district council, and wages earned by each type of worker is equal to their marginal productivity, the logarithm of wages for each group can be expressed as follows:

$$log(w_{1d}) = log(\alpha_o) + \alpha_0 log(\theta_1) + 1 - \alpha_1 - \alpha_0 log(K/E_d) + (\alpha_0 - 1)log(1-s_d) + \alpha_1 log(\theta_2 s_d \tag{3.3}$$

$$log(w_{2d}) = log(\alpha_1) + \alpha_1 log(\theta_1) + 1 - \alpha_1 - \alpha_0 log(K/E_d) + (\alpha_1 - 1)log(s_d) + \alpha_0 log(\theta_1(1-s_d)) \tag{3.4}$$

where $E_d = E_{1d} + E_{2d}$. To know what happens when the share of college educated workers in the district council increases, we take the first derivatives:

$$\frac{dlog(w_{1d})}{ds_d} = \frac{1-\alpha_0}{1-s_d} + \frac{\alpha_1}{s_d} + (\alpha_1 + \alpha_0)\gamma \tag{3.5}$$

$$\frac{dlog(w_{2d})}{ds_d} = \frac{\alpha_1 - 1}{s_d} - \frac{\alpha_0}{1-s_d} + (\alpha_1 + \alpha_0)\gamma \tag{3.6}$$

From equation 3.5, an increase in supply of college educated workers in a district council affects the wages of the non-college educated workers in the district council in two ways: through the neoclassical supply effects that arise from imperfect substitution between the college educated and non-college educated workers (given by the first two terms), and through human capital spillover (given by the last term). Overall, the unskilled workers benefit from these two effects. Similarly, the increase in the supply of college educated workers affects their wages in two ways: through the neoclassical supply effect which lowers their wages due to perfect substitution, and through the human capital spillover. The overall effect, however, is ambiguous since it is dependent on whether the spillover effect is large enough to overcompensate for the negative supply effect. Therefore, one would expect to find that an increase in skilled workers has a larger effect on wages of unskilled workers than on wages of skilled workers. Moretti's findings for the USA are in line with this expectation, but Heuermann (2011) finds the contrary for Germany.

Spillover effects from higher skilled workers are likely to occur when lower skilled workers learn or acquire some skills of doing a given activity while working with higher skilled workers,

and these skills leads to an increase in their productivity (rewarded with wages). Similarly, when higher skilled workers work together and learn from each other and as a consequence increase their productivity, then that increase in productivity is a spillover effect. By working with skilled workers, an unskilled or skilled worker is likely to learn how to run a given task either by observing how the skilled worker implements the task or by directly being taught by the skilled worker. For example, a receptionist in a company may learn how to manage an office by observing the office manager or by being taught the office manager. This is likely to increase her productivity in performing her responsibilities and in the general running of the office. The increase in the receptionist productivity can be credited to the skills acquired from the office manager, and are therefore spillover effects. Another example, when data analyst learn from each other how to use different data analysis software, which enable them to do more data analysis, and hence an increase in their productivity. This increase in data analysts productivity is a spillover effect from each other. The estimation of spillover effects from skilled workers would therefore give a quantitative indication of the benefit that other workers–whether skilled or unskilled–get from working with skilled workers. This estimation is important for South Africa particularly because higher education is largely perceived to only benefit the individual, and has no externalities.

The effects of an increase in skilled workers on wages in South Africa are likely to differ from those in the USA, mainly because labor market institutions differ substantially between the two countries. To contextualize this framework for South Africa, we consider a number of South Africa's labor market institutions. First, we consider the effects of a change in minimum wages. An increase in minimum wages is likely to reduce employment of unskilled workers by reducing demand for unskilled workers. According to Betcherman (2012), to increase productivity without changing employment levels, employers are likely to increase their demand for skilled workers as they substitute for unskilled workers. In South Africa, minimum wages are determined at national, regional, sectoral and occupational/skill level. Although violation of minimum wages is at 44% of covered worker (Bhorat, Kanbur and Mayet, 2012), in instances where there is compliance, wages increase, and compliance is likely to affect the distribution of workers employed. Bhorat et al. (2013) find an introduction of minimum wages increased wages in district councils that previously did not have minimum wages. There are other possible causes of a decline in demand for unskilled labor. In post-apartheid South Africa, the demand for low-skilled workers declined mainly due to the emphasis on a skill intensive tertiary sector, which has led to persistent concerns regarding the existence of a 'skills shortage' in the country (Rodrik, 2008). This has resulted in a skills and human capital premium where skilled workers are paid substantially higher wages than unskilled workers are. Rodrik (2008) identifies the causes of the skills shortage to be due to: an increase in susbstitutability towards skilled workers, an increase in capital intensive production technologies, and a structural change from low-skilled intensive economic activities. However, the formal sector in South Africa is characterized by a high proportion of semiskilled and lowskilled workers, at 42% (Rodrik, 2008). From this, one would expect a low substitutability between skilled and unskilled workers in South Africa relative to USA. It is therefore likely that wages of unskilled workers may

not rise as much as predicted by the theory.

Second, differences in individual job mobility are likely to lead to differences in spillover effects by country. Employment protection laws are likely to interfere with labor mobility differently from one country to another. South Africa has an affirmative action prioritizing employment of formerly disadvantaged groups, women, and people with disabilities as is stipulated in the Employment Equity Act No. 55 of 1998. The Act requires employers to consider a 'suitably qualified person' from these groups; employers are also required by the Skills Development Act to offer training to members of these groups in order to build their capacity in doing their jobs (Thomas and Jain, 2004). Therefore, it is possible that workers at the margin of the skill levels (that is,workers who are less than technically qualified for positions) get employed into skilled jobs. According to Boeri et al. (2006 as cited in Betcherman, 2012, p23) employment protection rules gravitate towards helping unskilled workers more than skilled workers. The implementation of the Employment Equity Act in South Africa has been slow, and uneven, particularly at highly skilled levels such as managerial positions, for both Blacks[2] and women (Booysen, 2007). Betcherman (2012) argues that employment protection legislation is more likely to have negative effects on the reallocation of labor. If affirmative action is being fully implemented, and a large number of less skilled but disadvantaged workers are deemed 'suitable', employers might substitute skilled workers with unskilled workers, leading to an increase in the wages of unskilled workers.

Third, the system of collective bargaining in place has an influence on wages, and can therefore confound the productivity of workers (reflected in the wages) in a given region. South Africa has a two-tier bargaining system: central bargaining which is conducted at the geographical level, and unilateral bargaining, which is conducted at the plant level. Different industries in one location may therefore be covered by different agreements at different points in time. Magruder (2012) finds that in any given town, industries covered by collective agreements have lower employment by about 8-13%, and higher wages by about 10-21%, than similar industries in uncovered towns. Plant-level bargaining focuses on reducing the wage gap between skilled and unskilled workers (Vettori, 2005 as cited in Bhorat, Goga and Van Der Westhuizen, 2012, p13). Although participation of both unions and employers is not mandatory, extension of agreement to non-union members is common. Bhorat, Goga and Van Der Westhuizen (2012) find a wage premium of 9-10% for non-union workers, and a 22% premium for union workers in public bargaining councils. According to Magruder (2012), education is not a predictor of bargaining council status in South Africa. This makes it hard to tell, a priori, whether the human capital externalities would be reflected in the wages of skilled or unskilled workers. Heuermann (2011) associates the higher spillover effects on wages of highly skilled workers with the effect of collective bargaining, which prevents wages of unskilled workers from changing with the prevailing market conditions in Germany.

Finally, mandated benefits such as, unemployment insurance, pension, childcare, parental leave, bonuses, and vacation pay, affect earnings, labor force participation, and employment. These ben-

[2]Blacks refer to Africans, Coloured, and Indians/Asians.

efits vary by type of workers' for instance, according to Boeri, Helppie, and Macis (2008 as cited in Betcherman, 2012, p35), fixed cost benefits (that is, cost paid per worker) relative to variable cost (that is, costs that increase with wages or hours worked) would lead to relatively higher labor costs for lowskilled workers, and are therefore likely to shift demand towards high skilled workers. In South Africa, Ranchhod (2009) finds that pension loss increased labor force participation for middle aged adults and older adults amongst resident household members.

In light of these specific labor market institutional effects, we estimate the effects of the district council share of college education on human capital externalities. To do this, as in Moretti (2004a) and (2004b), we assume constant spillover effects across education groups and estimate the effects of college education on wages of different education groups. This enables us to differentiate between imperfect substitution and human capital externalities. As discussed above, a positive and significant coefficient on the district council share of college graduate variable would be an indication of the existence of human capital externalities in South Africa. Therefore, our null hypothesis is that there are no external returns to education over and above private returns to education in South Africa.

3.3.2 Estimation and identification of human capital spillovers

To estimate human capital externalities, the existing literature follows a similar estimation approach of regressing the logarithm of wages on aggregate human capital, while controlling for individual characteristics, including individual's education. As in Moretti (2004b) and Heuermann (2011), we estimate a wage equation as follows:

$$log\,(w_{idt}) = X_{it}^{'}\beta_{dt} + \delta C_{dt} + Z_{dt}^{'}\alpha_t + \eta_d + \varepsilon_{idt} \tag{3.7}$$

where w_{idt} is the nominal monthly wage[3] of individual i in district council d at period t; $X_{it}^{'}$ is a vector of individual characteristics, including the individual's years of schooling, age, gender, race, marital status, job tenure, home location, home province, type of employment contract, whether one holds a second job, trade union membership, pension deduction, medical deduction, and unemployment insurance deduction ; C_{dt} is a measure of the level of aggregate education in district council d at period t; $Z_{dt}^{'}$ is a vector of district council time varying characteristics, including district level unemployment, racial composition, gender, sector and industry of employment; η_d is the district council fixed effects; and ε_{idt} is the error term. In this analysis, δ estimates the human capital spillovers and is our variable of interest. Given that the NIDS three wave panel was collected over a span of six years, we do not include year fixed effects since we consider this period to be too short for the majority in our sample to have changed their education level.

In estimating the external effects of higher education, we first need to deal with the issue of

[3]Nominal wages would best reflect the productivity of a region since employers would be willing to pay higher nominal wages to workers who guaranteed more productivity (see Moretti (2004a) for further discussion).

identification of the share of college graduates in the district council. This stems from the possibility that C_{dt} might be correlated with unobservable district council characteristics, and unobservable individual characteristics that potentially change over time. The error term is therefore a composite of three factors:

$$\varepsilon_{idt} = \gamma_d \theta_i + \sigma_{dt} + \nu_{idt} \tag{3.8}$$

where θ_i is the permanent unobservable component of individual human capital comprising ability, motivation or family background characteristics; γ_d is the district council specific coefficient that enables the permanent individual unobservable components to be valued differently in each district council; σ_{dt} are the district council time-varying labor demand and supply shocks in district council d at period t; and ν_{idt} is the transitory component of the logarithm of wages, which is assumed to be independently and identically distributed across all individuals, in every district council, and at each period.

In the presence of unobserved individual and district council characteristics, θ_i and γ_d would be correlated with C_{dt}, which would introduce bias in δ. To deal with this omitted bias, we follow two estimation strategies. First, we include individual fixed effects which deals with unobserved individual characteristics. Second, we estimate a district council fixed effect model to deal with omitted variable bias due to unobserved district council characteristics such as apartheid spatial planning. However, these fixed effects estimates are still likely to be biased by time-varying factors σ_{dt}, which could be correlated with changes in both human capital, and wages across the district councils. For instance, in case of a positive shock that increases demand for skilled workers, wages would increase leading to a positive bias in our estimate of δ. Similarly, if there was a supply shock that limited the number of skilled workers in a district council, the small share in skilled workers would lead to a negative bias in our estimate of δ. Therefore, we employ an instrumental variable approach to deal with endogeneity of the share of skilled workers in a district council due to supply and demand shocks of skilled workers.

3.3.2.1 The validity and exogeneity of the Instrumental Variable

In our search for an instrumental variable(s), our challenge is in finding a variable(s), that is correlated with the share of district council college graduates, but is uncorrelated to shocks in supply and demand of these workers (Angrist and Krueger, 2001). We follow the lead by Sand (2013), who uses city climate to instrument for the share of college graduates in a city; an instrument inspired by Dahl (2002). Climate has been shown to have no effect on wages (Rauch, 1993). Dahl (2002) finds that, in addition to comparative advantage in earnings, migration of college graduates from state to state, relative to migration of high-school graduates, is also motivated by amenities such as, climate, quality of life, and state spending and taxing. Given this idea that skilled workers place a higher value on amenities than unskilled workers do, we employ the annual average maximum

temperature, minimum temperature, and rainfall[4] in a district council as our instruments for the district council share of college graduates.

In post-apartheid South Africa, the probability of workers to migrate from one district to another increases with education level (Kok et al, 2003). In the period between 1990 and 2000, South African cities with a higher proportion of individuals with tertiary education grew faster (Naudé and Krugell, 2003); this is a likely indication of increased productivity. The opportunity to earn higher wages, and regional amenities such as safer neighborhoods with low crime rate determine individual's preference regarding district of residence (Choe and Chrite, 2014). However, for skilled workers, the choice of where to reside by city is often limited by, for instance, availability of jobs which could force most skilled workers to live in a given city. It is also important to note that, the choice of district council is artificial in that, it is essentially property prices which shape location. It is therefore possible that a skilled/unskilled worker could choose to reside in a given district council due to its favourable property prices and vice versa. According to Choe and Chrite (2014), individuals with at least post-secondary education when compared with those with lower levels of education are more likely to live in areas with higher proportions of individuals with similar education levels. They also find both groups are attracted to areas with positive amenities such as safer areas with low crime rate. Defining favorable climatic conditions as an amenity, and given that it is an non-economic factor that influences migration of college workers (Dahl, 2002), we deem it to be an good predictor of the share of district council college graduates. As for exogeneity of climate, we make a similar assumption as in Sand (2013), that there is no variation over time in the relationship between productivity and city-specific climate. A threat to the validity of climate as an instrument would be if any technologies are developed or adopted to suit the city-specific climate; an unlikely possibility in our period of study. Although rainfall or temperature are likely to influence an individual's productivity in the short-run, we consider it unlikely that an individual's productivity would be influenced by the climate over time. Therefore, this means that climate has to vary over time.

3.4 Background and Data

3.4.1 South African education system

There are three levels of formal education in South Africa, which include the General Education and Training (GET), Further Education and Training (FET), and Higher Education and Training (HET). The South African government is the main funder at all levels of education. Privately funded schools and institutions of higher learning exist, and their numbers have been growing in the post-apartheid era. According to the South African School Act No. 84 of 1996 education is compulsory

[4]The annual average is calculated from the monthly temperature given in degrees Celsius and rainfall given in millilitres. We recognise a longer average such as 5-10 year average would be more appropriate for the instruments, unfortunately we were unable to obtain these data.

from age seven to 15 or grade nine (RSA, 1996), which marks the end of General Education and Training. This level comprises seven years primary education; grade 1 through 7, and two years of secondary education; grade 8 and 9. After grade nine, students are allowed to choose to proceed to senior secondary education (grade 10 through 12) for three years, or join a further education and training institution, or join the labor market. At Further Education and Training level, students are offered vocational training, which is offered in what is commonly referred to as FET colleges. This level is considered parallel to the senior secondary level. Graduates from senior secondary level after matriculation (grade 12), and graduates from Further Education and Training, may either choose to proceed to higher education institutions, or join the labor market. For purposes of this study, we consider individuals who have attained a diploma from either a Further Education and Training college, or a higher education institution as being college graduates.

3.4.2 Education and employment scene in South Africa

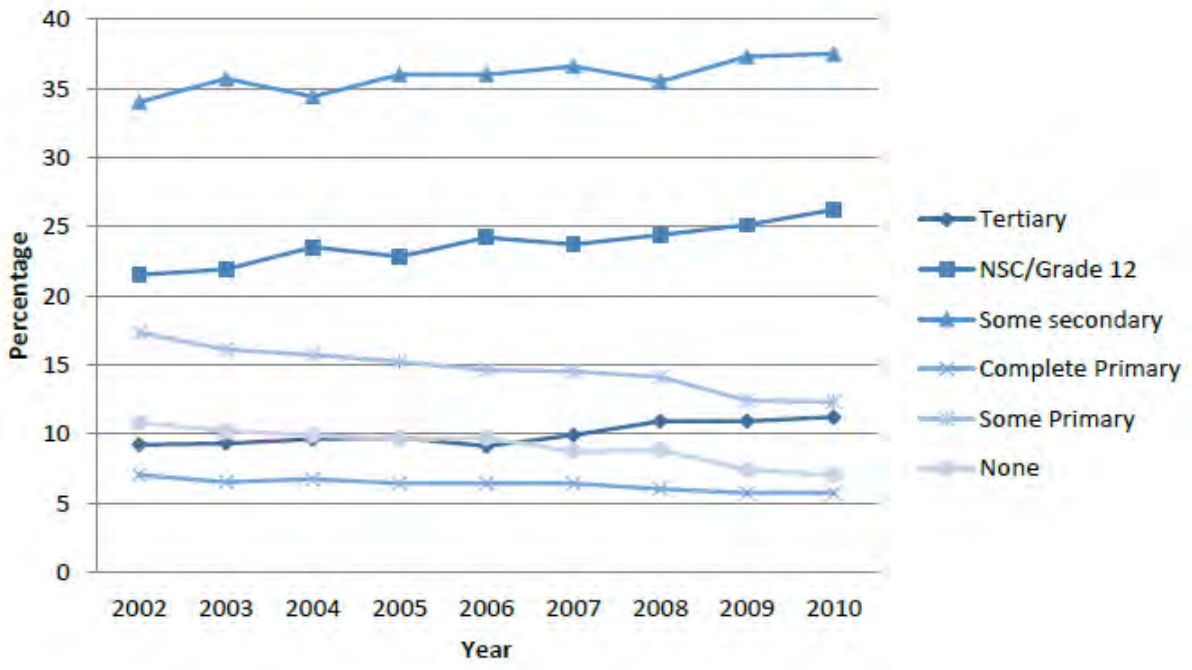
In post-apartheid South Africa enrollment has been on the increase at all levels of education. It has increased to an average of 98% for primary and secondary level. In higher education between 1993 and 2002, enrollment increased by 35% (Council for Higher Education, 2004) and almost doubled[5] by 2008. Attainment has also increased. For instance, in the period between 2002 and 2009, primary school completion increased from 79% to 87%, and secondary school completion rates averaged between 40% and 44% (DBE, 2011). Nonetheless, transition to tertiary education has been low. For instance, in 2011, only 4.8% of 20 to 29 year olds were enrolled at higher education institutions. High dropout rates and grade repetition have been major issues in schooling in South Africa. It has been highest at grade 11 and grade 12 (post compulsory education[6]), where about 58% of students drop out before attaining a National Leaving Certificate (*i.e.* matriculating). Resource constraints and imperfect future foresight on the value of schooling, ability, timing of jobs, and expected post-school earnings are blamed for incomplete schooling (Lam, Leibbrandt and Mlatsheni, 2008). As a consequence, the increase in mean years of schooling[7] has been marginal. For instance, in 2000, it was 8.2 years and increased to 8.5 years by 2012 (Human Development Report(HDR), 2013), which is below the number of compulsory years of schooling. Figure 3.1 below, gives a disaggregated picture of educational attainment at different levels, in percentage points from 2002 to 2011.

[5] See http://www.ieasa.studysa.org/resources/Study_SA/Facts_Figures_section.pdf

[6] Attending grade 1 to grade 9 is compulsory in South Africa.

[7] Mean year of schooling is the average number of years of schooling received in a lifetime by the people aged 25 years and older.

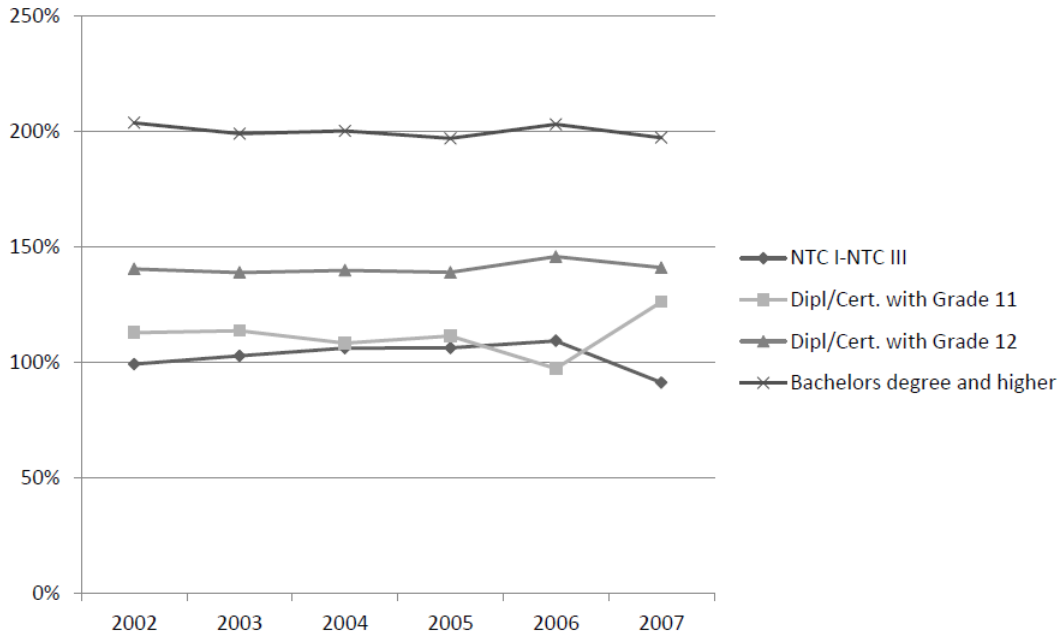
Figure 3.1: Educational attainment for persons aged 20 and above



Source: Author's calculations based on General Household Survey 2011

Figure 3.1 shows that, most South Africans have attained some secondary education, averaging at 36% in the period considered. Attainment of both grade 12 and tertiary level has been on the increase, with the increase reaching 7 and 2 percentage points, respectively, in about a decade. Secondary education attainment has been rising more rapidly than tertiary education attainment. In spite of this, overall the proportion of those with higher educational attainment has been on the rise, while that with lower attainment has been on the decline. In the last decade, as indicated in Figure 1, there has been a steady decline in the population of those with no formal education, complete primary education, and some primary and by 4.2, 1.5, and 5.6 percentage points, respectively. Educational attainment has increased despite the high unemployment rate in South Africa. Although the possibility of unemployed individuals re-enrolling into educational institutions cannot be ruled out, one would expect that this increase in educational attainment would have led to an increase in the number of highly skilled individuals. Unfortunately, the contrary is true, given the high skills shortage in South Africa (Department of Labour, 2005 and Letseka and Maile, 2008), which has resulted in a skills premium. The high wage for higher education graduates depicted in Figure 3.2, can be inferred to be indicative of the skills premium in South Africa.

Figure 3.2: Relative wages of skilled workers to unskilled workers

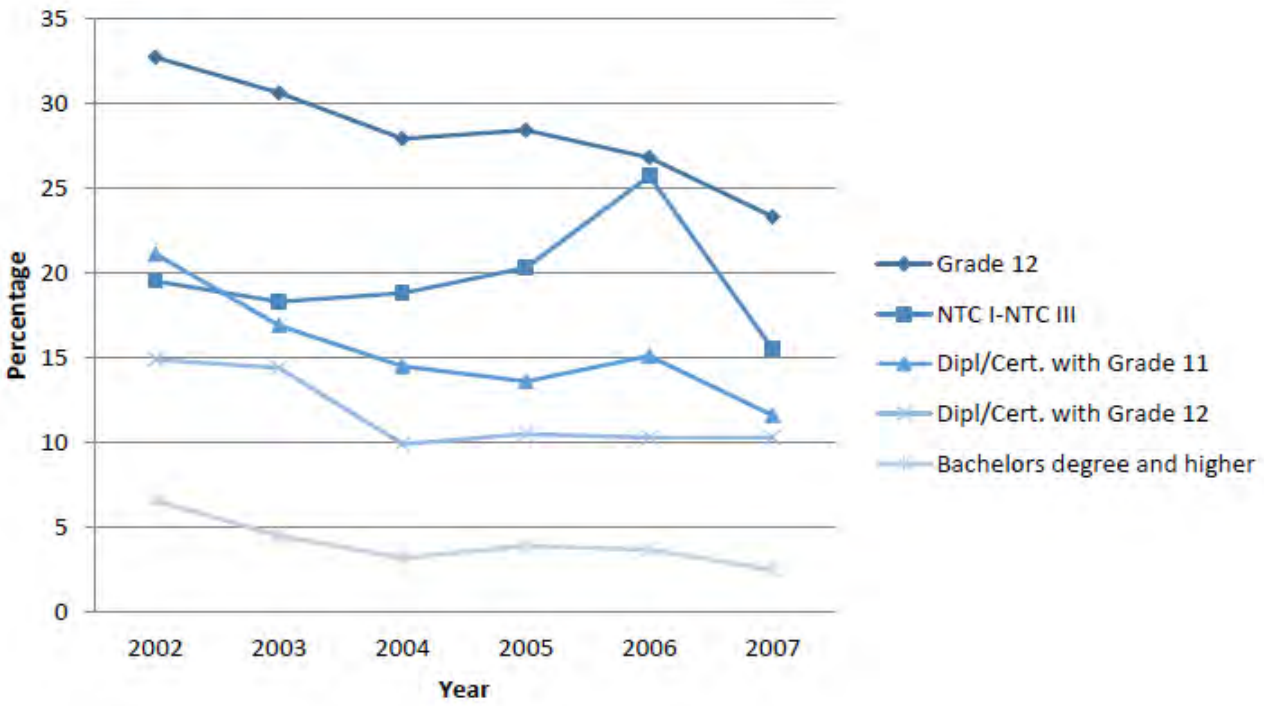


Source: Own calculations based on several Labour Force Surveys

As shown in Figure 3.2, individuals with a higher level of education are rewarded with higher wages in South Africa when compared to non-college graduates, and the differences in wages have remained almost constant over time. The relative wage increases with the level of education: while National Technical Certificate (NTC)[8] graduates earn the least when compared to other graduates, they earn about 100% more than primary and high school graduates. Degree holders are the highest earners, earning about 200% more than primary and high school graduates. It is interesting to note that, overall, there is no overlap in the wages, which could be interpreted to be an indication that either experience is not an important factor in the determination of wages, or experience does not have sufficient weight to influence wages to a level above the level associated with the education level attained. Given the overall high rate of unemployment in South Africa, statistics of unemployment rates amongst tertiary education graduates, as shown in Figure 3.3, indicate that this group is not immune, but it is least affected.

[8]NTC are offered at vocational colleges, that is Further Education and Training colleges.

Figure 3.3: Rate of unemployment



Source: Authors calculations based on several Labour Force Surveys

In general, unemployment has been persistently high in South Africa. As shown in Figure 3.3, unemployment varies by level of education, decreases with the level of diploma, and has been decreasing over time. From the figure, we note that unemployment is highest amongst grade 12 graduates and lowest amongst degree holders; a 20 percentage points difference. There is a significant unemployment margin between degree holders, and individuals with either grade 12, or at least a certificate. From this, one can infers that, attaining a degree significantly lowers chances of unemployment.

3.4.3 Data

The study uses the three waves National Income Dynamic Study (NIDS), which is a nationally representative household and individual level panel, that commenced in 2008. The sample of focus is that of working age population (15-64 years) who are in full- or part-time employment and earn a monthly wage. The logarithm of monthly wages in nominal Rand value is our dependent variable. As discussed in Moretti (2004a), the magnitude of nominal wages in a geographical area should mirror the level of productivity. We therefore do not adjust for the cost of living; our expectation is that college graduates are more likely to earn higher wages and are also more likely to live in district councils where the cost of living is higher. In the NIDS adult questionnaire of NIDS, respondents who were employed were asked for, either their exact take-home monthly wage, or to select a bracket that captured their wage. The years of schooling variable is calculated from the highest education attained response; one year of schooling is given for each level attained, and for a diploma, the official expected years to complete each type were considered. In the NIDS data there are 47 district councils or municipalities and six metropolitan municipalities. They are all referred to as district councils. We, therefore, have 53 district councils in the data. The district council share of college graduates is arrived at by calculating the share of workers with a college diploma in a district council, then assigning that share to individuals in the same district council. A similar approach is taken in calculating all the district council characteristics. We use the district council monthly maximum temperature, minimum temperature, and rainfall averaged for each year, in 2008, 2010, and 2012 to instrument for the district council share of college graduates. The South African Weather Service provided this climate data by district council.

Measurement error in the schooling variable could be a source of bias for our estimates of external returns. This is primarily a problem because measurement error in years of schooling may be correlated with the years of schooling, and is likely to have a downward bias in our estimate. This error has been shown to be mean reverting (Kane et al., 1999; Hertz, 2003; and Keswell and Poswell, 2004), since individuals with lower education levels are more likely to overstate their years of schooling, while those with higher education level are more likely to understate. However, according to Kane et al. (1999), individuals with complete college education are more likely to correctly report their schooling than individuals with less than college education. Given that we are using the panel structure of the data, and our variable of interest is the district council share of college graduates, our estimates of external returns to schooling are less likely to have any bias from mismeasurement of years of schooling.

Potential experience proxied by age less years of schooling and six years, is commonly used in the analysis of returns to schooling. Empirical evidence shows that it has positive, and significant effect on earnings. However, in the presence of selection, this proxy is a poor measure of job tenure, and skills acquired at the job among other factors it is usually meant to represent. Further, South Africa's history of apartheid means that a substantial portion of the population that was formerly disadvantaged may not have started school at age seven, but are more likely to have started school

at an older age. Also, they could have taken longer to attain their highest education if they either repeated a grade (considering the high grade repetition rate in the country), or if they were affected by the struggle against apartheid when schools were intermittent due to strikes. Evidence that the young and inexperienced South Africans' earn more relative to the older and experienced South Africans' (Mwabu and Schultz, 2000) seem to support this argument. In light of this, and given that we have a direct measure of tenure, we opt to exclude experience, but include tenure and its quadratic in our analysis.

3.4.4 Descriptive statistics

Table 3.1 presents the summary statistics of the estimation samples separately for each wave. The average years of schooling is about 11, indicating that the majority of workers have attained some secondary (or equivalent) education. The proportion of individuals in a district council with college education is at most, 11%. The average monthly earnings range between R4,700-6,100, and the average tenure is 7 years. The majority of the respondents across the waves are African, male, married, do not hold a second job, are not trade union members, have a written employment contract, and live in urban areas. A higher proportion live in Gauteng province, followed by the Western Cape and Kwa-Zulu Natal provinces. The proportion of the unemployed and discouraged workers by district council is at least twice as high as the proportion of workers unemployed but actively seeking employment. Most of the workers by district council are Africans, employed in semiskilled sectors, and working in non-manufacturing industry. There is a gender balance in the share of workers by district council.

Table 3.1: Summary statistics

	Wave 1(2008)	Wave 2(2010)	Wave 3(2012)
Measures of human capital			
District council college share	0.0846	0.0909	0.109
Years of schooling	10.86	11.16	11.26
	(0.088)	(0.112)	(0.088)
Individual characteristics			
Monthly wages(nominal Rand)	4728.1	5509.5	6130.0
	(212.9)	(318.6)	(300.4)
Tenure (years)	7.247	7.398	7.143
	(0.216)	(0.276)	(0.243)
Gender (Female)	0.433	0.433	0.449
Marital Status			
Married	0.426	0.416	0.383
Living with partner	0.105	0.100	0.079
Widow/widower	0.033	0.030	0.024
Divorced/separated	0.061	0.047	0.045
Never married	0.375	0.410	0.468
Population group			
African	0.712	0.706	0.733
Coloured	0.099	0.110	0.108
Indian/Asian	0.034	0.033	0.031
White	0.155	0.151	0.128
Home location(Urban)	0.766	0.771	0.774
Home province			
Western Cape	0.126	0.138	0.140
Eastern Cape	0.082	0.085	0.080
Northern Cape	0.021	0.023	0.022
Free State	0.055	0.050	0.058
KwaZulu-Natal	0.126	0.136	0.146
North West	0.091	0.091	0.081
Gauteng	0.337	0.336	0.337
Mpumalanga	0.087	0.072	0.070
Limpopo	0.075	0.070	0.067
Employment contract(written)	0.683	0.749	0.781
Trade union member	0.349	0.328	0.314
Pension deduction	0.499	0.466	0.445
Medical deduction	0.266	0.276	0.254
Unemployment insurance fund deduction	0.656	0.661	0.621
Holds a second job	0.017	0.013	0.009
District council characteristics			
Share of unemployed but searching	0.086	0.080	0.060
Share of unemployed and discouraged	0.219	0.200	0.265
Share in government	0.052	0.044	0.090
Share in Professional	0.144	0.137	0.079
Share in Technician	0.049	0.048	0.042
Share in semiskilled	0.533	0.475	0.505
Share in unskilled	0.222	0.293	0.284
Share in manufacturing sector	0.163	0.113	0.094
Share of female	0.495	0.499	0.506
Share of Africans	0.778	0.793	0.812
Share of Coloureds	0.109	0.115	0.108
Share of Indians/Asians	0.028	0.023	0.020
Share of Whites	0.086	0.080	0.060
Observations	**3089**	**3273**	**4184**

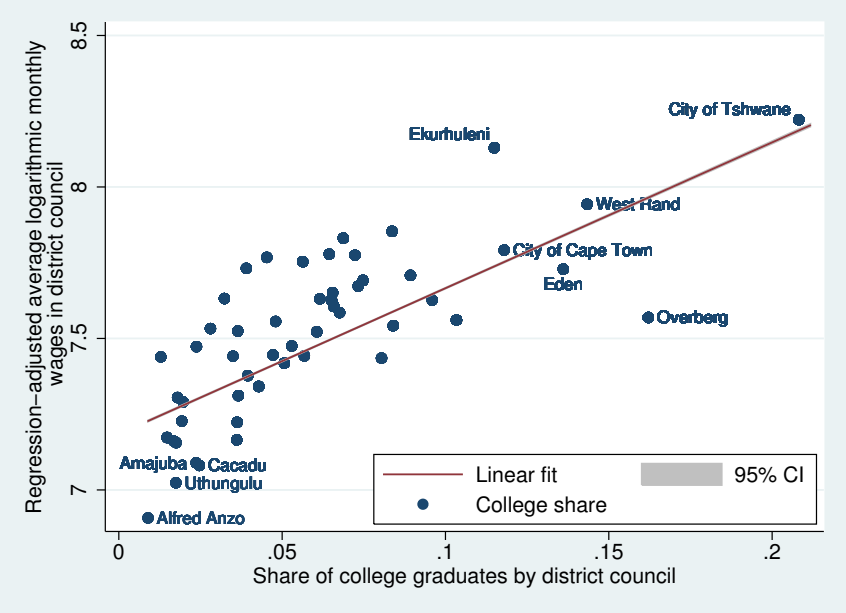
Notes: *(1)* Post stratification weights applied. *(2)* Standard deviations of continuous variables are in parenthesis. *(3)* The sample is of working population (age 15-64) in NIDS. Source: Authors calculation based on the three waves of NIDS.

In Figure 3.4, we present the correlation[9] between wages and the district council share of college graduates. In the figure, we specifically plot the district council share of college graduates against regression-adjusted logarithmic monthly wage that is obtained by holding constant indi-

[9]We are aware that the regression estimates could be possibly endogeneous due to apartheid spatial planning, and we deal with this concern by including district council fixed effects in the regressions presented later in the text. Here our main interest is show simple correlation without implying identification of the estimates.

vidual education, gender, race, marital status, tenure, home location, home province, trade union membership, whether an individual holds a second job, and whether or not an individual has medical aid, pension, and unemployment insurance fund deductions[10]. The main result from the figure is that, even after controlling for individual education, wages are higher in district councils with a higher share of college graduates such as the City of Tshwane (this includes Pretoria which is the executive (administrative) and de facto national capital). The City of Tshwane is ranked in the top (5^{th}) socio-economic quintile (Massyn et al., 2014). Similarly, wages are lower in district councils with a lower share of college graduates such as, Uthungulu and Umkhanyakude. These two district councils are ranked in the bottom two (2^{nd} and 1^{st}, respectively) socio-economic quintiles (Massyn et al., 2014). Although this correlation gives no evidence of human capital externalities, it is an indication that, variation in wages by district council is influenced by characteristics–both observed and unobserved–of workers and district councils that are potentially correlated with district council share of college graduates. Given this marked correlation, we empirically analyze its robustness in the next section.

Figure 3.4: Correlation between regression-adjusted average wages and the share of college graduates in 53 district councils, in Wave 1.



Notes: Regression-adjusted average wage obtained by holding constant individual education, gender, race, marital status, tenure, home location, home province, trade union membership, whether an individual holds a second job, and whether or not an individual has medical aid, pension, and unemployment insurance fund deductions. Post stratification weight is applied. The sample is of individuals aged 15-64 who earn a monthly wage. Both Waves 2 and 3 show a similar correlation.

[10]We get a similar graph when we control for district council fixed effects.

3.5 Empirical results

3.5.1 External returns to college education

In this section, we present estimates of external returns from equation 3.7 for a number of regressions. The results are presented in Table 3.2. In the table, each column represents a different regression with clustered standard errors (clustered at district council level). Column I and II give Ordinary Least Square (OLS) estimates for private returns, and both private and social returns respectively, while disregarding the panel nature of the data. The column I coefficient on years of schooling is in line with existing empirical evidence on private returns to education. It indicates that a one year increase in schooling raises wages by 10%[11]. Similarly, all other controls typically included in the estimation of private returns are consistent with findings in the literature. Women earn lower wages than men by 22%, and being married led to higher earning potential. Race is still a significant determinant of wages in post-apartheid South Africa. From the results, Africans earn significantly lower wages than all other population groups: Coloureds, Indians and Whites earn respectively, 10, 67 and 95% higher wages than Africans. Whilst a higher tenure raises wages by about 2%, earnings start to decline after 30 years of working. Workers on written contract earn higher wages, and a similar effect is observed for workers with trade union membership. Workers who have pension, medical, and unemployment insurance deductions earn higher wages. Although a very small percentage(1%) of our sample hold a second job, holding a second job raises wages for these workers by 43%. Workers in urban areas earn 16% higher wages than workers in rural areas.

As expected, we find the district rate of unemployment is an important determinant of wages. Differentiating the share of unemployed but searching workers from that of the unemployed and discouraged workers shows their conflicting effects on wages. From the results, an increase in the district share of workers unemployed but actively seeking employment reduces wages, while an increase in the district share of unemployed and discouraged workers raises wages. An increase in actively job searching unemployed workers would increase the supply of workers available, and is therefore likely to decrease wages offered; an opposite effect is expected for discouraged unemployed workers. An increase in the district council share of workers employed in the professional sector and in semi-skilled sectors lowers wages, relative to an increase in the share of workers in the elementary sector, which could be associated with a supply effect for skilled workers. An increase in the district share of workers employed in the manufacturing industry lowers wages relative to an increase in all other industries. An increase in the district share of White workers reduces wages, but an increase in the shares of Coloureds and Indians have no effect on individual wages relative to the share of Africans. The sign and level of significance of these controls are consistent in all other regressions.

[11] A regression without district council characteristics gives an almost identical estimate of private returns: a likely indication that these observed controls have minimal effects on private returns.

Table 3.2: The effect of district council share of college graduates on wages

Dependent variable:	(1)	(2)	(3)	(4)
Logarithm of monthly wages	OLS	OLS	Panel	IV
Years of schooling	0.1017***	0.1010***	0.0355**	0.0342*
	(0.004)	(0.004)	(0.017)	(0.018)
College share		3.7341***	5.9755***	6.3946***
		(0.671)	(0.773)	(1.842)
Gender (Female)	-0.2498***	-0.2496***		
	(0.015)	(0.015)		
Population group(African)				
Coloured	0.0967**	0.0958**		
	(0.040)	(0.040)		
Indian/Asian	0.5114***	0.5099***		
	(0.081)	(0.082)		
White	0.6668***	0.6660***		
	(0.041)	(0.042)		
Marital status (married)				
Living with partner	-0.1928***	-0.1974***	-0.0550	-0.0628
	(0.027)	(0.027)	(0.050)	(0.051)
Widow/widower	-0.0606	-0.0612	0.0218	0.0336
	(0.041)	(0.041)	(0.072)	(0.073)
Divorced/separated	-0.0639	-0.0650	0.1181	0.1169
	(0.044)	(0.044)	(0.076)	(0.077)
Never married	-0.1983***	-0.2034***	-0.0054	-0.0138
	(0.019)	(0.019)	(0.049)	(0.050)
Tenure	0.0238***	0.0238***	0.0118**	0.0119**
	(0.003)	(0.003)	(0.005)	(0.005)
Tenure squared/100	-0.0003***	-0.0003***	-0.0003**	-0.0003**
	(0.000)	(0.000)	(0.000)	(0.000)
Employment contract type(written)	0.2942***	0.2872***	0.1356***	0.1338***
	(0.018)	(0.018)	(0.027)	(0.027)
Trade union member	0.1580***	0.1602***	0.0843**	0.0844**
	(0.023)	(0.023)	(0.034)	(0.034)
Pension deduction	0.2578***	0.2633***	0.0810**	0.0776**
	(0.023)	(0.023)	(0.032)	(0.032)
Medical deduction	0.3539***	0.3520***	0.0181	0.0179
	(0.025)	(0.025)	(0.037)	(0.038)
Unemployment insurance fund deduction	0.0635***	0.0621***	0.0458	0.0425
	(0.018)	(0.018)	(0.029)	(0.029)
Holds a second job	0.3575***	0.3585***	0.2220***	0.2215***
	(0.068)	(0.068)	(0.083)	(0.082)
Home location(Urban)	0.1500***	0.1515***	-0.0086	-0.0056
	(0.026)	(0.026)	(0.085)	(0.085)

Notes: *(1)* Each column is an independent regression. *(2)* In all regressions cluster corrected standard errors are in parenthesis; ***, **, and * indicate significance at 1%, 5%, and 10% level respectively. *(3)* Sample of working population (age 15-64) in NIDS. Source: Authors calculation based on the three waves of NIDS.

Continued next page

Table 3.2 continued:

Dependent variable:	(1)	(2)	(3)	(4)
Logarithm of monthly wages	OLS	OLS	Panel	IV
District council-specific characteristics				
Unemployment				
Share of unemployed but searching	-0.7784***	-0.5465***	-0.9014***	-0.7976***
	(0.203)	(0.198)	(0.228)	(0.231)
Share of unemployed and discouraged	0.7750***	0.7018***	0.5287***	0.5297***
	(0.138)	(0.134)	(0.157)	(0.156)
Sector of employment				
Share of government	0.1829	0.0070	0.1902	0.1504
	(0.208)	(0.217)	(0.288)	(0.275)
Share of Professional	-0.8769***	-0.7369***	-0.9421***	-0.9867***
	(0.151)	(0.146)	(0.172)	(0.189)
Share of Technician	-0.1027	-0.0448	0.1245	0.2876
	(0.287)	(0.277)	(0.316)	(0.323)
Share of semiskilled worker	-0.3499***	-0.3448***	-0.4437***	-0.4718***
	(0.098)	(0.094)	(0.100)	(0.101)
Industry of employment				
Share of manufacturing sector	-0.8697***	-0.7646***	-0.8215***	-0.8059***
	(0.190)	(0.182)	(0.234)	(0.234)
Population characteristics				
Share of female	0.3773	-0.7094	0.2808	-0.2433
	(0.690)	(0.633)	(0.716)	(0.874)
Share of Coloureds	0.0184	0.2265	0.6974	0.6174
	(0.722)	(0.685)	(0.668)	(0.681)
Share of Indian/Asian	2.0509	1.0132	-0.2838	-0.4451
	(1.822)	(1.857)	(2.094)	(1.982)
Share of White	-3.2537***	-2.4706***	-2.4363**	-2.3899**
	(0.777)	(0.839)	(1.033)	(1.217)
Province dummies	Yes	Yes	Yes	Yes
District council fixed effects	Yes	Yes	Yes	Yes
Individual fixed effects	No	No	Yes	Yes
Constant	6.7874***	6.6290***	6.6193***	
	(0.721)	(0.672)	(0.800)	
Observations	10818	10818	10818	6319
Number of Groups			7091	2704
Adjusted R-Square	0.561	0.563	0.172	-0.499

Notes: *(1)* Each column is an independent regression. *(2)* In all regressions cluster corrected standard errors are in parenthesis; ***, **, and * indicate significance at 1%, 5%, and 10% level respectively. *(3)* Sample of working population (age 15-64) in NIDS. Source: Authors calculation based on the three waves of NIDS.

In column 2, the OLS estimates, in addition to years of schooling, include the district council share of college graduates. The estimate of private return is very similar to that in column 1, an indication that district council human capital does not affect private returns. The estimate of the district council share of college graduates gives strong evidence of human capital externalities over and above private returns to education. It suggests that a 1% increase in the district council share of college graduates raises wages by 3.73%, an indication that there are positive and large external returns to education. All individual level controls are very similar to those in column I, however, although district council controls maintain their signs and significance, they are on average, slightly lower.

In column 3, we control for the sorting of workers by unobservable characteristics–by including both district and individual fixed effects. There are two sources of identification in this specification. First, individuals may move from one district to another, and second, there is bound to be variation in the share of college graduates in a district over time (see Moretti, 2004a and Moretti, 2004b for a discussion). The coefficient of the share of college graduates increases by about 2 percentage points, an indication that an OLS regression on pooled data understates the effect of human capital externalities. This can be inferred to imply that, a worker's unobserved individual characteristics have a significant part to play in the determination of the extent to which the worker benefits from human capital externalities. The coefficient for private returns decreases to 3.6%, and is lower in precision. This suggests that there are significant unobservable characteristics in the determination of private returns. However, these individual and district council fixed effects estimates are still likely to be biased by time-varying factors, we consider an instrumental variable estimation next.

The inclusion of both district and individual fixed effects, as discussed above, offers a partial solution to the problem of endogeneity. This is because the coefficient of the district council share of college graduates is still likely to be biased by unobserved district council factors that change over time, and have significant effect of wages, such as local supply and demand shocks of college graduates. We therefore use the annual average maximum temperature, annual average minimum temperature, and the annual average rainfall to instrument for the district council share of college graduates. A reduced form regression of the instruments on the district share of college graduates indicates a causal relation with an R^2 of about 91.7%. An F-test of their joint relevance as indicated by the AP F-statistic is significant at the 1% level, and passes the Staiger and Stock (1997) rule of thumb ($F \geq 10$) for weak instruments. The Hansen J-tests of over-identification fail to reject the null hypothesis (F-stat=8.20 and p-value=0.19), an indication that the instruments are valid. We present the first-stage results in Table 3.3.

The results from the second stage are presented in column 4 of Table 3.2. To control for unobserved heterogeneity, both district council and individual fixed effects are included. From the results, a 1% increase in the district council share of college graduates increases external returns to education by 6.33%; a 0.16 percentage points increase from the fixed effects model without instruments in column 3. This is an indication that district council-specific supply and demand changes in college workers have a downward bias on estimates of external returns. This result shows that there are positive and large external returns to education, these returns are larger than private returns, and failure to control for individual fixed effects overstates private returns but understates external returns. Our estimates of external returns are relatively larger when compared with estimates from other developing and transition economies for instance, 0.77-1.75% for Turkey (Filiztekin, 2011), and about 1.33% for Russia (Muravyev, 2008). They are also larger than those from developed countries such as 1.3% for the U.S. (Moretti, 2004b), and 1.8% for Germany (Heuermann, 2011).

Table 3.3: IV-estimates: first stage

Dependent variable: District council share of college graduates	
Instruments	IV
Annual maximum temperature(average)	-0.6536***
	(0.0954)
Annual maximum temperature(average) Squared	0.0270***
	(0.0039)
Annual maximum temperature(average) cubed	-0.0004***
	(0.00005)
annual minimum temperature(average)	-0.0082***
	(0.0011)
Average annual rainfall	0.0002
	(0.0002)
Average annual rainfall squared	-0.000002*
	(0.000001)
Individual characteristics	Yes
District council characteristics	Yes
Province dummies	Yes
District council fixed effects	Yes
Individual fixed effects	Yes
Observations	6319
Adjusted R-squared	0.917
Number of groups	2704
F-statistic (model)	43.99
AP F-statistic(for excluded instruments)	29.68
AP p-value	0.0000

Notes: *(1)* Standard errors are in parenthesis; ***, **, and * indicate significance at 1%, 5%, and 10% level respectively. *(2)* AP F-statistic is the Angrist-Pischke F-statistic on excluded instruments in the first stage, and the AP p-value is the probability value in the Angrist-Pischke test for weak instruments. *(3)* Coefficient for constant are not reported. Source: Authors calculation based on the three waves of NIDS.

3.5.1.1 Robustness checks

To check for robustness of the results from the instrumental variable regression results in column 5 of Table 3.2, we re-estimate equation 3.7 in several specifications as presented in Table 3.4. First, we consider different definitions of human capital–for both individual and aggregate human capital. Most studies, in estimating human capital externalities, have considered the linear effects of years of schooling on wages. A few recent studies have considered nonlinear private returns by considering different years of schooling, and their results on external returns are conflicting. In two separate regressions–one on nonlinear years of education and another on college dummy– Moretti (2004b) finds there are nonlinear effects. However, Heuermann (2011), using quadratic specifications, finds no evidence of nonlinear effects. Following Moretti (2004a), we consider two specifications: a nonlinear years of schooling specification and a college dummy specification. The results are presented in columns 1 and 2 of Table 3.4. The coefficient for district council college share indicates that considering private returns in nonlinear form hardly affects the estimate and precision of external returns, however, the estimates from the college dummy are slightly lower. Acemoglu and Angrist (2001) argue that externalities from secondary education are as important as those from college education, particularly for developing countries where the bulk of human-capital accumulation is in secondary education, which is the case for South Africa. To check that using the share of college graduates is an appropriate measure for human capital externalities in South Africa, we use district council average years of education as our measure of external returns.

Table 3.4: IV-estimates: Robustness checks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Second stage										
College share	6.356***	6.192***		3.169	8.425***	-5.711*	6.907**	0.826	7.291***	5.255***
	(1.806)	(1.812)		(2.259)	(2.317)	(2.989)	(1.884)	(3.158)	(1.960)	(2.017)
Average years of schooling			0.3054**							
			(0.152)							
Instrumental variables(First stage)										
Annual maximum temperature(average)	-0.6539***	-0.6559***	-10.3976***	-0.6875***	-0.5920***	-2.0583***	-0.6329***	-0.6616***	-0.6884***	-0.5255***
	(0.0951)	(0.0964)	(1.4040)	(0.1087)	(0.0979)	(0.4946)	(0.0939)	(0.1760)	(0.1274)	(0.1038)
Annual maximum temperature(average) Squared	0.0270***	0.0271***	0.4178***	0.0282***	0.0246***	0.0845***	0.0261***	0.0274***	0.02843***	0.0219***
	(0.0039)	(0.0039)	(0.0555)	(0.0044)	(0.0040)	(0.0195)	(0.0038)	(0.0073)	(0.0051)	(0.0042)
Annual maximum temperature(average) cubed	-0.00037***	-0.00037***	-0.0056***	-0.0004***	-0.00034***	-0.0011***	-0.0004***	-0.0004***	-0.0004***	-0.0003***
	(0.00005)	(0.00005)	(0.0007)	(0.00006)	(0.00005)	(0.0003)	(0.00005)	(0.0001)	(0.00007)	(0.00006)
Annual minimum temperature(average)	-0.0081***	-0.0082***		-0.0074***	-0.0088***	-0.0046	-0.0083***	-0.0085***	-0.0078***	-0.0094***
	(0.0010)	(0.0011)		(0.0011)	(0.0011)	(0.0031)	(0.0011)	(0.0015)	(0.0013)	(0.0010)
Average annual rainfall	0.0002	0.0002	-0.0083*	0.00032**	0.00012	0.0007	0.00015	0.00001	0.0005**	0.000021
	(0.0001)	(0.0001)	(0.0049)	(0.00015)	(0.00015)	(0.0007)	(0.0002)	(0.0002)	(0.0002)	(0.00013)
Average annual rainfall squared	-0.000002*	-0.000002*	0.00017**	-0.000002**	-0.000001	-0.000005	-0.000001	-0.000000	-0.000005**	-0.000000
	(0.000001)	(0.000001)	(0.0001)	(0.000001)	(0.000001)	(0.000006)	(0.000001)	(0.000001)	(0.000002)	(0.000001)
Average annual rainfall cubed			-0.000001**							
			(0.000000)							
Controls included										
Individual characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District council characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	6319	6331	6319	3169	3150	409	5910	2026	4088	4949
Number of groups	2704	2709	2704	1349	1355	177	2527	874	1745	2116
Adjusted R-Squared (first stage)	0.923	0.923	0.965	0.925	0.921	0.937	0.922	0.911	0.890	0.923
AP F-stat. of excluded instruments	17.58	28.58	29.63	22.61	28.46	8.26	28.92	13.3	19.14	21.41
AP p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: *(1)* Each column is an independent regression. In all regressions cluster corrected standard errors are in parenthesis; ***, **, and * indicate significance at 1%, 5%, and 10% level respectively. *(2)* AP F-statistic is the Angrist-Pischke F-statistic on excluded instruments in the first stage, and the AP p-value is the probability value in the Angrist-Pischke test for weak instruments. *(3)* Coefficient for constant are not reported. *(4)* Similar controls as in Table **2** are included. Source: Authors calculation based on the three waves of NIDS

The results are presented in column 3, and they indicate that there are positive external returns to an increase in the district council average years of schooling. However, the returns are less precisely measured.

Next, we consider the effect of district council college share in sub-samples by gender, race, and geographical location of the district council, that is, rural or urban. The results on gender indicate that while there are no external returns to women, a 1% increase in the district council share of college graduates' raises external returns to men by 8.43%; a 2 percentage points higher effect than the effect in the overall sample. Considering the number of observations for men and women samples are equal, this is an interesting finding, and can be inferred to imply that an increase in the share of college graduates benefits men more than it benefits women. This is likely to be associated with a possibility that, a higher proportion of men hold highly skilled jobs, where human externalities are more likely to be realized, and the contrary is the case for women.

Given the South African history of apartheid, sample estimation by population grouping/race is relevant. We group the formerly disadvantaged groups together as Black—Africans, Coloureds, and Indians. We find that an increase in the district council share of college graduates' reduces wages for Whites but increases wages for Blacks. Considering the Whites were previously advantaged in wages and educational attainment, the negative coefficient on the White sample could be an indication that an increase in the district council share of college graduates is characterized by an increase in the proportion of nonwhites with college education. We infer the positive effect on wages for Blacks to be a reinforcement of the preceding argument, and given that the majority of the individuals in our sample are Black, the increase in the district council share of college graduates could be viewed as synonymous with an increase in the average share of Blacks with college education, and is therefore in agreement with Michaud and Vencatachellum (2003) who find the the increase in supply of workers of one race did not dominate the demand for workers of that race.

The geographical location specifications in column 8 and 9 indicate that there are no external returns to workers in rural areas, but there are large and significant external returns to workers in urban areas. Our final estimation is of a sub-sample that excludes individuals residing in cities included in our sample, namely: City of Cape Town, City of Johannesburg, City of Tshwane, Ekurhuleni, Ethekwini, and Nelson Mandela Bay. This is because these districts are likely to have outliers in several dimensions such as educational attainment, size, public facilities and services, and amenities. Therefore, they are potentially likely to be driving our estimates. In column 10, the estimate of external returns is significant but about 1 percentage point smaller than the estimate including the cities. This is an indication that the effect of these cities is marginal, and that there are external returns to education even in smaller district councils. In view of the above evidence, we conclude that our estimate of the external returns from college education are robust. We next consider externalities from college graduates.

3.5.2 Externalities from college education

The above estimates of external returns give strong evidence that there are external returns to education in South Africa. They, however, do not show whether there are externalities (spillover effects) from the district council share of college graduates. To be able to identify the spillovers, and given our sample size, we divide our sample by education level into two groups: workers with less than a college education and workers with at least some college education. Table 3.5 presents estimates of the effects of changes in the district council share of graduates on each group. The first stage estimates are also presented in the lower panel of the table. Reduced form regressions of the instruments on the district council share of college graduates indicate a causal relation for each group with an R^2 of about 92% for both. F-tests of their joint relevance as indicated by the AP F-statistic are significant at 1% level, and they pass the Staiger and Stock (1997) rule of thumb ($F \geq 10$) for weak instruments.

The estimates from the table indicate that there are spillover effects. A 1% increase in the district council share of college graduates increases wages for workers with less than college education by 5.9%, and for workers with at least some college education by 7.6%. These effects are, however, contrary to the theory (Moretti, 2004a) predicting that, the effect should be larger for lower skilled workers than for higher skilled workers. Although these effects differ from those by Moretti (2004b) for U.S., they are similar to the findings by Heuermann (2011) for Germany. In our view, the spillover effects are appropriate for South Africa based on several reasons. The existing skills premium implies that highly skilled workers earn higher wages, and therefore an increase in workers with higher skills is unlikely to lower wages while there is still demand for highly skilled workers. This means that without the neoclassical supply effect on wages, we are likely to be capturing a pure spillover effect in the district council share of college graduate estimate, on the group with at least some college education. It is also likely the case that the wages of workers with less than college education are prevented from adjusting to the market conditions by trade union wage setting. The level of unionization is higher amongst low skilled workers in South Africa. Heuermann (2011), who finds similar effects, argues that this is likely to be the case in Germany. In the appendix we present a Table 3.7 which show how the coefficients vary as we added controls. From the table, the coefficient of college share to be positive and highly significant an indication that it is robust.

Table 3.5: The effect of college share on different education groups

	Less than college	At least some college
	(1)	(2)
	Monthly Wage (in log)	Monthly Wage (in log)
Second stage		
College share	5.950116***	7.825675***
	(2.104125)	(2.960854)
First stage (Instrumental variables)		
Maximum Temperature (T)	-0.644731***	-0.713140***
	(0.095870)	(0.146910)
Maximum Temperature^2 (T^2)	0.026655***	0.029342***
	(0.003891)	(0.005916)
Maximum Temperature^3 (T^3)	-0.000364***	-0.000399***
	(0.000053)	(0.000079)
Minimum Temperature (T)	-0.008097***	-0.008664***
	(0.001031)	(0.001774)
Rainfall	0.000147	0.000456**
	(0.000143)	(0.000216)
Rainfall^2	-0.000001	-0.000003**
	(0.000001)	(0.000001)
Individual characteristics	Yes	Yes
District Council characteristics	Yes	Yes
Province dummies	Yes	Yes
District council fixed effects	Yes	Yes
Individual fixed effects	Yes	Yes
Observations	4347	1582
Adjusted R-squared	-0.542	-0.570
Number of groups	1890	673
F-statistic (model)	15.00	11.24
AP F-statistic (for excluded instruments)	27.67	13.84
AP p-value	0.000	0.000

Notes: (1) Standard errors are in parenthesis; ***, **, and * indicate significance at 1%, 5%, and 10% level respectively.
(2) AP F-statistic is the Angrist-Pischke F-statistic on excluded instruments in the first stage, and the AP p-value is the probability value in the Angrist-Pischke test for weak instruments.
(3) Coefficient for constant are not reported. Source: Authors calculation based on the three waves of NIDS.

3.6 Summary and conclusion

Using a panel of three waves of the National Income Dynamics Study, we focus our analysis on the estimation of external returns to education and spillover effects from college graduates. Using a Mincerian regression augmented with the district council share of college graduates, we analyze a sample of working individuals aged 15-64 years in full-time or part-time employment and with some formal education. Using the average annual maximum temperature, the average annual minimum temperature, and the average annual rainfall to instrument for the district council share of college graduates, we find that a 1% increase in the district council share of college graduates increases wages of workers by 6.4%. These effects are fairly robust to different definitions of individual education and aggregate education, and also to several sample specifications.

Further, our estimation of externalities indicate that there are large spillover effects from an increase in the college graduates. Interestingly, we find that contrary to theory, workers who are college graduates benefit more from an increase in the share of college graduates than workers with less than college education. We infer this to be either a likely consequence of the high skills deficit which leads to a no supply effect on wages of highly skilled workers, or a consequence of unionization, resulting in low skill wages not responding to market changes, or both.

Women earn lower wages than men and being married led to higher earning potential. Race is still a significant determinant of wages in post-apartheid South Africa. A higher tenure raises wages, workers on written contract earn higher wages, and a similar effect is observed for workers with trade union membership. Workers who have pension, medical, and unemployment insurance deductions earn higher wages. Workers in urban areas higher wages than workers in rural areas.

The district rate of unemployment is an important determinant of wages. An increase in the district council share of workers employed in the professional sector and in semi-skilled sectors lowers wages, while an increase in the district share of workers employed in the manufacturing industry lowers wages. An increase in the district share of White workers reduces wages.

Appendix 1: Robustness

Table 3.7: The effect of college share on different education groups by controls

	Less than college		At least some college	
	(1) Monthly Wage (in log)	(2) Monthly Wage (in log)	(3) Monthly Wage (in log)	(4) Monthly Wage (in log)
Second stage				
College share	6.957466***	6.931167***	11.147658***	8.092403**
	(2.123180)	(2.068263)	(3.014294)	(3.174891)
First stage (Instrumental variables)				
Maximum Temperature (T)	-0.639713***	-0.671391***	-0.711923***	-0.740315***
	(0.125373)	(0.100142)	(0.184968)	(0.144882)
Maximum Temperature^2 (T^2)	0.026251***	0.027626***	0.029298***	0.030450***
	(0.005144)	(0.004052)	(0.007587)	(0.005838)
Maximum Temperature^3 (T^3)	-0.000355***	-0.000376***	-0.000397***	-0.000414***
	(0.000070)	(0.000055)	(0.000103)	(0.000078)
Minimum Temperature (T)	-0.009701***	-0.008207***	-0.010381***	-0.008304***
	(0.001134)	(0.001040)	(0.002221)	(0.001748)
Rainfall	0.000201	0.000151	0.000379*	0.000468**
	(0.000150)	(0.000145)	(0.000201)	(0.000215)
Rainfall^2	-0.000001	-0.000001	-0.000002	-0.000003**
	(0.000001)	(0.000001)	(0.000001)	(0.000001)
Individual characteristics	Yes	No	Yes	No
District Council characteristics	No	Yes	No	Yes
Province dummies	Yes	Yes	Yes	Yes
District council fixed effects	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	Yes	Yes	Yes
Observations	4347	4890	1582	1752
Adjusted R-squared	-0.622	-0.565	-0.652	-0.569
Number of groups	1890	2094	673	737
F-statistic (model)	11.17	24.57	112.5	19.13
AP F-statistic (for excluded instruments)	33.61	25.57	16.56	14.12
AP p-value	0.000	0.000	0.000	0.000

Notes: (1) Standard errors are in parenthesis; ***, **, and * indicate significance at 1%, 5%, and 10% level respectively.

(2) AP F-statistic is the Angrist-Pischke F-statistic on excluded instruments in the first stage, and the AP p-value is the probability value in the Angrist-Pischke test for weak instruments.

(3) Coefficient for constant are not reported. Source: Authors calculation based on the three waves of NIDS.

Chapter 4

Unemployment Duration and Exit States by Gender in South Africa

4.1 Introduction

Most studies of unemployment duration consider either the probability of entering unemployment or the probability of exiting unemployment. These studies therefore explicitly assume that an unemployed individual has one option, to enter employment. The reality is that there are multiple options. An individual can exit unemployment into self-employment, or become economically inactive (by either going back to school or by simply being out of the labour market). Considering these other exit states enriches the characterisation of the unemployed and would be useful in advising policy for dealing with unemployment. In this study, we estimate the determinants of the probability of leaving unemployment into either employment or becoming economically inactive, and focus on the variation in the determinants by gender. A number of studies have considered exit states by gender (such as Böheim and Taylor, 2000; Gonzalo and Saarela, 2000; Tansel and Tasci, 2004; Lentz and Tranæs, 2005; and Ollikainen, 2003). Others have controlled for gender by including a gender dummy variable (such as Dendir, 2006; Galiani and Hopenhayn, 2003; and Grogan and Van den Berg, 2001). Studies that control for gender using a dummy variable disregard the possibility that the effects of some of the other variables may vary by gender. Given the existing evidence on gender differences in the labour market, we examine gender factors that differentiate unemployment exits into employment or economic inactivity in South Africa.

Unemployment spells differ for men and women, and similarly, factors that influence duration of transit from unemployment into one exit state or another also vary by gender. The gender differences in transition into either employment or economic inactivity have to some degree been associated with traditional gender roles. For instance, on the one hand, a short transition into employment for men has been linked to men being compelled to take a job as soon as an offer is received, since they are expected to provide for a family. On the other hand, women may transit into economic inactivity or into part-time employment because they are expected to be at home

more often to take care of children (Ollikainen, 2003). In addition, Gonzalo and Saarela (2000) argue that men are traditionally more attached to the labour market. The odds of finding a job are against women with children, which they infer to be an indication of gender discrimination. Affirmative action however, now provides more employment opportunities for women, and policies have been introduced that are more supportive of working women with children, such as the Basic Conditions of Employment Act (BCEA) of 1997 (also amended in 2002). Consequently, there has been an increase in labour force participation amongst women in the last few decades (Kingdon and Knight, 2007). Understanding the gender-related factors that differentiate the transition of unemployed South Africans into employment or economic inactivity is a key policy issue, and in its own right is a valid and important research question.

The level of unemployment in post-apartheid South Africa has been persistently high, and the estimation of the relationship between unemployment duration and the probability of finding a job, that is, duration dependence, is important. In the last decade, the unemployment rate by narrow definition(which is the official definition) has averaged 25%. The unemployment rate (broad definition) was 35% in 2013, which is a decline from 39% in 2003 (Statistics South Africa, 2013). The unemployment rate (narrow definition) amongst women was higher than the average rate by 3 and 2 percentage points in 2003 and 2013 respectively, and it was higher than the men's rate by 4 and 6 percentage points in 2003 and 2013 respectively. This high and persistent unemployment is a major problem for the country. Unemployment has been associated with both economic and social costs (Böheim and Taylor, 2000 ; Ollikainen, 2003). As Arulampalam and Stewart (1995) argue, in addition to income losses, unemployment causes misery for the unemployed.

In addition to the problem of a high unemployment, Wittenberg (2002) finds South Africa has a long-term labour absorption problem which largely affects African men. He also finds that the country has "a severe short-term unemployment 'spike"' which largely affects African youth Wittenberg (2002). He concludes that although the majority of these African youth are likely to transit into employment, some are likely to become discouraged, while others transit into non-participation. Given the high unemployment rate and long unemployment duration, it is likely that the unemployed have pessimistic expectations of finding work (Kingdon and Knight, 2006). A study of the determinants of unemployment duration by multiple exits therefore facilitates the identification of characteristics that distinguish individuals who transit into employment from those who transit into economic inactivity.

In addition to the differences in unemployment rate by gender, estimation of unemployment duration by competing risks is particularly valuable for South Africa given the existing literature gap. Although empirical evidence indicates that unemployment duration varies by gender, age and race, there is hardly any evidence on competing exits. Wittenberg (2002) finds that while the majority of men aged 30 to 50 are economically inactive, only 20-30% of African women in this age cohort are economically inactive. The share is higher for Indian women, and for White and Coloured women the economically inactive share increases steadily after the mid-twenties

age range. Wittenberg associates these differences with women's reproduction choices. In a job search, the opportunity cost of such a search may vary by gender, among other factors (Kingdon and Knight, 2006). However, Posel et al. (2014) find search status has no effect on chances of employment by gender, age cohort, and location.

Existing studies on unemployment duration for South Africa have mainly focused on exit into employment. This is likely to give an incomplete picture, particularly considering the evidence that individuals are discouraged before they even start looking for a job (Wittenberg, 2002). It is, however, important to consider unemployed individuals as being in transition to either securing a job or out of the labour market (being economically inactive) (Wittenberg, 2002). The consideration of exit from unemployment into economic inactivity is important since it is an alternative that is considered by much of the working age population. Wittenberg (2002) finds that after age 40, the stock of unemployed African men falls, and that of discouraged workers also falls, while that of economically inactive individuals increases.

There is limited literature in this area. However, existing past studies indicate that the transition from school to work in South Africa is difficult (Wittenberg, 2002; Lam et al., 2008) and that unemployment is duration dependent (Brick and Mlatsheni, 2007). Whilst the aggregate unemployment rate is high in South Africa, the estimates for young people are even higher. In 2005, the unemployment rate of the 15-24 age group was 52.9%, and 31% for those aged 25-34 (SSA, 2006). This is a considerably higher youth unemployment rate than that of sub-Saharan Africa as a whole, which averages 21% (Guarcello et al., 2005). Prolonged initial joblessness has been found to permanently impair individuals' productivity and affect their employment prospects and wages (Guarcello et al., 2005). Young adults' duration of unemployment should therefore be of policy concern to any government. In this study, therefore, we interrogate the potential determinants of unemployment duration by competing risks, and estimate and characterise the duration by gender.

The remainder of the study is organized as follows: Section 2.2 reviews the relevant literature, section 2.3 outlines the analytical framework and estimation strategy. A description of the data follows in section 2.4, followed by descriptive statistics in section 2.5. Section 2.6 presents a discussion of empirical results from both the non-parametric and semiparametric approaches. We summarise the results and conclude in the last section.

4.2 Evidence on unemployment duration

The literature on unemployment duration is extensive. Studies of unemployment duration by gender indicate unemployment duration varies between men and women. The difference is particularly interesting when different exits are considered. Studying unemployment exits for the U.K., Böheim and Taylor (2000) find negative duration dependence into full-time employment for both men and women. Similarly, in a study of a local Finnish labour market, Gonzalo and Saarela (2000) find exit into employment had a negative duration dependence for both men and women. However

they find that employers penalise men more than women in offering jobs, which the authors infer to be due to men's higher attachment to the labour market. Gonzalo and Saarela find positive duration for women in non participation, but find no duration dependence in non participation for men. Ollikainen (2003) considers the Finish labour market at the national level, and finds negative duration dependence with men being more likely than women to exit unemployment after one year. He also finds that men have a longer transition into economic inactivity than women. Men and women in the U.K. have the same unemployment duration into employment, however, women have a short unemployment duration when they exit into part-time employment and out of the labour market (Böheim and Taylor 2000). For the Danish labour market, Lentz and Tranæs (2005) find negative duration dependence, but the duration dependence disappears once unobserved heterogeneity is controlled for.

Various factors affect transition from unemployment into different states such as employment (either full-time or part-time), self-employment, and economic inactivity, for women and men. The evidence on the effect of age on exit into employment is conflicting. Ahn and Ugidos-Olazabal (1995) find probability of employment decreases substantially after the age of 45, but the probability of non participation is higher for the younger and older population groups. However, Ahn and Ugidos-Olazabal find age has no effect on the probability of employment for women, but significantly affects their probability of exiting the labour market. The evidence on transition out of unemployment not only varies by gender, but also seems to be inconsistent. For instance, for Finland Gonzalo and Saarela (2000) find a bell-shaped relationship for both women and men. However, Ollikainen (2003) finds transition into employment is bell-shaped for men but it increases for women. Gonzalo and Saarela (2000) find exit into non participation declines with age, and is higher for younger men than women. However, Ollikainen finds transition into economic inactivity to have a bell shape for women, while it increases for men. The two studies, however, find the significance of unemployment duration declines with age. Lentz and Tranæs (2005) finds unemployment duration decreases with age.

Although the evidence on the effect of education is mixed, it generally indicates that educated women have the highest chances in securing employment. Both Ollikainen (2003) and Böheim and Taylor (2000) find that educated women have a shorter transit into full-time employment. Uneducated women have the least transition into employment (Ahn and Ugidos-Olazabal, 1995). However, although Ollikainen (2003) finds education shortens the transition period into employment for both men and women, Böheim and Taylor (2000) find formal education has no significant effect on transition out of unemployment for men.

The effects of family characteristics such as marital status, employment status of spouse, presence of children, and financial status, are significant in influencing differential exit states. Evidence on the effect of marital status is mixed and differs by unemployment exit considered. In a study of unemployment duration in Spain, Ahn and Ugidos-Olazabal (1995) find married women have a lower transition into employment, but a higher transition into non participation than single

women. Gonzalo and Saarela (2000) find marriage significantly influenced employment probability, as married women and men take a shorter time to transit into employment. However, findings from a number of studies differ. Arulampalam and Stewart (1995) find marriage has no effect on the probability of being employed, and has no effect on transition into either full-time or part time employment. Böheim and Taylor (2000) find marital status affects transition into self-employment or economic inactivity, and Ollikainen (2003) finds that it marital status has no effect on transition into economic inactivity.

Having a working spouse has been shown to have different effects. Arulampalam and Stewart (1995) find unemployed men married to working women had a shorter unemployment spell when compared to single and men married to unemployed women. Similarly, Böheim and Taylor (2000) find men with a working spouse have a shorter transition into part-time employment, and the effect on women is also positive but small. Evidence on the effect of presence of children is mixed. Having younger children lengthens unemployment duration for women while it has no effect on unemployment duration for men (Ahn and Ugidos-Olazabal, 1995; Ollikainen, 2003). Both Gonzalo and Saarela (2000), and Böheim and Taylor (2000) find the number of children had no effect on exit into full-time employment for both women and men, but it significantly affects transition into non participation for women (Gonzalo and Saarela, 2000) and transition into part-time employment for men (Böheim and Taylor, 2000). However, Lentz and Tranæs (2005) finds that having children has no effect for both men and women when unobserved heterogeneity is controlled for. Family responsibility is a determinant of transition: Transition into employment for male or female household heads is double that of non-household heads (Ahn and Ugidos-Olazabal, 1995).

Evidence of the effect of unemployment benefits on unemployment duration is vast. The conditions for being a recipient, and the length of time that benefits are accessible varies from country to country. However, most evidence indicates that these benefits have negative effects and there is often a spike in exit towards the end of the benefit period (Ahn and Ugidos-Olazabal, 1995). Ahn and Ugidos-Olazabal (1995) argue that generous family support contributed to the high unemployment in Spain in 1985. For Finland, Ollikainen (2003) finds unemployment benefits have a highly negative effect. Arulampalam and Stewart (1995) find that for UK men, unemployment income influences the probability of employment only after a given threshold.

In most developing countries, unemployment benefits tend to be non-existent, or where they exist, not generous in period of coverage or value. Most unemployed individuals are thus dependent on informal jobs, financial support from family and friends such as spouse's income, family wealth, and non-labour incomes. Both spouse's income and family wealth are expected to have a negative effect on employment probability. Lentz and Tranæs (2005) find wealth has a negative effect on unemployment duration for both men and women. However, spouse's income increases employment probability for men but not for women. There is limited evidence on the effect of non-labour income. Böheim and Taylor (2000) find having non-labour income reduces the probability of full-time employment, but has no effect on part-time employment, self-employment, or

economic inactivity. Location of residence has been shown to affect the probability of finding employment. According to Ollikainen (2003), transition into employment is shorter for men in urban areas, but men and women residing in rural areas have a shorter transition into employment than those residing in urban areas.

Different approaches have been taken in the estimation of unemployment duration. The common practice has been the use of the Kaplan-Meier nonparametric estimator, followed by either a semi-parametric or a parametric empirical specification, or both. Different models are likely to produce different results. Using a split population model and a piecewise constant hazard model, Ollikainen (2003) finds a larger difference in exit probabilities between women and men than from using just the split population model. Ollikainen (2003) argues for the piecewise constant hazard model population, since, in addition to the hazard model, it gives more information.

Controlling for both supply and demand factors is ideal in the estimation of unemployment duration. Most studies, however, fail to include demand factors such as the unemployment rate. Controlling for labour market conditions gives insight into the behaviour of the unemployed (Gonzalo and Saarela, 2000). Böheim and Taylor (2000) find that a high unemployment rate significantly increased the probability of transiting into economic activity for both men and women, and transition into full-time employment for women.

Empirical literature on unemployment duration for transitioning and developing countries, particularly on competing exits by gender, is limited but growing. Most studies find unemployment duration to be long. Dendir (2006) analyses unemployment duration in urban Ethiopia and finds gender and ethnic background have no effect on unemployment duration. However, he finds other factors such as age, education, marital status, location, and support mechanism while unemployed, have significant effects. Dendir also finds long mean unemployment duration (that is, three years for complete spells and 4.7 years for incomplete spells). However, he finds non-constant duration dependence, first negative then positive duration dependence, when he considers parametric and semi-parametric estimations, respectively. Focusing on incidence and duration of unemployment of young men in urban Ethiopia, Serneels (2007) finds long unemployment spells of 45 months , a similar finding to Dendir (2006).

Two other studies that investigate unemployment duration by gender in a developing country are those of Tansel and Tasci (2004), and Du and Dong (2009). In their study of determinants of unemployment duration for men and women separately, Tansel and Tasci find women have longer unemployment durations than men. They also find that the probability of employment for both men and women varies by marital status, age, education level, and residential area unemployment rate. As in Dendir (2006), Tansel and Tasci (2004) find non-constant duration dependence (first negative then positive) with an overall U-shape. They associate the changes with scaring effect that is, unemployment may lead to lower future wages or chances of employment due to lack of work experience and likely deterioration of general skills during the unemployment period. They also link the changes with subsequent decimation of unemployment benefits or diminution of savings

and/or family support. For China, Du and Dong (2009) find women have longer unemployment duration, and market structures and institutional incentives played significant roles in lengthening unemployment duration for women. There is, however, some evidence of short unemployment duration. Galiani and Hopenhayn (2003), investigating this in Argentina, find short unemployment duration when re-incidence is not considered. In an analysis of unemployment duration in Russia, Grogan and Van den Berg (2001) find unemployment spells are significantly shorter for women than men, and highly educated workers have shorter spells. Their results are consistent for different types (definitions) of unemployment, although they find the mean duration is on average short but varies depending on which definition of unemployment is used.

Unemployment rate varies by age, race, gender, location, and education level. Young South Africans have a harder time securing jobs. Regardless of race, it is difficult for 15-24 year olds to find a job (Ranchhod and Dinkelman, 2007). However, it is easier for 25 year olds to find a job than for 30 year olds (Wittenberg, 2002). Wittenberg relates these differences to stigmatisation, whereby an employer sees an unemployed 30 year old as a signal of bad quality in the individual. This could also be a result of skills erosion, where a delay in getting a job compromises future job prospects. Evidence shows African South Africans take longer to secure employment compared to other population groups (Wittenberg, 2002; Ranchhod and Dinkelman, 2007; Lam et al., 2010). Using three large household surveys, Wittenberg (2002) finds that at age 25, 100% of White men have secured employment. He finds that only 70% of African men would have secured a job at age 30, and only 50% of African women would have secured a job at age 50.

Evidence suggests that unemployment duration in South Africa varies by race. In studying duration to first job using data for the Western Cape Province, Lam et al. (2010) finds Africans take longer to secure a first job even when education level is considered, but find no difference in hazard rate between Whites and Coloureds. Ranchhod and Dinkelman (2007) find that while 25% of the unemployed but actively searching Whites found jobs, only 10% of similar Africans found jobs. Labour force participation varies by gender and it is therefore expected that transition into employment for unemployed men and women would vary. Ranchhod and Dinkelman (2007) find that unemployed men (both actively seeking and discouraged) find jobs at a higher rate than similar women, and that these jobs are more likely to be in the formal sector. The rate of securing jobs in the informal sector is higher for women than for men (Ranchhod and Dinkelman, 2007). Education is argued to be an important factor in employment. Wittenberg (2002) finds that the effect of education in securing employment is only realised with post secondary education. Similarly, Lam et al. (2010) find that getting at least a matriculation diploma substantially increases employment prospects for graduates. The number of studies on unemployment duration for South Africa is growing.

This summary of factors that affect unemployment duration indicates that both demand and supply side factors are important determinants. One can deduce from the summary that firstly, transition out of unemployment varies by exit state and gender. Secondly, the effect of age is con-

flicting by exit states and gender. Ollikainen (2003) and Lentz and Tranæs (2005) find transition into employment decreases with age for men, but Ollikainen (2003) finds it increases with age for women. Gonzalo and Saarela (2000) find transition into economic inactivity decreases with age. Thirdly, education significantly affects unemployment duration, and educated women have a higher exit into employment than men (Böheim and Taylor, 2000; Ollikainen, 2003). Fourthly, both employment benefits and non-labour incomes negatively affect employment probability. Finally, the literature in developing countries although limited, indicates longer duration, and similar evidence exists for South Africa. Unemployment duration has been shown to vary by gender, race, age and education level. The literature is, however, lacking in studies on the influence of competing risk by gender, an area to which this study contributes.

4.3 Analytical framework and estimation strategy

The analytical framework of this study is based on the theory of job search as discussed in Mortensen (1986). The cost of searching for work includes a financial cost and search time costs. Based on social role expectations for each gender, men and women face different time constraints. Unemployed men and women may allocate time differently to their job search but they may face similar financial costs. The time allocated to search has a significant influence on the probability of securing a job. According to the theory of job search, the time it takes to exit unemployment is determined by the probability of receiving a job offer. This probability is based on the general labour market environment, and the individual's probability of accepting the offer, which is dependent on their reservation wage. In a similar manner to Ollikainen (2003), we further assume that the probability that an individual receives a job offer, and accepts the offer, could in part be determined by gender, among other demand and supply-side factors. For instance, an employer might choose the best candidate based on gender, and the reservation wage might differ between men and women.

In the estimation of duration of unemployment, we do not model the separate probabilities of receiving a job offer, and the offer being accepted. We instead adopt the reduced form specification, and consider the conditional probability of leaving unemployment, that is, the hazard function. The hazard function is the product of these two probabilities, that is, the probability that a job offer is received, and the probability that the offer is accepted (see Kiefer (1988) p656 for a discussion). We define the hazard function as:

$$\lambda(t) = \frac{f(t)}{S(t)} \tag{4.1}$$

where $f(t)$ is the probability density function, and $S(t) = 1 - F(t)$ is the survival function which yields the probability of surviving (staying unemployed) until period t. In the estimation of exit from unemployment, we consider a competing risk model because we consider two possible exits, namely employment and economic inactivity (non-participation in the labour force). As a result, the hazard rate of exiting unemployment is given by the sum of the probability of transiting to employment, and the probability of transiting to economic inactivity. We assume the competing

risks are independent, conditional on explanatory variables which, given the assumption of continuous duration, means we estimate two separate models for each exit (Addison and Portugal, 2003). When estimating spells ending with employment, spells ending with non-participation are considered censored, and vice versa. We first plot the the Kaplan-Meier survival function for men and women to compare their unemployment duration.

To investigate the determinant's of exit from unemployment into either employment or economic inactivity by gender, we consider the semi-parametric Cox proportional hazard model as outlined in Kiefer (1988). Suppose that n unemployed individuals faces two exits $j = j_1,\ j_2$ out of unemployment, and an individual's exit status is identified by a dummy variable, $d_i = 1$ for individuals with complete spell, and $d_i = 0$ for individuals with censored spell. To estimate the hazard rates, we maximize the following log likelihood function for each transition state:

$$\sum_j \sum_i \left[d_{ij} * log\lambda_j (t_i, x_i, \beta_i) \right] + \sum_j \sum_i \left[logS_v (t_i, x_i, \beta_i) \right] \tag{4.2}$$

where for individual i, t_i is the unemployment duration, x_i is a vector of covariates, and β_i are the coefficients to be estimated. Therefore, considering we have right censored observations, $d_{ij} = 1$ for the exit status for which the hazard is being calculated, and $d_{ij} = 0$ otherwise. This means that the censored individuals only contribute to the log likelihood a factor S, as indicated by the second term of equation 4.2, while the uncensored individuals contribute to both terms, that is, the factor S and the density of the observed exit status (see Cox and Oakes (1984) for details). Censored observations are therefore included, since excluding them leads to an upwards asymptotic bias of the estimated hazard (Kiefer, 1988).

In estimating individual $i's$ hazard rate to exit to state j, we follow Cox's (1972) proportional hazards model, which takes the following form:

$$\lambda(t, x, \beta) = \lambda_0(t)\phi(x, \beta) \tag{4.3}$$

where $\lambda_0(t)$ is the baseline hazard common to all individuals, β are the coefficients to be estimated, and x_i is a vector of covariates (such as individual characteristics, family background, and whether the individual is a recipient of any of the four social grants, district council unemployment share, and employment share in manufacturing in the district council) that are constant over the duration. We assume that the hazard rate is continuous over time. This overall hazard of exiting unemployment is obtained from the sum of the two transition probabilities:

$$\lambda(t, x, \beta) = \sum_{j=1}^{2} \lambda_j(t, x, \beta) \tag{4.4}$$

We estimate the hazard rates for each exit state using equation 4.2 by maximizing the partial likelihood function, which allows for the estimation of β without specifying the form of the baseline

hazard function $\lambda_0(t)$.

If the hazard is non-constant, it means there is duration dependence. Positive dependence will occur where the estimate of the baseline hazard is $\lambda_0(t) > 0$, and negative dependence will occur where the estimate of the baseline hazard is $\lambda_0(t) < 0$. However, incorrect inference on duration dependence, and on the effect of explanatory variables included, is likely to result from heterogeneity due to omitted variables and/or measurement errors (Kiefer, 1988). Although the presence of heterogeneity could result in a downward bias of duration dependence, and a bias towards zero for effects of covariates (Lancaster, 1990; Kiefer, 1988), we do not control for this. This is because ignoring individual heterogeneity is equivalent to possible misspecifications and distortions that could arise from controlling for it (Böheim and Taylor, 2000). Besides which, there is evidence to suggest that the standard methods of correcting for heterogeneity do not improve the identification of duration (Addison and Portugal, 1998).

4.4 Data

The study uses the National Income Dynamic Study (NIDS), which is a nationally representative household and individual level panel, that commenced in 2008. We focus our sample on individuals aged 15 to 64 years who responded to the three waves of NIDS (balanced panel), and had an identifiable unemployment spell. In defining the sample of unemployed, NIDS follows the ILO definition, which the South African government adopted in 1998 as its official definition. This official definition (also referred to as strict/narrow definition) considers an unemployed individual to be someone who did not have work in the last seven days prior to the interview date, wanted work and was available to work in that week, and actively tried to find work or start a self-employment activity in the preceding four weeks (Statistics South Africa, 1998). There has been a debate on the difference between the official definition and the expanded definition. On the one hand, Posel et al. (2014) argue that disregarding unemployed individuals who are not actively looking for a job, but are likely to be passively looking for work, could be misleading given that majority of the unemployed in NIDS Wave 1 and 2 found jobs through social networks rather than through job search. On the other hand, Ranchhod and Dinkelman (2007) find that unemployed individuals who actively search for employment have a higher chance of finding employment than unemployed and discouraged individuals. They are also more likely to find formal jobs, while the discouraged are more likely to find informal jobs. In NIDS, respondents were asked three questions regarding their employment status, which are considered in identifying the unemployment spell. The first is on unemployment duration before the current job for those employed at the time of the survey. The second is the period since the last job. The third identifies individuals who have never worked, but have been wanting to work. Data collected by these three questions give us both complete and right censored spells. Unemployment spells longer than 49 years, which is the time between the legal age of starting work (15) and retirement at age 64, are dropped from the estimation.

We consider non-labour income by including dummy variables that indicate whether an individual is a recipient of any of the four[1] government social grants. These grants are disability, child support, foster care, and care dependency. To control for family wealth, we include a dummy indicating whether an individual lives in a dwelling that is owned by a family member.

In addition, we control for personal and family characteristics that include education level attained, marital status, relationship with household head, population group, age , home location, and home province. These variables are included as categorical variables. In addition to the above supply side effects, we also control for demand side effects by including the unemployment level of the district council , and also the share of individuals employed in the manufacturing sector in the district council.

The demand- and supply-side factors we control for are not exhaustive. Due to data limitations we are unable to control for factors such as job search intensity[2] and networks[3], among others. These factors, and unobserved factors such as an individual's motivation to find work are likely to introduce bias in our estimates. Since it is impossible a priori to know the direction of the likely bias, we recommend that any inference on the estimates be made with caution.

4.5 Descriptive statistics

In this section, we present the summary statistics of unemployment duration, individual and family characteristics, and some demand-side factors. As discussed in the previous sections , we explore the variation in post-unemployment states by gender because this may provide more interesting insights into labour supply decisions and unemployment outcomes for women relative to men. We first present the distribution of unemployment duration by exit states and by gender followed by the distribution of the individual and family characteristics, and demand side factors of the sample.

The summary statistics of unemployment duration by exit states and by gender are presented in Table 4.1. Although, it would be ideal to weight these statistics for both non-response and attrition across the waves, we are unable to do this since a panel weight for the three waves is currently not available for NIDS. The table shows that the majority of the individuals in our sample (61%) are women. Unemployment spells that ended due to employment are 48%, about 10 percentage points higher than those that ended in inactivity. About 14% of the individuals were still unemployed at the end of the study period in 2012.

[1] There are currently 5 government social grants. We, however, exclude state old age pensions.

[2] We do not control for intensity of search because very few individuals responded to the question of search resulting in a small sample size when included in the analysis.

[3] Although NIDS includes a question on the networks used to search for jobs, it changed the question in Wave 3 which makes it impossible to reconcile Wave 3 responses with responses from the other two waves.

Table 4.1: Distribution of unemployment spells by exit state and gender (%) for the period 2008-2012

	Censored	Employed	inactive	All Exits	Total by gender
		Mean(s.e.)			Frequency (%)
Male	4.295	5.349	4.578	4.892	3884
	(0.114)	(0.090)	(0.235)	(0.068)	(39.34)
Female	5.341	6.747	5.928	6.059	5988
	(0.104)	(0.101)	(0.189)	(0.069)	(60.66)
Mean difference	-1.046***	-1.398***	-1.351***	-1.168***	9872
by gender	(0.161)	(0.139)	(0.337)	(0.101)	(100)
Total by exit(Freq.)	3753	4740	1379	9872	
(%)	(13.97)	(48.02)	(38.02)	(100)	

Notes: The Sample is of working population (age 15-64). Source: Authors calculation based on Wave 1, 2, and 3 balanced of NIDS.

From the table, it is evident that women have a longer average unemployment duration at a mean of six years, compared to that of men (about one year). Interestingly, women take longer to leave the labour market than men by at least one year. This could be an indication that men are easily discouraged in job search, while women are more persistent. The mean of unemployment duration in years by exit states and gender is also presented in Table 4.1. The mean differences[4] by gender for all exit states are statistically significant at 1% level. This suggests that gender is an important factor in determining unemployment duration. We depict the distribution of unemployment duration in Figure 4.1.

[4]We assume the variance for men and women is the same.

Figure 4.1: Distribution of unemployment duration by gender for the period 2008-2012

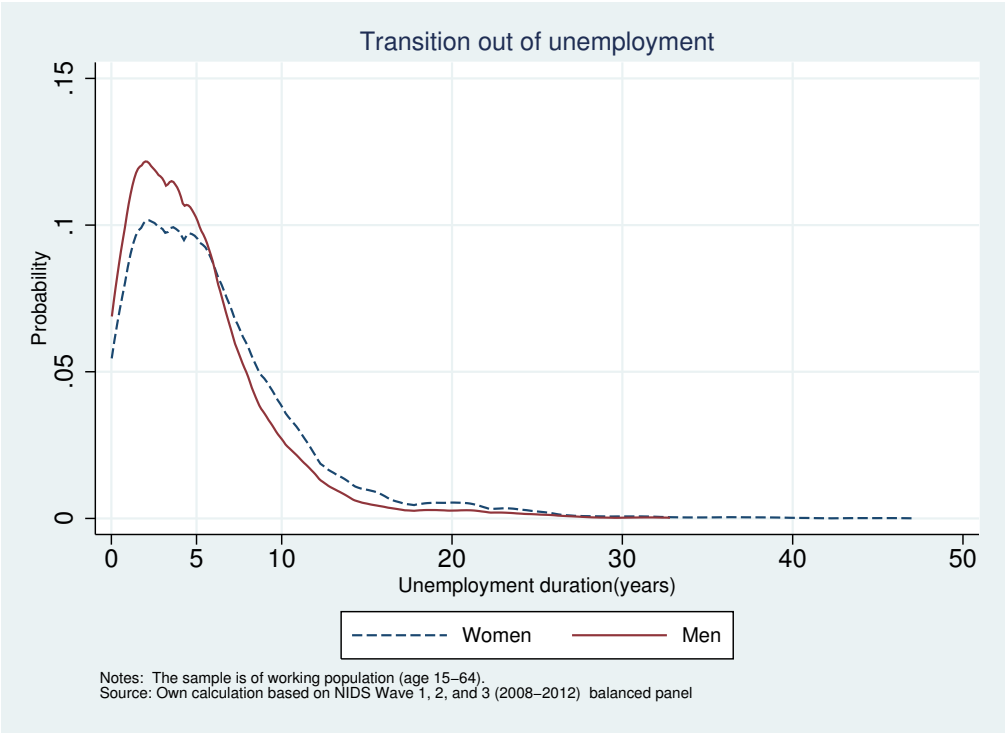


Figure 4.1 presents kernel densities of unemployment duration by gender. The figure indicates that unemployment duration for both men and women is skewed to the left and the two overlap over time. It first increases at an increasing rate, peaks after about 2 years of unemployment then steadily declines just before 5 years of unemployment. As the figure shows, at the beginning of an unemployment spell, men have a higher probability of leaving unemployment than women, but after about 6 years of unemployment the probability of leaving unemployment is higher for women than it is for men.

Next we consider summary statistics of individual characteristics and labour market characteristics that we include in the full analysis by gender. These are presented in Table 4.2. The sample comprises women and men who are primary educated, African, not married, household heads, from Kwa-Zulu Natal province, living in urban areas, and living in a dwelling owned by a household member. In the sample, most women are aged 26-35, while most men are aged 15-25.

The majority of individuals in the sample are not recipients of government social grants, since only less than 2% of the sample are recipients. The exception, however, is the child support grant, where 48% of women are recipients. Both men and women in our sample live in district councils with a share of unemployment at 35%, and share of residents employed in manufacturing at 11%.

Table 4.2: Individual characteristics of the labor market full sample and by gender for the period 2008-2012

Variables	Both	Female	Male	Mean difference by gender
	Mean (Standard error)			
Highest education level attained				
No schooling	0.0582	0.0549	0.0633	0.0084*
	(0.002)	(0.003)	(0.004)	(0.005)
Primary	0.366	0.357	0.379	0.0220**
	(0.005)	(0.006)	(0.008)	(0.010)
Some secondary	0.241	0.253	0.222	-0.0306***
	(0.004)	(0.006)	(0.007)	(0.009)
Grade 12	0.207	0.201	0.217	0.0160*
	(0.004)	(0.005)	(0.007)	(0.008)
Tertiary	0.128	0.134	0.118	-0.0158**
	(0.003)	(0.004)	(0.005)	(0.007)
Population group (Race)				
African	0.843	0.832	0.859	0.0271***
	(0.004)	(0.005)	(0.006)	(0.008)
Coloured	0.140	0.150	0.124	-0.0265***
	(0.004)	(0.005)	(0.005)	(0.007)
Indian/Asian	0.0070	0.0070	0.0080	0.0010
	(0.001)	(0.001)	(0.001)	(0.002)
White	0.0099	0.0105	0.0090	-0.0015
	(0.001)	(0.001)	(0.002)	(0.002)
Age group				
Age 15-25	0.273	0.250	0.308	0.0581***
	(0.005)	(0.006)	(0.007)	(0.009)
Age 26-35	0.299	0.311	0.280	-0.0316***
	(0.004)	(0.006)	(0.007)	(0.009)
Age 36-45	0.232	0.247	0.208	-0.040***
	(0.004)	(0.006)	(0.007)	(0.009)
Age 46-55	0.157	0.156	0.157	0.001
	(0.004)	(0.005)	(0.006)	(0.008)
Age 55-64	0.0397	0.0349	0.0471	0.0122***
	(0.002)	(0.003)	(0.003)	(0.004)
Marital status (married=base)	0.254	0.251	0.259	0.0076
	(0.004)	(0.006)	(0.007)	(0.009)
Household head	0.408	0.371	0.463	0.0920***
	(0.005)	(0.006)	(0.008)	(0.0101)
Household member owns dwelling	0.789	0.803	0.769	-0.034***
	(0.004)	(0.005)	(0.007)	(0.008)
Government social grant				
Disability grant	0.0185	0.0202	0.0160	-0.004
	(0.001)	(0.002)	(0.002)	(0.003)
Child support grant	0.298	0.483	0.0108	-0.473***
	(0.005)	(0.007)	(0.002)	(0.008)
Foster care grant	0.0112	0.0174	0.0018	-0.016***
	(0.001)	(0.002)	(0.001)	(0.002)
Care dependency grant	0.004	0.005	0.00154	-0.004***
	(0.001)	(0.001)	(0.001)	(0.001)

Notes: The sample is of working population (age 15-64) in NIDS balanced panel. Source: Authors calculation based on the three waves of NIDS.

Continued next page.....

Table 3 continued...

Variables	Both	Female	Male	Mean difference by gender
	Mean (Standard error)			
Home location (Urban=base)	0.526	0.525	0.527	0.0021
	(0.005)	(0.007)	(0.008)	(0.010)
Home province				
Western Cape	0.102	0.114	0.0850	-0.0286***
	(0.003)	(0.004)	(0.005)	(0.006)
Eastern Cape	0.108	0.106	0.111	0.0048
	(0.003)	(0.004)	(0.005)	(0.006)
Northern Cape	0.0739	0.0750	0.0723	-0.0026
	(0.003)	(0.003)	(0.004)	(0.0054)
Free State	0.0829	0.0845	0.0803	-0.0042
	(0.003)	(0.004)	(0.004)	(0.006)
KwaZulu Natal	0.243	0.245	0.240	-0.0044
	(0.004)	(0.006)	(0.007)	(0.009)
North West	0.107	0.0989	0.120	0.0216***
	(0.003)	(0.004)	(0.005)	(0.006)
Gauteng	0.118	0.106	0.135	0.0295***
	(0.003)	(0.004)	(0.006)	(0.007)
Mpumalanga	0.0794	0.0808	0.0772	-0.0036
	(0.003)	(0.004)	(0.004)	(0.006)
Limpopo	0.0854	0.0903	0.0778	-0.0126**
	(0.003)	(0.004)	(0.004)	(0.006)
Share of unemployed in district council (dc)	0.349	0.349	0.349	-0.0001
	(0.001)	(0.002)	(0.002)	(0.002)
Share employed in manufacturing in dc	0.107	0.106	0.107	0.001
	(0.001)	(0.001)	(0.002)	(0.002)
Number of observations	9872	5988	3884	9872

Notes: The sample is of working population (age 15-64) in NIDS balanced panel. Source: Authors calculation based on the three waves of NIDS.

The last column of Table 4.2 gives the mean differences of these characteristics by gender. The majority of the characteristics are statistically significant. The only variables with insignificant mean differences are marital status, home location, and the labour market demand-side factors. The latter include share of unemployed and share employed in manufacturing in the district council of residence.

4.6 Empirical results

In this section, using both non-parametric and semi-parametric[5] approaches, we estimate the probability of exiting unemployment into either employment or economic inactivity. In the non-parametric approach, using the Kaplan-Meier estimator, we consider a selected number of characteristics we consider have important effect's on transition into either employment or economic inactivity. In addition, these factors have a significant mean difference by gender, as shown in Table 4.2. They include gender, education level , race, age group, and being a recipient of a government social grant. In the semi-parametric approach, we use the Cox proportional hazard model in which we control for individual and family characteristics, and demand-side factors. The findings from these

[5]We also estimate the Weibull hazard (a parametric model), and find similar results to the results from the Cox proportional model. We do not present these results here, but they could be provided on request.

analyses are presented in the following sub-sections.

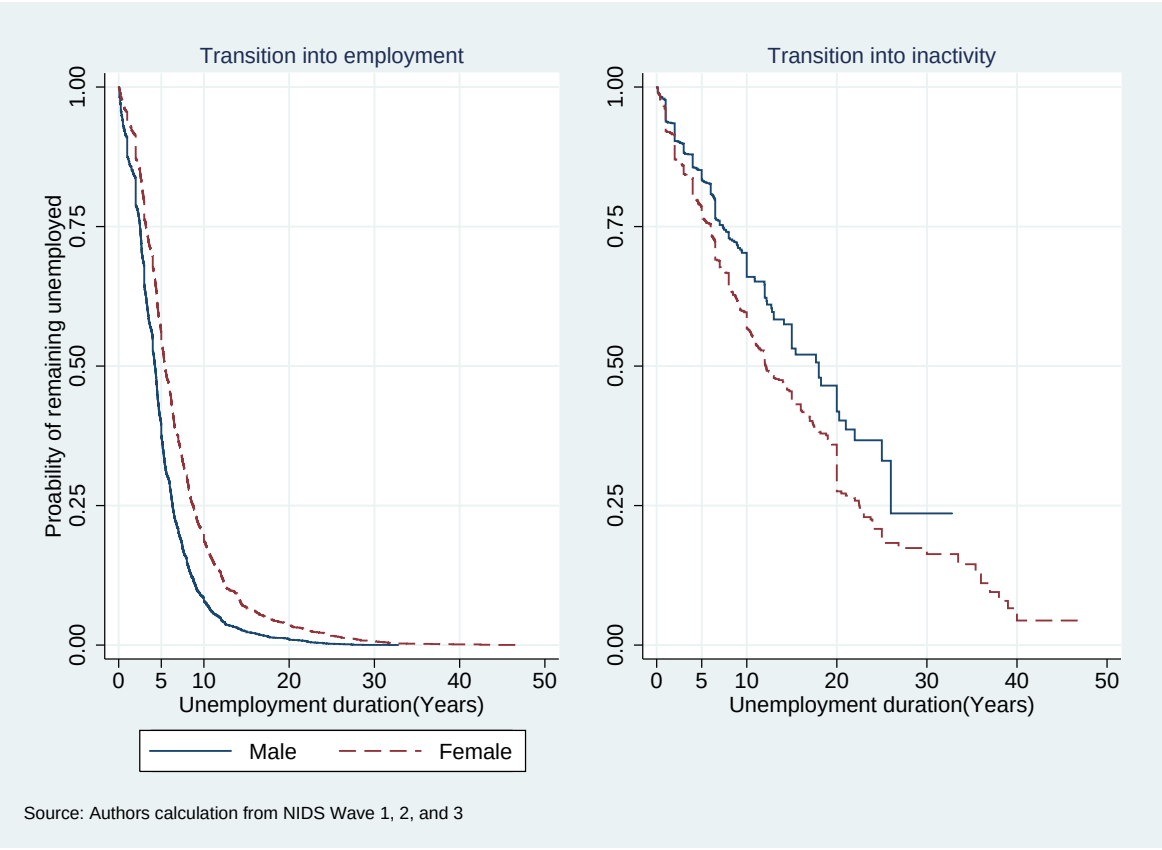
4.6.1 Non-parametric estimates

Using the Kaplan-Meier estimator, we first start by estimating unemployment spell by gender. This, as noted earlier, is the factor we use to study variation in post-unemployment states. We then present the Kaplan-Meier estimates of exit states by education level , race, age group, and receipt of government social grants and compare them with the estimates by gender.

The survival function of unemployment from the Kaplan-Meier estimator by gender[6] is presented in Figure 4.2. This shows that for men the probability of exiting into employment after 5 years is 65%. For women, the probability of exiting into employment after 5 years is 45%. For men, The probability of exiting into economic inactivity after 5 years for men is 15%, while for women it is 22%. Three conclusions can be drawn from these results. First, the transition out of unemployment differs by exit states for men and women. This is in agreement with other empirical evidence (Gonzalo and Saarela, 2000; Ollikainen, 2003). Second, men have a higher probability of exiting into employment, while women have a higher probability of exiting into economic inactivity. Third, the rate of transition into employment is steeper than the rate of transition into economic inactivity for both men and women.

[6]The log-rank tests for both exits reject the null hypothesis of equality of survival functions at the 1% level.

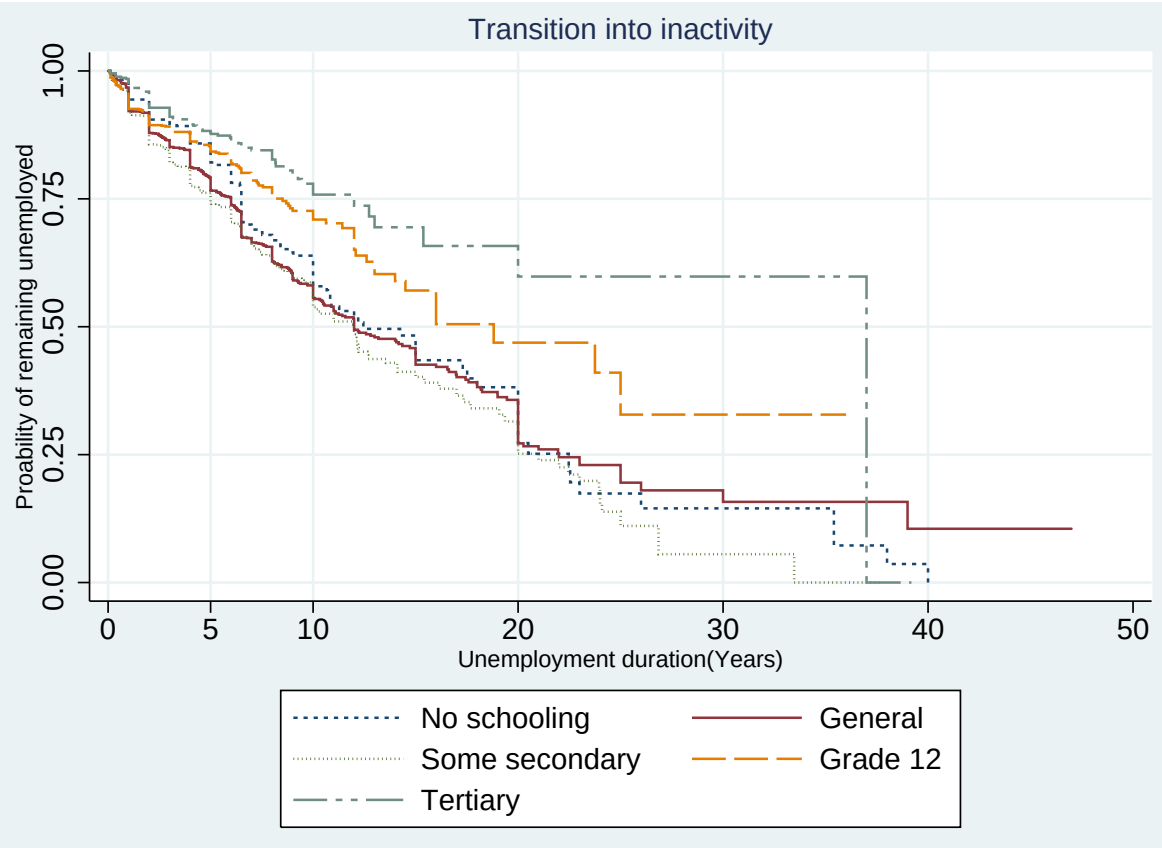
Figure 4.2: Kaplan-Meier survival estimates, by exit and gender for the period 2008-2012



Source: Authors calculation from NIDS Wave 1, 2, and 3

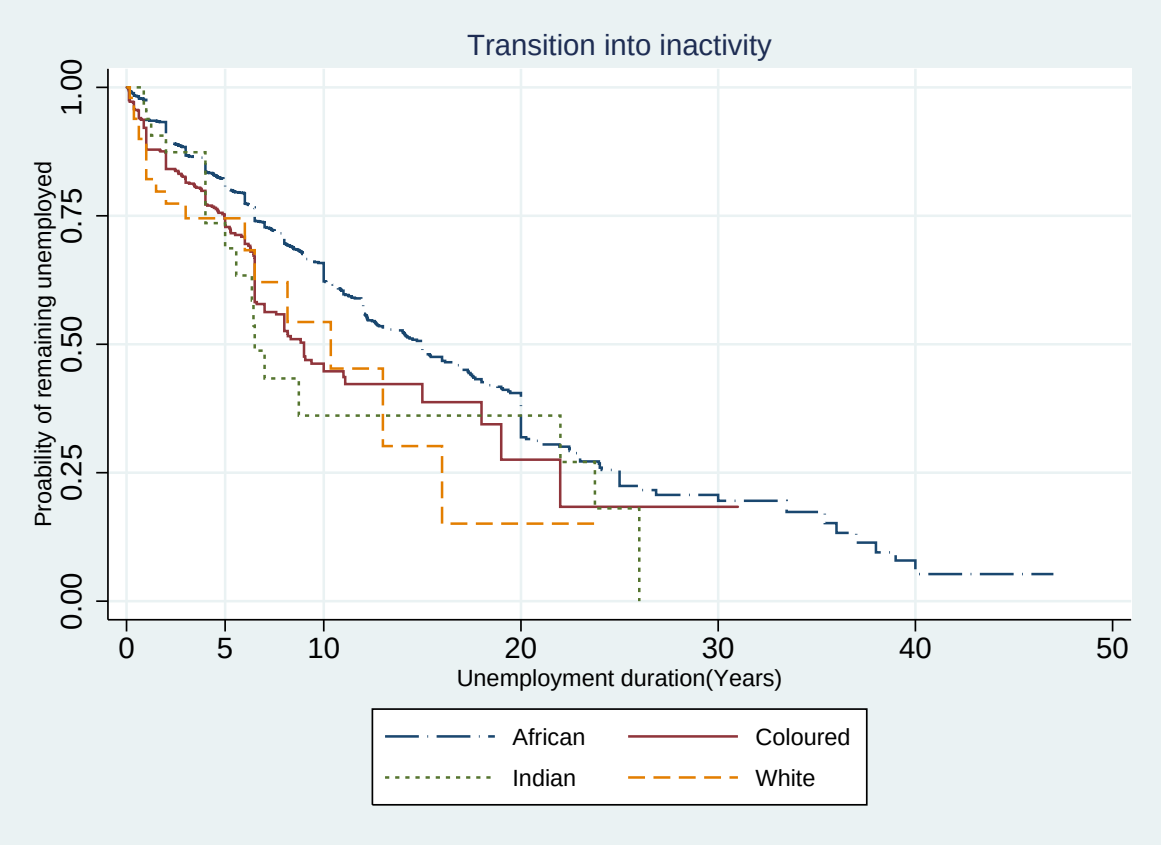
The effect of education on the probability of exiting unemployment is presented in Figure 4.3. For those with a tertiary education, the probability of exiting into employment after 5 years is 70%. For those with a grade 12, some secondary or general education, or who have no schooling, the probability of exiting into employment after 5 years is 55%, 50%, and 42%, respectively. The figure also shows that, using these education categories, the probability of exiting into economic inactivity after 5 years is 12%, 17%, 31%, 24%, and 20%, respectively. Four conclusions can be drawn from these results. First, the transition out of unemployment differs by exit states for different levels of education. Second, individuals with tertiary education have the highest transition into employment and the lowest transition into economic inactivity. Third, the rate of transition into employment is steeper than the rate of transition into economic inactivity for all education levels. Fourth, the difference in transition rate into employment and into economic inactivity is not clear for individuals with general education and for those with secondary education. Overall, the transition rate out of unemployment by education is less clear when compared with the transition rate by gender, at least for some levels. This is in agreement with existing evidence that shows the effect of education by gender is largely mixed (Böheim and Taylor, 2000; Ollikainen, 2003).

Figure 4.3: Kaplan-Meier survival estimates, by exit and education level for the period 2008-2012



The effect of race on exiting unemployment is presented in Figure 4.4. For White or Coloured South Africans, the the probability of exiting into employment after 5 years is 70%; conversely, if you are Indian, or African, the probability of exiting into employment after 5 years 65% and 50%, respectively. For transition into economic inactivity, for Indian South Africans, the probability of exiting into economic inactivity after 5 years of unemployment is 35%. For Coloured, White or African South Africans, the probability of exiting into economic inactivity after 5 years of unemployment is 28%, 26%, and 15%, respectively. We deduce from these findings that the transition out of unemployment differs by exit states for the different races. The rate of transition into employment is also steeper for all races relative to the rate of transition into economic inactivity. The difference in transition rate into employment and into economic inactivity is not clear for the different races. At most points such as after 5 years, Whites have the highest transition rate into employment and the lowest transition rate into economic inactivity. Africans have the lowest transition into employment and into economic inactivity an indication that Africans remain unemployed for longer, and supports the evidence by Ranchhod and Dinkelman (2007) and Lam et al. (2010). Overall, the transition rate by race when compared with the transition by gender is less clear for all races.

Figure 4.4: Kaplan-Meier survival estimates, by exit and race for the period 2008-2012



The effect of age on exiting unemployment is presented in Figure 4.5. The difference in the probability of exiting into employment is hardly differentiable by age group, however after 13 years the probability of exiting into employment is lower for those aged 15-25. Similarly, the rate of transition into economic inactivity by age group is hardly differentiable except for those aged 15-25 who have a faster transition. We conclude that the transition out of unemployment differs for the different age groups, which is in agreement with findings by Ahn and Ugidos-Olazabal (1995). The rate of transition into employment is also steeper for all age groups relative to the rate of transition into economic inactivity. Lentz and Tranæs (2005) show unemployment decreases with age. However, the difference in transition rate into employment and into economic inactivity is not clear for different age groups. Overall, the transition rate by age group is not as obvious as the transition rate by gender for all age groups.

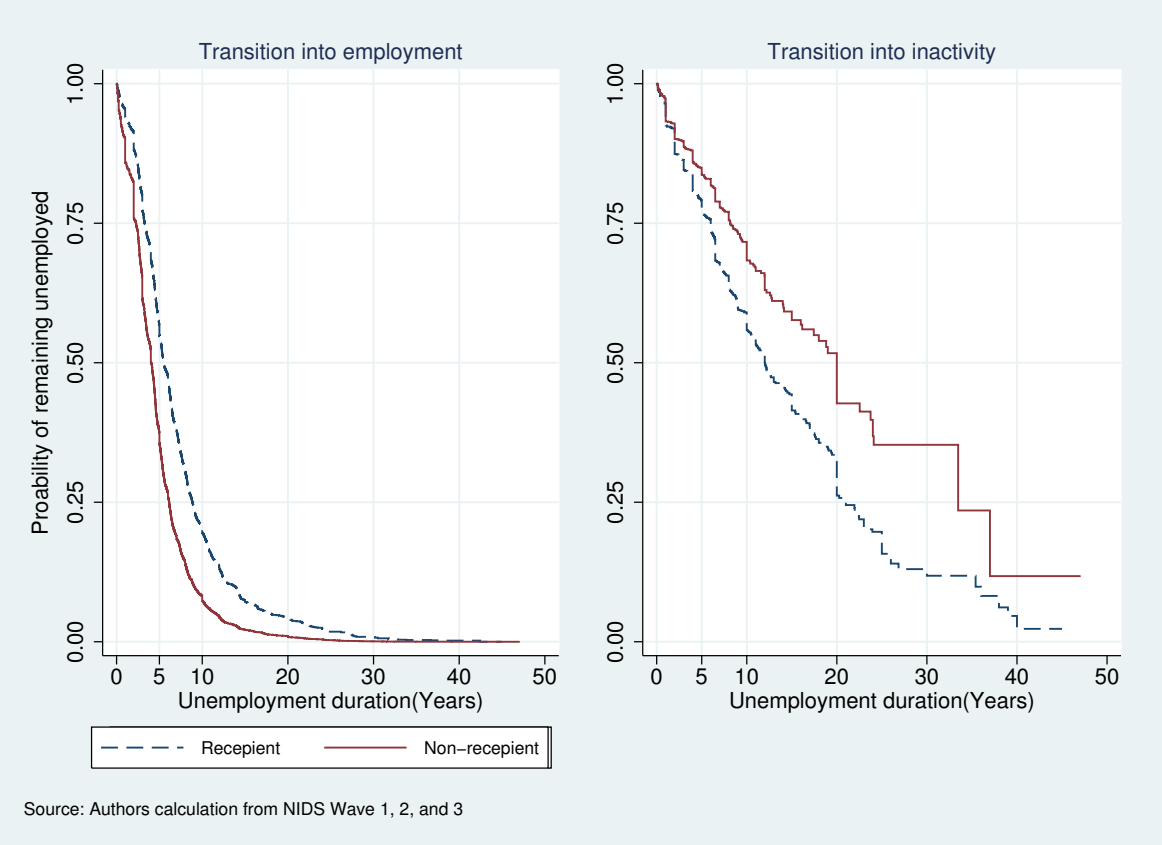
Figure 4.5: Kaplan-Meier survival estimates, by exit and age for the period 2008-2012

Source: Authors calculation from NIDS Wave 1, 2, and 3

In the full analysis we consider four social grants. These are disability, child support, foster care, and care dependency grants. However, for ease of analysis we consolidate them into an indicator variable that identifies recipients of any one of the four grants. The effect of government social grants on exiting unemployment is presented in Figure 4.6. This shows that for grant recipients, the probability of exiting into employment after 5 years of unemployment is 45%. For non-recipients , the probability of exiting into economic inactivity after 5 years of unemployment is 65%. The figure also shows that, if you are a recipient of a social grant, the probability of exiting into economic inactivity after 5 years is 22%.If you are a woman, the probability of exiting into economic inactivity after 5 years is 15%. Three conclusions can be drawn from this. First, the transition out of unemployment differs by exit states for recipients and non-recipients of government social grants. Second, non-recipients have a higher probability of exiting into employment, while recipients have a higher probability of exiting into inactivity. Third, the rate of transition into employment is faster for both recipients and non-recipients relative to the rate of transition into economic inactivity. This is in agreement with expectation that non-labour income reduces the probability of full-time employment. It is, however, contrary to findings that receipt of government grants has no effect on the probability of being economically inactive (Böheim and Taylor, 2000). For South Africa, Williams (2007) and Eyal and Woolard (2011) find child support grants increased female labour force participation. Overall, the effect of social grants on transition out of

unemployment is similar to the effect of gender. Social grants would therefore be an interesting factor to analyse variation in exit states. However, the sample is mostly non-recipients. This makes gender a preferred variable because of the high response from both men and women, among other reasons.

Figure 4.6: Kaplan-Meier survival estimates, by exit and social grant for the period 2008-2012



The above analyses suggest that transition out of unemployment varies by exit states and the rate of transition into employment is steeper than the rate of transition into economic inactivity. Each characteristic affects transition differently, but the effect of gender and social grants are most distinct. We choose to focus on the effect by gender rather than social grants, for to reasons noted earlier.

4.6.2 Cox proportional hazard estimates

In this section, we present the estimates of the probability of leaving unemployment, based on the Cox proportional hazard model, by exit state and gender. In the model we include several controls. These are education, race, age group, marital status, household head status, location, home province, social grants, share of unemployed in the district council of residence, and share employed in manufacturing in the district council of residence. The estimates from these controls are discussed next.

The estimated hazard rates from the Cox proportional hazard models by exit state for both women and men are presented in Table 4.3. The estimates from the full sample show that the hazard rate into employment is decreasing for women, while the hazard rate into economic inactivity is increasing, holding all other factors constant. Education has been empirically shown to significantly influence the probability of employment for both men and women. Our estimates indicate that the effect of education on exit from unemployment to either employment or inactivity differs by gender. Having a primary education relative to having no schooling significantly increases the exit rate into employment for women, but has no effect on men, and no effect on exit into inactivity for either gender. The effect of post primary education on exit to employment for both women and men, however, increases with the level attained. Its effect on exit to economic inactivity decreases with the level attained.

A grade 12 or tertiary education relative to no schooling increases the hazard rate into employment for both women and men, but the rate is higher for women. These results have two likely implications. One is that it is an indicator of higher returns in education, particularly in terms of the probability of employment. Second, it is perhaps an indication of employment equity which places a premium on hiring women. In an economy which is skills-biased in labour demand preferences it can be an easier and quicker transition into employment for educated women relative to educated men. On the other hand, a grade 12 or tertiary education decreases the hazard rate into inactivity for both women and men, but the hazard rate is higher for women than for men. It is interesting to note that the hazard rate of transition into inactivity does not differ substantially between women and men with tertiary education. The higher rate of exit into employment for women can indicate they are more favored by employers .

Race has been shown to be an important factor in the determination of employment in South Africa. Yet, we find race to be a significant factor only for women. Being a Coloured woman relative to being an African woman, increases the exit rate into employment. Contrary to expectation, we find being a Coloured or a White woman increases the exit rate into inactivity relative to being African. These results are likely to indicate that the importance of race in the South African labor market is declining particularly for men.

Table 4.3: Cox hazard model estimates

Variables	Employed			Inactivity		
	Both exits	Female	Male	Both exits	Female	Male
Female	-0.3147***			0.3839***		
	(0.042)			(0.080)		
Education level(No schooling=base)						
Primary	0.1679**	0.2412**	0.0428	0.0212	-0.0102	0.1142
	(0.083)	(0.117)	(0.104)	(0.110)	(0.131)	(0.226)
Some secondary	0.2490***	0.2683**	0.1966*	-0.0446	-0.0959	0.1398
	(0.092)	(0.126)	(0.118)	(0.119)	(0.135)	(0.243)
Grade 12	0.3930***	0.4722***	0.2593**	-0.6395***	-0.7144***	-0.4590*
	(0.092)	(0.129)	(0.111)	(0.131)	(0.148)	(0.268)
Tertiary	0.6254***	0.7435***	0.4281***	-0.7603***	-0.8116***	-0.6842**
	(0.089)	(0.127)	(0.114)	(0.168)	(0.200)	(0.327)
Population group(African=base)						
Coloured	0.1846**	0.2274**	0.1350	0.3750***	0.3952**	0.2775
	(0.079)	(0.099)	(0.118)	(0.142)	(0.169)	(0.289)
Indian/Asian	0.1682	0.0623	0.3729*	0.4275	0.7224*	-0.1822
	(0.161)	(0.198)	(0.214)	(0.337)	(0.424)	(0.371)
White	-0.0686	-0.0007	-0.0565	1.0239***	1.3469***	0.5090
	(0.165)	(0.171)	(0.326)	(0.309)	(0.333)	(0.636)
Age group (age 15-25=base)						
Age 26-35	0.1025**	0.2098***	0.0181	-0.8013***	-0.7999***	-0.8080***
	(0.052)	(0.076)	(0.071)	(0.076)	(0.085)	(0.160)
Age 36-45	0.1158**	0.2240***	-0.0044	-1.0648***	-1.0840***	-1.0763***
	(0.058)	(0.084)	(0.079)	(0.096)	(0.107)	(0.213)
Age 46-55	0.1346**	0.2965***	-0.0914	-0.8712***	-0.9336***	-0.6467***
	(0.065)	(0.093)	(0.091)	(0.111)	(0.119)	(0.222)
Age 55-64	0.1729*	0.3253**	-0.0402	-0.7059***	-0.6788***	-0.6686*
	(0.088)	(0.128)	(0.123)	(0.164)	(0.198)	(0.346)
Marital status (married=base)	0.0406	-0.0812	0.2388***	-0.0630	-0.0344	-0.1551
	(0.043)	(0.058)	(0.061)	(0.079)	(0.091)	(0.183)
Household head (head=base)	0.3118***	0.2726***	0.2827***	-0.3524***	-0.2772***	-0.5139***
	(0.034)	(0.049)	(0.052)	(0.073)	(0.083)	(0.147)
Household member own dwelling	-0.2545***	-0.2112***	-0.3171***	0.1502	0.0771	0.3677**
	(0.043)	(0.057)	(0.062)	(0.099)	(0.105)	(0.182)
Government social grant						
Disability grant	-0.5680***	-0.2995*	-1.0105***	0.5287***	0.5468***	0.7851**
	(0.138)	(0.173)	(0.223)	(0.164)	(0.177)	(0.365)
Child support grant	-0.1841***	-0.1748***	-0.2797	0.0174	0.0255	0.7389*
	(0.050)	(0.052)	(0.284)	(0.073)	(0.074)	(0.391)
Foster care grant	-0.1249	-0.1563	-0.0754	-0.5669*	-0.6030**	0.4279
	(0.131)	(0.138)	(0.495)	(0.290)	(0.292)	(1.102)
Care dependency grant	-0.3266	-0.2673	-0.6734	0.3579	0.5475	-44.2040
	(0.233)	(0.243)	(0.580)	(0.362)	(0.349)	(.)

Notes: *(1)* Each column is an independent regression. *(2)* In all regressions cluster corrected standard errors are in parenthesis; ***, **, and * indicate significance at 1%, 5%, and 10% level respectively. *(3)* Sample of working population (age 15-64). Source: Authors calculation based on Wave 1, 2, and 3 balanced of NIDS.

Continued next page....

Table 4 continued...

	Employment			Inactivity		
	Both	Female	Male	Both	Female	Male
Home location (Urban=base)	0.2259***	0.3168***	0.1046	-0.0248	-0.1402	0.2582
	(0.051)	(0.066)	(0.067)	(0.104)	(0.108)	(0.176)
Home province (Western Cape=base)						
Eastern Cape	-0.1152	-0.1096	-0.1244	-0.4488**	-0.4102*	-0.3844
	(0.113)	(0.129)	(0.165)	(0.210)	(0.232)	(0.359)
Northern Cape	-0.1639	-0.3022**	-0.0152	-0.2466	-0.1263	-0.4971
	(0.108)	(0.131)	(0.146)	(0.194)	(0.199)	(0.448)
Free State	-0.1649	-0.2313*	-0.1087	-0.4493**	-0.3897	-0.5165
	(0.111)	(0.131)	(0.163)	(0.219)	(0.245)	(0.365)
KwaZulu Natal	-0.0185	0.0507	-0.1393	0.0825	-0.0628	0.5443
	(0.104)	(0.121)	(0.157)	(0.188)	(0.211)	(0.341)
North West	-0.1355	-0.2898*	0.0135	-0.1929	-0.2305	0.0267
	(0.127)	(0.156)	(0.168)	(0.210)	(0.226)	(0.407)
Gauteng	-0.1492	-0.1550	-0.1539	-0.5533***	-0.4386*	-0.6302
	(0.095)	(0.123)	(0.141)	(0.203)	(0.228)	(0.390)
Mpumalanga	0.1309	0.1077	0.1411	-0.4851**	-0.4807**	-0.5103
	(0.104)	(0.129)	(0.152)	(0.230)	(0.243)	(0.454)
Limpopo	-0.2564**	-0.2512*	-0.3240*	-0.8128***	-0.8519***	-0.5979
	(0.120)	(0.147)	(0.192)	(0.256)	(0.283)	(0.413)
Share of unemployed in district council (dc)	-1.6260***	-1.3421***	-1.9613***	1.3208***	1.3301***	1.0915
	(0.234)	(0.297)	(0.290)	(0.360)	(0.366)	(0.696)
Share employed in manufacturing in dc	-1.5674***	-2.0235***	-0.8905**	1.3812***	-1.4293***	-1.3009*
	(0.290)	(0.344)	(0.395)	(0.424)	(0.433)	(0.774)
Number of observations	9872	5988	3884	9872	5988	3884
Ln L	-34620.2	-17921.0	-13465.1	-10074.1	-6824.5	-2422.5
Ln L(0)	-35261.4	-18248.1	-13712.4	-10393.8	-7025.7	-2547.8

Notes: *(1)* Each column is an independent regression. *(2)* In all regressions cluster corrected standard errors are in parenthesis; ***, **, and * indicate significance at 1%, 5%, and 10% level respectively. *(3)* Sample of working population (age 15-64). Source: Authors calculation based on Wave 1, 2, and 3 balanced of NIDS.

The results show that older women have a higher rate of transition into employment. This is a likely indication that employers have an increased preference for older women. However, age has no effect on exit into employment for men, which is unexpected. The effect of age on transition into economic inactivity is negative for both men and women, which at first increases then decreases with age.

Given the interesting and somewhat unexpected effect of higher education, race and age by gender, on employment, we take a further look at the transition rate out of unemployment for a prototype male and female. This individual is African, aged 26-35, with a tertiary education. We chose Africans and the age group 26-35 because they are the most common category in the sample, while we chose tertiary education following the unexpected results above. We calculate the predicted survival probabilities for African men and women aged 26-35 with a tertiary education from the estimated coefficients given in Table 4.3 and the baseline hazard estimates, which we do not report here. We present the survival predicted probability for this individual in Figure 4.7.

Figure 4.7: Predicted probability of Africans aged 26-35 with a tertiary education

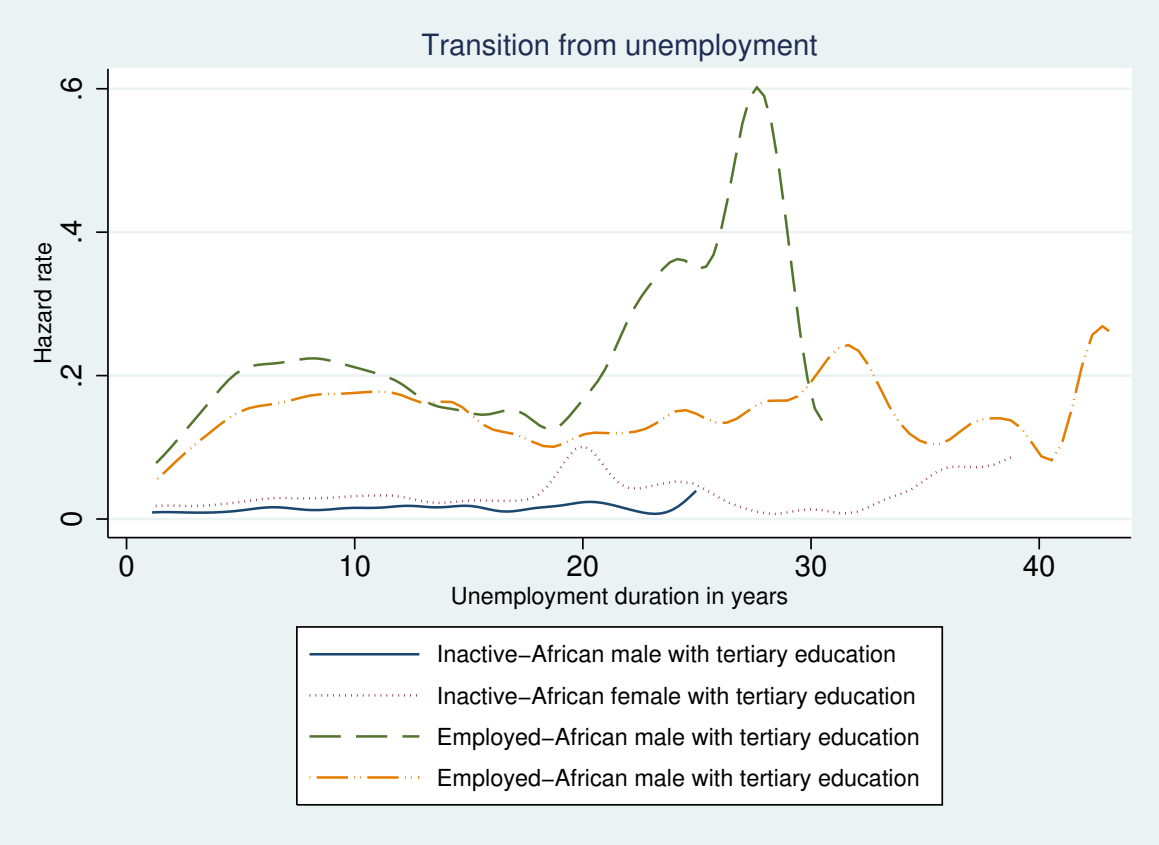


Figure 4.7 shows that African men aged 26-35, with a tertiary education have a higher probability of exiting into employment than similar women. At 5 years of unemployment, the difference in employment probability between men and women is about 4 percentage points. This difference reduces to zero after 12 years of unemployment, but increases after 20 years, and reaches 52 percentage points after 28 years. This suggests that when selecting from a pool of young Africans with tertiary education employers prefer men. However, this bias against women with higher levels of education is less obvious when the average characteristics are considered. The transition into economic inactivity for African women aged 26-35 with a tertiary education is consistently higher than that for similar men. For both men and women, the transition rate into inactivity is almost zero, and constant as the unemployment spell increases, except for women at 20 years of unemployment where it suddenly peaks. Overall, the transition rate into economic inactivity for this typical individual is very small and consistently lower than the transition rate into employment. We can deduce from this that African men and women with tertiary education aged 26-35 are less likely to become economically inactive, and their likelihood of employment remains high even as their unemployment spell increases. This suggests that the demand for educated African workers aged 26-35 is high, and employers are not deterred by long unemployment spells. We further interrogate the effect of age , and consider its effect at two different spells of unemployment, 1 year and 2 years. This is presented in figure 4.8.

Figure 4.8: Transition probability from unemployment by age and gender for the period 2008-2012

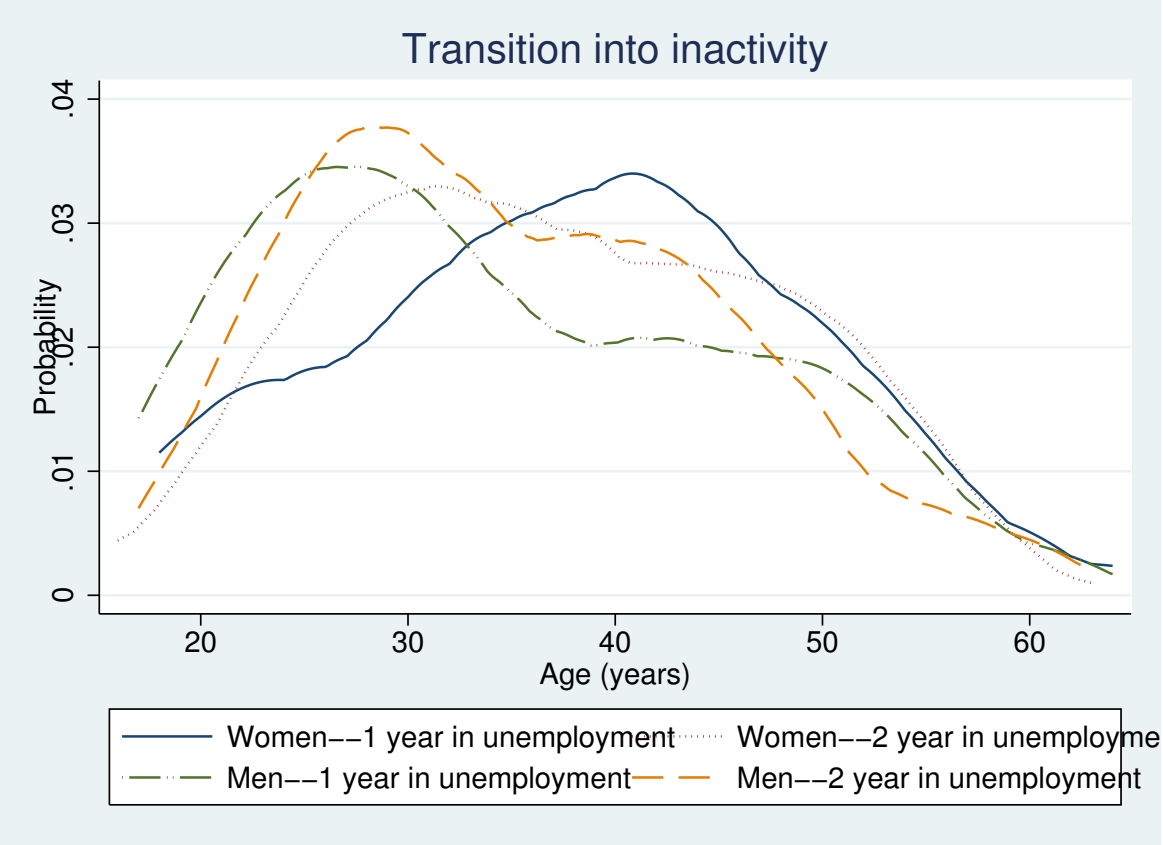


Figure 4.8 indicates the effect of age on unemployment becomes less important with increased age. The effect of age on transition into employment exhibits a bimodal distribution for both men and women, peaking at around 20 and at around 50. Workers aged 20-25 have the highest exit into employment, but men have a higher exit than women for both spells. The high exit rate for these young workers could be because these individuals are likely to receive more job offers. Interestingly, women have a higher exit rate into employment after 2 years of unemployment than after 1 year of unemployment, until about age 40. After age 30, employment probability seems to be equal for men and women, which could be associated with women having passed their most reproductive years. The effect of age on transition into economic inactivity shows that the exit rate for both men and women first increases steadily then declines. Men below the age of 30 are more likely to exit into economic inactivity than women of the same age, and men unemployed for 1 year are more likely to exit into inactivity than those unemployed for 2 years. This suggests that younger men are easily discouraged in job search after a short spell of unemployment. The exit rate for women is slower, and particularly so after 1 year of unemployment. Older women unemployed for 1 year take longer to exit into economic inactivity, which is a likely indication that women are persistence in job search after their reproductive years. For both exit states, younger men seem to have higher exit rates. Differences in transition probabilities for men and women may be a proxy for demand. This includes demand for educated African workers, and the pressure of employment

equity to hire people from formerly disadvantaged groups. The demand for highly skilled workers is also likely to push young men to become economically inactive by re-entering the education system. However, these differences by gender seem to disappear with increase in age. The figure also shows the exit rate into employment is higher than the exit rate into economic inactivity at all ages.

The results show that for a man, being a married increases the exit rate into employment. However, marital status has no effect on transition to employment for women, and on transition to economic inactivity for both men and women. These results suggest that prospective employers place a higher value on the reliability of married men, among other characteristics that employers may be looking for in an employee. The lack of effect on women is unexpected since being a married woman could be a proxy for a second income earner, which in turn may reflect higher reservation wages or a lower intensity of search[7] for married women. The lack of effect of marital status on the hazard rate into inactivity is equally unexpected considering that, for instance, being married is likely to come with child care responsibility which could lead to one spouse, typically the woman, being economically inactive. The relationship with the household is an important factor. Being a household head increases the exit rate into employment and decreases the exit rate into inactivity, for both men and women. This may be explained by the expectation that a household head has to take the responsibility of supporting dependent family members, which forces household heads to put more effort into securing employment, , and delaying exit into inactivity.

The financial status of a family, or having a source of income, whether labour income or non-labour income, has been shown to influence job search. Using living in a dwelling that is owned by a household member as an indicator of family wealth, we find, as expected, that the transition into employment for both men and women is reduced. Lentz and Tranæs (2005) find similar results. This has been associated with reduced mobility, since individuals living in their own houses are less likely to move further from home to find work, and is in line with both theory and empirical evidence (Böheim and Taylor, 2000; and in Oswald (1996 and 1998) as cited in Böheim and Taylor 2000). We infer this to mean that for both men and women, family wealth acts as a form of insurance which reduces their search intensity. Unexpectedly, the results indicate a higher reduction in transition into employment for men than for women. Living in a dwelling owned by a household member, however, significantly increases the transition rate to inactivity for men, but has no effect for women. This result is also unexpected since women, rather than men, are more likely to stay at home in wealthier households. The factors driving this results are not clear and warrant further research.

The effect of social grants show that being a recipient of a disability grant, on the one hand decreases the transition rate to employment for both men and women, albeit, the reduction in the rate for women is less than half that of men. On the other hand, disability grants decrease the transition rate to economic inactivity for both women and men at a similar rate. Interestingly, being

[7] See footnote 3 for a note on our reason for not controlling for search intensity.

a recipient of a child support grant decreases the transition rate to employment for women, but decreases the transition rate to inactivity for men. Being a recipient of a foster care grant decreases transition to inactivity for women, but has no effect for men, or on transition into employment. Being a recipient of a care dependency grant increases the transition rate to inactivity for women, but reduces the transition into economic inactivity for men. Child care dependency grants have no effect on employment probability. However, grant receipt increases the probability of transition into economic inactivity for women and reduces the transition into economic inactivity for men. These results on both foster care and child dependency are in line with expectation, considering women are often the care givers in a household. These results indicate that non-labour income, in this case government social grants, are important factors in determining unemployment duration, but they have varying effects by both exit and gender. Both men and women who have non-labour income take longer to transition to employment. This indicates that non-labour income increases their reservation wage, making these individuals more selective in accepting job offers. These results related to social grants are in line with the findings of Böheim and Taylor (2000), who find non-labour income reduces transition into employment, but has no effect on transition into inactivity. However, for South Africa, existing evidence, which mainly investigates government old age pensions, is mixed. Bertrand et al. (2003) find having a pensioner in a household, particularly a woman pensioner, decreases household labour supply of prime-aged adults. However, Klasen and Woolard (2009) do not find a direct disincentive to job search from the pension and remittance income, but Ardington et al. (2009) find having a pensioner in the home increased the probability of prime-aged adults working.

Living in urban areas relative to living in rural areas increases the exit rate to employment for women, but has no significant effect on the employment of men, and exit to inactivity of both women and men. Living in the Northern Cape, Free State, North West, or Limpopo provinces relative to living in Western Cape province decreases the transition rate to employment for woman, but has no effect on transition of men. Living in the Eastern Cape, Gauteng, Mpumalanda, or Limpopo provinces relative to living in Western Cape province decreases the transition rate to economic inactivity for woman, but has no effect on transition of men.

The level of unemployment in the area of residence may influence the probability of securing employment, and the decision to exit the labour market. We find that high levels of unemployment in the district council of residence decreases the transition to employment for both men and women, although the effect is more pronounced for men . This implies that individuals living in district councils with higher unemployment have a longer unemployment duration. Tansel and Tasci (2004) find a similar effect in Turkey. For transition rate to inactivity, we find district council unemployment rates reflect increases in the rate of this transition for both women and men, and at a similar rate. The effect of an increase in the share of individuals employed in manufacturing, shows that it decreases the transition to employment and to inactivity for both women and men, but the effect is larger for women's transition to employment. This negative effect on transition

to inactivity makes sense since it implies availability of manufacturing jobs reduces the likelihood of leaving the labour market, for both women and men. However, the implication of the negative effect on transition to employment is not obvious, and would warrant further investigation.

Overall, the factors that are driving the difference in exiting unemployment for both women and men include education, relationship with household head, being a recipient of a government social grant, unemployment share in residential district council, and the share of those employed in manufacturing in a district council.

4.7 Summary and conclusion

We use the three waves of the National Income Dynamics Study (NIDS) to study the time it takes men and women to transit from unemployment into either employment or economic inactivity. From a sample of 15 to 64 year olds in all three waves, we estimate a non-parametric model and a semi-parametric model. The results from the Kaplan-Meier estimator and those from the Cox proportional hazard model on the full sample show that while men have a higher probability of exiting into employment, women have a higher probability of exiting into economic inactivity. The finding by exit states and gender from the Cox proportion hazard model show transition rate into employment increases as education level increases for both men and women, and transition rate into economic inactivity decreases as education level increases. We find higher education increases the hazard rate into employment for both women and men, but the rate is higher for women than for men. This could be an indication of higher returns in education, particularly in terms of the probability of employment, or an indication of employment equity which places a premium on hiring women. We also find that at continued higher education levels the hazard rate into inactivity decreases for both women and men, but the hazard rate is higher for women n.

Our findings also show race to be a significant factor, but only for women. The effect of age on unemployment becomes less important as the individual becomes older. The results show a that an African man aged 26-35 with a tertiary education has a higher probability of exiting into employment than a similar woman. . However, this bias against women is less obvious at higher levels of education. For both men and women with these characteristics, the transition rate into inactivity is almost zero, and constant as the unemployment spell increases.

The results of age on transition rate after 1 and 2 years of unemployment show women have a higher exit rate into employment after 2 years of unemployment than after 1 year of unemployment until about age 40. On the other hand, men below the age of 30 are more likely to exit into economic inactivity than women of the same age, and men unemployed for 1 year are more likely to exit into inactivity than those unemployed for 2 years. Workers aged 20-25 have the highest exit into employment, but men have a higher exit than women for both spells. The high exit rate for these young workers could be because these individuals are more skilled, and are therefore likely to receive more job offers.

Marital status is largely not important, but household headship is important. Living in urban areas increases the transition rate to employment for women. Being a recipient of a government social grant decreases the transition rate to employment for both men and women. We also find that living in areas with higher levels of unemployment decreases the rate of transition into employment and increases the rate of transition into inactivity.

Overall, we find both transition rates, to employment and to economic inactivity, are duration dependent. Men have a shorter transition into employment, and women have a shorter transition into economic inactivity. This is likely an indication that men are favoured in the labour market and are therefore more likely to receive job offers. It could also be thought that men have a lower reservation wage than women, and are therefore likely to accept the first job offer they receive. For transition into economic inactivity, this could be an indication that women are expected to stay at home as either caregivers or home makers, but they are also persistent in their job search, and therefore take longer to exit the into economic inactivity.

Chapter 5

Summary of Findings and Conclusion

State-provided school inputs of expenditure per pupil and pupil-teacher ratios influence educational attainment of Africans in South Africa. Both inputs are most important in the attainment of secondary education, but they have declining effects on attainment of tertiary education. Lower pupil-teacher ratios increase the probability of attaining lower levels of education, but decrease the probability of attaining higher levels of education. They therefore have a 'lagged effect' on educational attainment. Contrary to other empirical evidence, the ratio does not have cumulative effects on attainment. The effect of the ratio is in agreement with the views that younger students are more reliant on teachers , and at higher levels of education other factors become more important. Increased expenditure per pupil has a negative effect on attainment of general education, suggesting existence of compensatory funding for poor schools at this level. An increase in the amount of expenditure per pupil does not lead to an increase in education attainment, thus inefficiency in the use of these funds is likely. This calls for measures to ensure these funds are used more productively. State-provided inputs play a significant role in determining educational attainment, but other factors such as individual and socio-economic characteristics are also important. The results also suggest that the secondary-higher education nexus is driven by factors outside of expenditure per pupil by state, and the pupil-teacher ratio

The possibility of inefficiency in the use of public funds for education should not deter continued public funding since the presence of social returns to education justifies the funding. There are social (external) benefits to education over and above private returns to education, and the external benefits to education are significantly larger than the private benefits. The spillover effects from higher education are also high, but contrary to theory's prediction, the spillover effects are higher for college graduates than for those with less education than college diplomas. There are two main likely drivers of this effect: The presence of a high skills premium resulting from inadequate supply of skilled workers, and high unionisation amongst less skilled workers, which could prevent wages from being market driven. Labour market institutions therefore seem to play a significant role in shaping wage outcomes in South Africa.

Despite the existing deficit in highly skilled workers, high unemployment has been persistent in

South Africa. Long unemployment duration is experienced by both men and women. In transition from unemployment, men are more likely to transit into employment, while women are more likely to transit into economic inactivity. The effects of labour supply factors on transition into either employment or economic inactivity differ for men and women. These factors include education level attained, age, race, and non-labour income, among others. Women with a tertiary education have a higher probability of employment that men with similar education, and are less likely to exit the labour market. There are, therefore, higher returns in education for women. Employment equity also influences transition to employment for two men with education, as this places a premium on the hiring of women. Labour demand factors, such as unemployment rate at district council level and employment in the manufacturing sector, also affect these transitions in a gender-specific way.

Bibliography

Acemoglu, D. and Angrist, J.: 2001, How large are human-capital externalities? evidence from compulsory-schooling laws, *NBER Macroeconomics Annual 2000, Volume 15*, MIT PRess, pp. 9–74.

Addison, J. T. and Portugal, P.: 1998, Some specification issues in unemployment duration analysis, *Labour Economics* **5**(1), 53–66.

Addison, J. T. and Portugal, P.: 2003, Unemployment duration competing and defective risks, *Journal of Human Resources* **38**(1), 156–191.

Ahn, N. and Ugidos-Olazabal, A.: 1995, Duration of unemployment in spain: relative effects of unemployment benefit and family characteristics, *Oxford Bulletin of Economics and Statistics* **57**(2), 249–264.

Akerhielm, K.: 1995, Does class size matter?, *Economics of Education Review* **14**(3), 229–241.

Angrist, J. D. and Lavy, V.: 1999, Using maimonides' rule to estimate the effect of class size on scholastic achievement, *The Quarterly Journal of Economics* **114**(2), 533–575.

Angrist, J. D. and Pischke, J.-S.: 2008, *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

Angrist, J. and Krueger, A. B.: 2001, Instrumental variables and the search for identification: From supply and demand to natural experiments, *Technical report*, National Bureau of Economic Research.

Ardington, C., Case, A. and Hosegood, V.: 2009, Labor supply responses to large social transfers: Longitudinal evidence from south africa, *American Economic Journal: Applied Economics* **1**(1), 22–48.

Arulampalam, W. and Stewart, M. B.: 1995, The determinants of individual unemployment durations in an era of high unemployment, *The Economic Journal* pp. 321–332.

Aturupane, H., Glewwe, P. and Wisniewski, S.: 2013, The impact of school quality, socioeconomic factors, and child health on students academic performance: evidence from sri lankan primary schools, *Education Economics* **21**(1), 2–37.

Behrman, J. R.: 2010, Investment in education inputs and incentives, *Handbook of Development Economics* **5**, 4883–4975.

Bertrand, M., Mullainathan, S. and Miller, D.: 2003, Public policy and extended families: Evidence from pensions in south africa, *the world bank economic review* **17**(1), 27–50.

Betcherman, G.: 2012, Labor market institutions: A review of the literature.

Bhorat, H., Goga, S. and Van Der Westhuizen, C.: 2012, Institutional wage effects: Revisiting union and bargaining council wage premia in south africa, *South African Journal of Economics* **80**(3), 400–414.

Bhorat, H., Kanbur, R. and Mayet, N.: 2012, Minimum wage violation in south africa, *International Labour Review* **151**(3), 277–287.

Bhorat, H., Kanbur, R. and Mayet, N.: 2013, The impact of sectoral minimum wage laws on employment, wages, and hours of work in south africa, *IZA Journal of Labor & Development* **2**(1), 1–27.

Bhorat, H. and Oosthuizen, M.: 2008, Determinants of grade 12 pass rates in the post-apartheid south african schooling system, *Journal of African Economies* **18**(4), 634–666.

Bhorat, H., Van der Westhuizen, C. and Jacobs, T.: 2009, *Income and non-income inequality in post-apartheid South Africa: what are the drivers and possible policy interventions?*, Development Policy Research Unit, University of Cape Town.

Boes, S. and Winkelmann, R.: 2006, Ordered response models, *Modern Econometric Analysis*, Springer, pp. 167–181.

Böheim, R. and Taylor, M. P.: 2000, Unemployment duration and exit states in britain, *Technical report*, ISER Working Paper Series.

Booysen, L.: 2007, Barriers to employment equity implementation and retention of blacks in management in south africa, *South African Journal of Labour Relations* **31**(1), 47–71.

Brick, K. and Mlatsheni, C.: 2007, Examining the degree of duration dependence in the western cape labour market, *Technical report*, Southern Africa Labour and Development Research Unit, University of Cape Town.

Brown, M., Daniels, R., de Villers, L., Leibbrandt, M. and Woolard, I.: 2012, National income dynamics study wave 2 user manual., *technical manual*, South African Labour and Development Research Unit.

Burger, R. and Jafta, R.: 2006, Returns to race: Labour market discrimination in post-apartheid south africa, *Department of Economics, Stellenbosch University, Working Papers* **4**, 2006.

Cahuc, P. and Zylberberg, A.: 2004, *Labor economics*, 1 edn, MIT press.

Card, D. and Krueger, A. B.: 1992, Does school quality matter? returns to education and the characteristics of public schools in the united states, *Journal of Political Economy* **100**(1), 1–40.

Case, A. and Deaton, A.: 1999, School inputs and educational outcomes in south africa, *The Quarterly Journal of Economics* **114**(3), 1047–1084.

Chernichovsky, D.: 1977, Indexing household investment in formal education of school-age children in developing economies, *Technical report*, The World Bank.

Choe, C. and Chrite, E. L.: 2014, Internal migration of blacks in south africa: An application of the roy model, *South African Journal of Economics* **82**(1), 81–98.

Ciccone, A. and Peri, G.: 2006, Identifying human-capital externalities: Theory with applications, *The Review of Economic Studies* **73**(2), 381–412.

Council for Higher Education, .: 2004, South african higher education in the first decade of democracy, *Technical report*, Council for Higher Education (CHE), Pretoria.

Council on Higher Education, .: 2013, Vitalstats public higher education 2011, *Technical report*, Council on Higher Education.

Cox, D. R. and Oakes, D.: 1984, *Analysis of survival data*, Vol. 21, CRC Press.

Dahl, G. B.: 2002, Mobility and the return to education: Testing a roy model with multiple markets, *Econometrica* **70**(6), 2367–2420.

Das, J., Dercon, S., Habyarimana, J., Krishnan, P., Muralidharan, K. and Sundararaman, V.: 2011, School inputs, household substitution, and test scores, *Technical report*, National Bureau of Economic Research.

Dearden, L., Ferri, J. and Meghir, C.: 2002, The effect of school quality on educational attainment and wages, *Review of Economics and Statistics* **84**(1), 1–20.

Dendir, S.: 2006, Unemployment duration in poor developing economies: Evidence from urban ethiopia, *The Journal of Developing Areas* **40**(1), 181–201.

Department of Basic Education, .: 2011, Macroindicators trends in schooling summary report 2011, *Technical report*, Department of Basic Education.

Department of Education, .: 2003, A review of financing, resourcing and costs of education in public schools, *Technical report*, Department of Education, pretoria, South Africa.

Du, F. and Dong, X.-y.: 2009, Why do women have longer durations of unemployment than men in post-restructuring urban china?, *Cambridge Journal of Economics* **33**(2), 233–252.

Eyal, K. and Woolard, I.: 2011, Female labour force participation and south africaś child support grant, CSAE.

Filiztekin, A.: 2011, Social returns to education in a developing country.

Financial and Fiscal Commission, .: 2012, Technical report: Submission for 2013/14 division of revenue., *technical report*, Pretoria: Financial and Fiscal Commission.

Galiani, S. and Hopenhayn, H. A.: 2003, Duration and risk of unemployment in argentina, *Journal of Development Economics* **71**(1), 199–212.

Glewwe, P.: 2002, Schools and skills in developing countries: Education policies and socioeconomic outcomes, *Journal of economic literature* **40**(2), 436–482.

Glewwe, P. and Jacoby, H.: 1994, Student achievement and schooling choice in low-income countries: Evidence from ghana, *Journal of Human Resources* pp. 843–864.

Glewwe, P. and Kremer, M.: 2006, Schools, teachers, and education outcomes in developing countries, *Handbook of the Economics of Education* **2**, 945–1017.

Glewwe, P. W., Hanushek, E. A., Humpage, S. D. and Ravina, R.: 2011, School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010, *Technical report*, National Bureau of Economic Research.

Glick, P. and Sahn, D. E.: 2000, Schooling of girls and boys in a west african country: the effects of parental education, income, and household structure, *Economics of education review* **19**(1), 63–87.

Gonzalo, M. T. and Saarela, J.: 2000, Gender differences in exit rates from unemployment: evidence from a local finnish labour market, *Finnish Economic Papers* **13**(2), 129–139.

Graddy, K. and Stevens, M.: 2003, The impact of school inputs on student performance: An empirical study of private schools in the uk.

Gradin, C.: 2014, Race and income distribution: Evidence from the usa, brazil and south africa, *Review of Development Economics* **18**(1), 73–92.

Grogan, L. and Van den Berg, G. J.: 2001, The duration of unemployment in russia, *Journal of Population Economics* **14**(3), 549–568.

Guarcello, L., Manacorda, M., Rosati, F., Fares, J., Lyon, S. and Valdivia, C.: 2005, School-to-work transitions in sub-saharan africa: An overview, *ILO-UNICEF-World Bank* .

Gustafsson, M.: 2007, Using the hierarchical linear model to understand school production in south africa, *South African Journal of Economics* **75**(1), 84–98.

Hall, K. and Giese, S.: 2008, Addressing quality through school fees and school funding, *South African child gauge* **2009**, 35–40.

Hanushek, E.: 2013, Economic growth in developing countries: The role of human capital, *Economics of Education Review* .

Hanushek, E. A.: 1995, Interpreting recent research on schooling in developing countries, *The world bank research observer* **10**(2), 227–246.

Hanushek, E. A. and Woessmann, L.: 2011, How much do educational outcomes matter in oecd countries?, *Economic Policy* **26**(67), 427–491.

Hertz, T.: 2003, Upward bias in the estimated returns to education: Evidence from south africa, *American Economic Review* pp. 1354–1368.

Heuermann, D.: 2011, Human capital externalities in western germany, *Spatial Economic Analysis* **6**(2), 139–165.

Heyneman, S. P.: 2004, International education quality, *Economics of Education Review* **23**(4), 441–452.

Johnson, D. W.: 1981, Student-student interaction: The neglected variable in education, *Educational researcher* pp. 5–10.

Kane, T. J., Rouse, C. E. and Staiger, D.: 1999, Estimating returns to schooling when schooling is misreported, *Technical report*, National bureau of economic research.

Keswell, M. and Poswell, L.: 2004, Returns to education in south africa: A retrospective sensitivity analysis of the available evidence, *South African Journal of Economics* **72**(4), 834–860.

Kiefer, N. M.: 1988, Economic duration data and hazard functions, *Journal of economic literature* pp. 646–679.

King, E. M. and Lillard, L. A.: 1983, *Determinants of schooling attainment and enrollment rates in the Philippines*, Rand Corporation Santa Monica, Calif.

King, E. M. and Lillard, L. A.: 1987, Education policy and schooling attainment in malaysia and the philippines, *Economics of Education Review* **6**(2), 167–181.

Kingdon, G. and Knight, J.: 2006, The measurement of unemployment when unemployment is high, *Labour Economics* **13**(3), 291–315.

Kingdon, G. and Knight, J.: 2007, Unemployment in south africa, 1995–2003: causes, problems and policies, *Journal of African Economies* **16**(5), 813–848.

Kirby, S. and Riley, R.: 2008, The external returns to education: Uk evidence using repeated cross-sections, *Labour Economics* **15**(4), 619–630.

Klasen, S. and Woolard, I.: 2009, Surviving unemployment without state support: unemployment and household formation in south africa, *Journal of African economies* **18**(1), 1–51.

Krueger, A. B.: 2003, Economic considerations and class size*, *The Economic Journal* **113**(485), F34–F63.

Lam, D., Leibbrandt, M. and Mlatsheni, C.: 2008, Education and youth unemployment in south africa, *Technical report*, Southern Africa Labour and Development Research Unit, University of Cape Town.

Lam, D., Leibbrandt, M. and Mlatsheni, C.: 2010, Human capital, job search, and unemployment among young people in south africa, *meeting of Population Association of America, Dallas*.

Lancaster, T.: 1990, *The analysis of transition data*, Cambridge University Press.

Lange, F. and Topel, R.: 2006, The social value of education and human capital, *Handbook of the Economics of Education* **1**, 459–509.

Lee, J.-W. and Barro, R. J.: 2001, Schooling quality in a cross–section of countries, *Economica* **68**(272), 465–488.

Leibbrandt, M., Woolard, I. and de Villiers, L.: 2009, methodology: report on nids wave 1, *technical paper no. 1*, South African Labour and Development Research Unit.

Lentz, R. and Tranæs, T.: 2005, Marriage, wealth, and unemployment duration: A gender asymmetry puzzle.

Levinsohn, J. and Pugatch, T.: 2010, The role of reservation wages in youth unemployment in south africa: A structural approach, *University of Michigan, mimeo* .

Liu, Z.: 2007, The external returns to education: evidence from chinese cities, *Journal of Urban Economics* **61**(3), 542–564.

Magruder, J. R.: 2012, High unemployment yet few small firms: The role of centralized bargaining in south africa, *American Economic Journal: Applied Economics* **4**(3), 138–166.

Mani, S., Hoddinott, J. and Strauss, J.: 2009, Determinants of schooling outcomes: Empirical evidence from rural ethiopia, *Fordham University, Department of Economics Discussion Paper* (2009-03).

Massyn, N., Day, C., Peer, Naziaand Padarath, A., Barron, P. and English, R.: 2014, District health barometer 2013/14, *Technical report*, Health System Trust.

Michaud, P.-C. and Vencatachellum, D.: 2003, Human capital externalities in south africa*, *Economic Development and Cultural Change* **51**(3), 603–628.

Mingat, A. and Tan, J.-P.: 1996, The full social returns to education: Estimates based on countries' economic growth performance, *Human Capital Development Papers. Washington: World Bank* .

Moloi, M. Q. and Chetty, M.: 2011, Quality of primary school inputs in south africa, *Policy brief*, department of Basic Education.

Moretti, E.: 2004a, Human capital externalities in cities, *Handbook of regional and urban economics* **4**, 2243–2291.

Moretti, E.: 2004b, Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data, *Journal of econometrics* **121**(1), 175–212.

Mortensen, D. T.: 1986, Job search and labor market analysis, *Handbook of labor economics* **2**, 849–919.

Muravyev, A.: 2008, Human capital externalities evidence from the transition economy of russia1, *Economics of Transition* **16**(3), 415–443.

Murtin, F.: 2013, Improving education quality in south africa.

Mwabu, G. and Schultz, T. P.: 2000, Wage premiums for education and location of south african workers, by gender and race, *Economic Development and Cultural Change* **48**(2), 307–334.

Naudé, W. A. and Krugell, W. F.: 2003, An inquiry into cities and their role in subnational economic growth in south africa, *Journal of African Economies* **12**(4), 476–499.

Nazier, H.: 2013, Higher education externalities in egyptian labr market, *Advances in Management and Applied Economics* **3**(2), 193–217.

Ollikainen, V.: 2003, The determinants of unemployment duration by gender in finland, *VATT Discussion Papers* **316**.

Posel, D., Casale, D. and Vermaak, C.: 2014, Job search and the measurement of unemployment in south africa, *South African Journal of Economics* **82**(1), 66–80.

Psacharopoulos, G. and Patrinos, H. A.: 2004, Returns to investment in education: a further update, *Education economics* **12**(2), 111–134.

Pugatch, T.: 2012, Bumpy rides: School to work transitions in south africa, *Technical report*, Discussion Paper series, Forschungsinstitut zur Zukunft der Arbeit.

Ranchhod, V.: 2009, Household responses to adverse income shocks: Pensioner out-migration and mortality in south africa.

Ranchhod, V. and Dinkelman, T.: 2007, Labour market transitions in south africa: what can we learn from matched labour force survey data?

Rauch, J. E.: 1993, Productivity gains from geographic concentration of human capital: evidence from the cities, *Journal of urban economics* **34**(3), 380–400.

Republic of South Africa, .: 1998b, National norm and standards for school funding, south african school act, 1996, *Technical report*, Department of Education.

Rodrik, D.: 2008, Understanding south africa's economic puzzles*, *Economics of Transition* **16**(4), 769–797.

Rolleston, C. and Krutikova, S.: 2014, Equalising opportunity? school quality and home disadvantage in vietnam, *Oxford Review of Education* (ahead-of-print), 1–20.

SALDRU: 2013, National income dynamics study 2012, wave 3, *Technical report*, Southern African Labour and Development Research Unit.

Sand, B. M.: 2013, A re-examination of the social returns to education: Evidence from us cities, *Labour Economics* **24**, 97–106.

Schultz, T. P.: 2004, Evidence of returns to schooling in africa from household surveys: Monitoring and restructuring the market for education, *Journal of African Economies* **13**(suppl 2), ii95–ii148.

Schündeln, M. and Playforth, J.: 2014, Private versus social returns to human capital: Education and economic growth in india, *European Economic Review* **66**, 266–283.

Selod, H. and Zenou, Y.: 2003, Private versus public schools in post-apartheid south african cities: Theory and policy implications, *Journal of Development Economics* **71**(2), 351–394.

Serneels, P.: 2007, The nature of unemployment among young men in urban ethiopia, *Review of Development Economics* **11**(1), 170–186.

Spreen, C. A. and Vally, S.: 2006, Education rights, education policies and inequality in south africa, *International Journal of Educational Development* **26**(4), 352–362.

Statistics South Africa, .: 1998, Unemploymnet and employment in south africa, *Technical report*, Statistics South Africa.

Statistics South Africa, .: 2013, National and provincial labour market trends 2003-2013, *Statistical release p0211.4*, Statistics South Africa (SSA).

Tansel, A.: 1997, Schooling attainment, parental education, and gender in cote d'ivoire and ghana, *Economic Development and Cultural Change* **45**(4), 825–856.

Tansel, A.: 2002, Determinants of school attainment of boys and girls in turkey: individual, household and community factors, *Economics of Education Review* **21**(5), 455–470.

Tansel, A. and Tasci, H. M.: 2004, Determinants of unemployment duration for men and women in turkey.

Taylor, N., Fleisch, B. and Shindler, J.: 2008, Changes in education since 1994, *Higher Education* **32**, 3–5.

Thomas, A. and Jain, H. C.: 2004, Employment equity in canada and south africa: progress and propositions, *The International Journal of Human Resource Management* **15**(1), 36–55.

United Nation, .: 2010, Keeping the promise: united to achieve the millennium development goals, *technical manual*, United Nation.

United Nation, .: 2013, The millennium development goals report 2013, *Technical report*, United Nation.

Van der Berg, S.: 2007, Apartheid's enduring legacy: Inequalities in education, *Journal of African Economies* **16**, 849–880.

Van der Berg, S. and Moses, E.: 2011, Inertia in south african education quality: evidence from sacmeq ii and iii, *Policy brief*, department of economics, University of Stellenbosch.

Van der Berg, S., Taylor, S., Gustafsson, M., Spaull, N. and Armstrong, P.: 2011b, Improving education quality in south africa.

Wangenge-Ouma, G. and Cloete, N.: 2008, Financing higher education in south africa: Public funding, non-government revenue and tuition fees, *South African Journal of Higher Education* **22**(4), 906–919.

Williams, M. J.: 2007, *The social and economic impacts of South Africaś child support grant*, PhD thesis, Williams College Doctoral Dissertation.

Williams, R.: 2006, Generalized ordered logit/partial proportional odds models for ordinal dependent variables, *Stata Journal* **6**(1), 58–82.

Wilson, K.: 2001, The determinants of educational attainment: Modeling and estimating the human capital model and education production functions, *Southern Economic Journal* pp. 518–551.

Wittenberg, M.: 2002, Job search in south africa: A nonparametric analysis*, *South African Journal of Economics* **70**(8), 1163–1196.

wittenberg, M.: 2009, National income dynamics study wave 2 user manual., *technical paper 2l*, South African Labour and Development Research Unit.

Wossmann, L.: 2003, Schooling resources, educational institutions and student performance: the international evidence, *Oxford Bulletin of Economics and Statistics* **65**(2), 117–170.

Yamauchi, F.: 2011, School quality, clustering and government subsidy in post-apartheid south africa, *Economics of Education Review* **30**(1), 146–156.

Zhao, M. and Glewwe, P.: 2010, What determines basic school attainment in developing countries? evidence from rural china, *Economics of Education Review* **29**(3), 451–460.