Measuring social and economic mobility in South Africa: New approaches to well-known survey data concerns

By Asmus Zoch

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

Supervisor: Prof. Rulof Burger

Co-Supervisor: Prof. Servaas van der Berg

December 2016

Declaration

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

With regard to chapter 2, the nature and scope of my contribution in Chapter 2 were as follows:

Nature of Contribution	Extent of Contribution
1. Data cleaning and all Stata coding 2. Estimation and data analysis 3. Write up of literature review 4. Write up of results 5. All additional manuscript text 6. All tables and figures	75%

The following co-authors have contributed to chapter 2:

Name	e-mail address	Nature of Contribution	Extent of Contribution
Rulof Burger (Thesis supervisor)	rulof@sun.ac.za	1. Conceptualising methodology 2. Review and comments on chapter content	15%
Patrizio Piraino	patrizio.piraino@uct.ac.za	1. Conceptualising research question 2. Review and comments on chapter content	10%

Signature of candidate: *Asmus Zoch*
Date: 26/07/2016

Declaration by co-authors:

The undersigned hereby confirm that

1. the declaration above accurately reflects the nature and extent of the contributions of the candidate and the co-authors to chapter 2,
2. no other authors contributed to chapter 2
3. potential conflicts of interest have been revealed to all interested parties and that the necessary arrangements have been made to use the material in chapter 2.

Signature	Institutional affiliation	Date
Rulof Burger:	Stellenbosch University	-
Patrizio Piraino:	Cape Town University	Declaration with signature in possession of candidate and supervisor

Abstract

The aim of this dissertation is measuring economic and social mobility in South Africa. The work from this thesis shows that various problems with survey responses can produce misleading impressions of the South African labour market and of income dynamics. This highlights the importance of measuring variables of interest accurately and to carefully consider the ways in which unreliable responses can bias the results of conventional estimators. It will be demonstrated that even the most appropriate estimator and identification strategy can fail to yield unbiased estimates if important measurement issues are ignored. To address these shortcomings this thesis applies and adapts new approaches to remedy well-known survey data reliability concerns. The most important findings of the three chapters are as follows:

First, in the context of high unemployment and weak labour market attachment for many South African youth, the formulation of survey questions matters and approaches commonly used to elicit reliable responses in developed countries cannot be assumed to work equally well. This is particularly true for subjective measures. Hence, answers to the traditional question on reservation wages may fail to provide meaningful answers. It appears that different formulations and ordering of the reservation wage question can trigger different cognitive processes in the respondent that elicit different answers. However, using a series of questions intended to elicit a more accurate response, the new reservation wage measure seems to be more internally consistent and the regression results to be in line with labour market search models.

Second, this thesis shows that estimating the speed of convergence between the poorest and richest households using micro growth regressions without controlling for measurement error would overestimate income mobility significantly. Therefore, a newly developed GMM estimator was applied to four large national panel studies to obtain less biased β estimates. The findings of four large representative national panel studies from the USA, South Africa, Chile and Tanzania show that naïve OLS regression coefficients would overestimate the extent of income mobility by a factor of about 4-6. The hypothesis of no measurement error can be rejected for all the countries observed. While the data reliability for the US, Chile and Tanzania correspond to their levels of economic development, South Africa's data reliability appears to be unexpectedly high. The nonparametric estimates also show that the speed of convergence varies over the income distribution and that income is more reliably captured for richer than for poorer households.

Third, the relative importance of family, neighbourhood and school quality in explaining variation in socio-economic outcomes are evaluated. Using spatial merging techniques to combine different data sets, new school wealth quintiles have been created that predict individual learner and school outcomes more accurately than the old school quintiles. This chapter provides evidence of the importance of quality education in explaining university enrolment. In addition, there seem to be a significant premium for quality education in labour markets earnings regressions, which confirms the long-term importance of schooling.

.

Abstrak

Die doel van hierdie tesis is om die ekonomiese en sosiale mobiliteit in Suid-Afrika te meet. Die werk wat gedoen is, toon dat verskeie probleme met opname-response tot 'n wanindruk oor die Suid-Afrikaanse arbeidsmark en inkomste-dinamiek kan lei. Dit beklemtoon hoe belangrik dit is om die veranderlikes van belang akkuraat te meet, en om die maniere waarop onbetroubare response die uitkomste van konvensionele beramers kan beïnvloed, versigtig te oorweeg. Daar sal geïllustreer word dat selfs die mees toepaslike beraming- en identifikasiestrategie tot bevooroordeelde beramings kan lei wanneer belangrike metingskwessies verontagsaam word. Ten einde hierdie tekortkominge te bowe te kom, word nuwe en aangepaste benaderings in die studie gebruik om bekende betroubaarheidskwessies rakende opname-data uit die weg te ruim. Die belangrikste bevindinge van die drie hoofstukke is soos volg:

Eerstens, in 'n omgewing waar baie Suid-Afrikaanse jongmense onderhewig is aan 'n hoë voorkoms van werkloosheid en swak koppeling aan die arbeidsmark, maak die formulering van opname-vrae saak. Daar kan nie aanvaar word dat metodes wat algemeen gebruik word om betroubare response in ontwikkelde lande te verkry, net so doeltreffend sal wees nie, veral in die geval van subjektiewe maatstawwe. Antwoorde op die tradisonele vraag oor reserwelone kan dus niksseggend wees. Dit kom voor asof die vraag, wanneer dit anders geformuleer of georganiseer word, aanleiding gee tot ander kognitiewe prosesse in die respondent, wat ander antwoorde ontlok. Deur 'n reeks vrae te gebruik wat daarop gemik is om 'n meer akkurate respons uit te lok, is die nuwe maatstaf vir reserwelone op die oog af intern meer konsekwent en die regressie-resultate in ooreenstemming met werksoekmodelle.

Tweedens wys hierdie tesis dat inkomstemobiliteit aansienlik oorskat word by die beraming van konvergensiespoed tussen die armste en rykste huishoudings indien mikro-groeiregressies gebruik word sonder om metingsfoute in ag te neem. 'n Nuwe GMM-beramer is derhalwe op vier groot, nasionale paneelstudies toegepas om minder bevooroordeelde β-beramings te kry. Die bevindinge van vier groot, verteenwoordigende paneelstudies in die VSA, Suid-Afrika, Chili en Tanzanië toon dat naïewe OLS- (Ordinary Least Squares) regressie-koëffisiënte die mate van inkomstemobiliteit met 'n faktor van 4-6 oorskat. Die hipotese van geen metingsfoute kan verwerp word in die geval van al die betrokke lande. Terwyl die betroubaarheid van die data in die geval van die VSA, Chili en Tanzanië ooreenstem met die vlak van hul ekonomiese ontwikkeling, lyk die data-betroubaarheid van Suid-Afrika onverwags hoog. Die nie-parametriese beramings wys ook dat die konvergensiespoed met betrekking tot die inkomsteverspreiding wissel, en dat inkomste meer betroubaar weergegee word by ryker as by armer huishoudings.

Derdens word die relatiewe belangrikheid van die gesin, die woonbuurt en die kwaliteit van die skool geëvalueer ten einde variasies in sosio-ekonomiese uitkomste te verklaar. Deur ruimtelike samesmeltingstegnieke te gebruik om verskilende datastelle te kombineer, is nuwe welvaartskwintiele vir skole geskep. Hierdie nuwe kwintiele voorspel individuele leerder- en skooluitkomste akkurater as die ou skoolkwintiele. Die bewys van hoe belangrik kwaliteitonderrig is, word verskaf wanneer universiteitsinskrywings verduidelik word. Dit kom boonop voor asof

kwaliteitonderwys in arbeidsmarkte met verdienstefunksies sterk in aanvraag is. Dit bevestig die langtermyn belangrikheid van onderwys.

Acknowledgements

I want to say thank you because I feel very lucky and privileged for the time I had as a PhD student in Stellenbosch. I know how much struggle a PhD can contain and I want to say thank you to everyone that made my journey great fun, educationally enriching and overall a wonderful experience. Numerous people have been very kind, helpful and supportive but there are some people who I would like to thank especially.

First and foremost I would like to thank my supervisor Professor Rulof Burger who has been an amazing mentor and great help in tackling any econometric problem. He has invested more time and effort than you could ever expect from a supervisor and has been a friend since the day we started working together. I am forever indebted to my second supervisor Professor van der Berg who was the first person in South Africa who believed in me and always had my back when I needed him since then. Special thanks go to the other Burgers, Professor Ronelle and Dr. Cobus Burger. They have been a source of inspiration and great support from the beginning till the end of my journey.

To all my other fantastic colleagues from RESEP at Stellenbosch, PODER fellows and friends in Göttingen I say thank you for all the helpful comments, ideas and time you invested in helping me with my research. I show extensive gratitude to my three examiners that spent so much time and energy to provide me with lots of helpful feedback to further improve and finalize this thesis. Finally, I would like to thank my family, as well as my dearest friends who have always believed and encouraged me in my work.

Table of Contents

Chapter 1

Introduction and research question

South Africa is an extraordinarily unequal country. The richest 10% of the population own 90-95% of the assets and earn 55-60% of all income (Orthofer, 2016), which is higher than in any other country for which comparable data exists. Even though legalised racial discrimination ended with political transition in 1994, the distributions of wealth and income continue to be highly dependent on race. Moreover, the place of birth and the socio-economic status of parents remain a strong predictor of a child's probability of completing or continuing beyond secondary education (WB, 2012; Zoch, 2015). A relatively small decrease in inequality between races since the end of apartheid has coincided with rising inequality within the racial groups (Leibbrandt et al., 2011). Despite the positive indication that wealth and poverty are being distributed less along racial lines and that a new affluent black elite and middle class have emerged, poor black households are falling behind in relative terms. Against this backdrop of high inequality and persistent racial bias, this dissertation investigates different aspects of social and economic mobility with an emphasis on reliable measurement using survey data. It will demonstrate that various well-known problems with survey responses can produce a misleading impression of the South African labour market and income dynamics. Explicitly addressing these shortcomings can supply more reliable results that challenge previous findings and have important policy implications. This is achieved by introducing new approaches to analyse existing panel data and survey data linked using spatial merging techniques.

This chief scientific contribution of this research is to the literature on measurement error in survey data and how this affects our ability to obtain reliable estimates (e.g. Bound and Krueger, 1991; Bound et al., 2001; Akee, 2011; and Glewwe, 2012). On the basis of three chapters on social- and economic mobility, it examines the importance of measuring variables of interest accurately and carefully considering the ways in which unreliable responses can bias the results of conventional estimators. The dissertation will demonstrate that even the most appropriate estimator and identification strategy will sometimes fail to yield unbiased estimates if important measurement issues are ignored.

These findings offer a reminder of at least three lessons. First, the formulation of survey questions matters and approaches commonly used to elicit reliable responses in developed countries cannot be assumed to work equally well in different developing country contexts. This is particularly true for subjective measures, and it is worth exploring whether different question formulations that require different cognitive processes provide consistent responses. Second, the effect of measurement error is well-documented to be exacerbated in panel data estimators that study the changes in variables between successive waves, but this fact is often stated without carefully exploring the magnitude of the resulting bias. This dissertation shows that the bias can be severe under certain circumstances, and lead to very misleading estimates. Third, omitting variables that are difficult to observe but highly correlated to observable variables of interest, results in the well-known omitted variable bias. Using additional data sources to construct proxy variables will help alleviate this bias, and can provide a much more accurate understanding of the causal channels that determine outcomes.

The three thesis chapters that follow investigate how survey reliability concerns can affect the measurement of social- and economic mobility. Chapter 2, the first substantive chapter, explores how reliable self-reported reservation wages are. Chapter 3 estimates income mobility in a cross country comparison. The fourth Chapter estimates the effect of neighbourhoods and school quality on education outcomes and earnings. The remainder of this introductory chapter first gives some background to the South African economy in order to contextualise this dissertation and to demonstrate the relevance of the three research chapters. Thereafter, for each chapter the methodology, general findings and policy implications will be shorty discussed.

1.1. Context

While the South African economy has experienced modest GDP growth for the last two decades, around 3.4 percent per annum since 1995 (WB, 2012), it is evident that economic growth was not fast enough to create enough jobs for the increased labour supply after the fall of apartheid. As a consequence unemployment (by the broad definition) rose from 31% in 1995 to 42% in 2003 and was about 36% in 2007 (Burger et al., 2015). Besides structural and institutional reasons, unemployment also increased due to the mismatch of mostly unskilled worker supply and demand for high-skill labour (Banerjee et al. 2008). Branson and Wittenberg (2007) concludes that the recent generation of young labour work participants are better qualified and left school at an earlier age than their parents' generation. Nevertheless, South African youth unemployment rate for 2005 was about 50% for the cohort aged 15-24 which is much higher than in other comparable sub-Saharan countries (Rankin and Roberts, 2011). One explanation for this phenomenon could be household formation, where the unemployed may migrate and attach themselves to households with some kind of employment income or state transfers (Klasen and Woolard, 2009). In particular, social grants that are significantly higher than in other sub-Saharan countries and could increase the reservation wage and lower labour force participation (Leibbrandt et al., 2011; Abel, 2013). Similarly, high expectations following the political transformation could have led to unrealistically high perception of wage possibilities in South Africa and further increase the reservation wage of the youth (Banerjee et al. 2008). Interestingly previous studies that have looked at the direct effect of reservation wages on employment probability haven't found any adverse effects (Kingdon and Knight, 2001; Nattrass and Walker, 2005). Hence, chapter 2 will reconsider whether reservation wage can explain labour market behaviour if measured more precisely.

While school to work transition is a key element in understanding who is getting ahead and who is falling behind, chapters 3 and 4 will deal more directly with the concept of social- and economic mobility. As mentioned before, South Africa is one of the most unequal societies in the world, with a Gini coefficient of 0.70 in 2008 (WB, 2012). There are many different reasons for this situation. Yet, at the heart of the inequality lies the rising inequality within the labour market, both due to the increase in unemployment as well as rising earnings inequality. This again can be attributed to the above mentioned mismatch between low skill labour supply and high skill labour demand, which is partly caused by the failure of the education system to provide students from low socio-economic status with quality education (Van der Berg, 2007, 2009). In this context, chapter 3 will measure how income mobile South African households are. As upward economic mobility continues to be one of the most important indicators for economic development, the chapter sets out to estimate the true speed of convergence between the poorest and richest household after controlling for measurement error in self-reported per capita household income.

A key feature of the apartheid regime was that non-whites were restricted in the location they were allowed to live in as well as limited in their access to affordable schooling. This segregation between races has seen long-lasting impacts that can still be observed in today's distribution of neighbourhoods, travel costs to central formal jobs and education outcomes (e.g. Van der Berg, 2007, 2009; Lam et al. 2011). Yet, not much is known about the direct effect of spatial location and the relative importance of family, neighbourhood and the school children are attending (Altonji and Mansfield, 2011; Jenning et al., 2015). Specifically, no existing empirical studies have attempted to distinguish between the effects of residential neighbourhoods or family background. However, the answer to this question is important for at least two reasons. First, there are important policy implications of knowing whether it is the school or the family background which is determining education outcomes. If parents' social economic status is only significant because it is predicting the quality of the school their offspring go to, improving the school quality will have large effects, in particular for the poorest students. Second, in the light of South African inequality problem, knowing how much neighbourhood and school quality will determine a child's chances of reaching matric, enrolling into university and their future earnings is an important step to reduce the inequality gap of future generations and making South Africa a more fair society.

1.2. How reliable are self-reported reservation wages?

Chapter 2 demonstrates how different formulations of the same survey question can trigger different cognitive processes in the respondent that elicit different answers. This is in consistent with findings in the behavioural literature that answers to subjective questions can be influenced by a multitude of seemingly irrelevant factors (Bertrand and Mullainathan, 2001). In particular respondents that have little experience thinking about a question may provide less reliable information, unless the question is specifically formulated to require a thoughtful response. Given the enormous problem of youth unemployment in South Africa, estimating labour market behaviour and understanding the school to work transition is of the utmost importance. However, answers to the traditional question on reservation wages may fail to provide meaningful answers if the youth have a very weak attachment to the labour market and are not used to scrutinising their preferences to determine their lowest acceptable wage. This issue will be revisited using a series of questions in the Cape Area Panel Study (CAPS) of South Africa, which was intended to elicit a more accurate response to the lowest wage an individual will accept. It can be shown that the more reliable reservation wage measures behaves according to labour market theory and influence labour market outcomes. The new measure evolves over time in a way that is consistent with workers updating their beliefs regarding the wage offer distribution (i.e. learning), whereas the traditional measure does not. In addition, a simple model of employment probability shows that the new measure can predict the likelihood of finding employment or remaining employed, but the traditional measure cannot. This implies that South African youth indeed are waiting out for well paid jobs. This is in line with the findings by Rankin and Roberts (2011) showing that there is an over-demand for jobs at large formal companies, which are the only ones that can fulfil the overoptimistic wage aspirations of most workers. More generally, the chapter provides new evidence how economic models and research results can be influenced by noisy data due to respondent misperceptions and unreliable self-reports. It highlights the importance for economists to carefully design survey questions so that individuals give reliable answers that are not influenced by a variety of cognitive and non-cognitive factors.

1.3. Estimating income mobility

Chapter 3 studies the extent of economic mobility by estimating the speed of convergence between the poorest and richest households using micro growth regressions and a three wave panel data set. It is well-established that measurement error in household income can cause income to appear more mobile than is actually the case (Antman and Mckenzie, 2007a). This is particularly problematic in developing countries where household consumption and income data from household surveys can be noisy (Fields, 2008a). Hence, controlling for such measurement error, Antman and McKenzie (2007a) show that the naïve OLS estimates would overestimate mobility in Mexico by about 33%, while Glewwe (2012) finds that between 15 to 42% of estimated Vietnamese economic mobility is due to measurement error. In an attempt to build on such studies, chapter 3 applies a newly developed GMM estimator to four large panel datasets from different developing countries and the USA, each containing three waves of household income data. This method aims to quantify the effect of household income measurement error when estimating economic mobility with micro growth regressions. This approach is more efficient than the two-stage least squares (2SLS) estimator and allows testing the validity of the underlying identifying assumptions. The cross-country comparison is particularly interesting since it compares the speed of convergence in South Africa to other developing countries. Furthermore, applying the method to the USA will allow the juxtaposition of the data reliability estimated from the GMM estimator to that obtained from validation studies in the USA. The findings show that naïve OLS regression coefficients overestimate the extent of income mobility for all four countries. The GMM estimates of the expected half-life of income gaps for the US are consistent with estimates from the intergenerational literature which are less vulnerable to measurement error. The results provide support for the assumption that the GMM estimator produces realistic estimates for the speed of convergence within a country. The findings from all countries strongly indicate the presence of measurement error, in particular, for Chile and Tanzania. The nonparametric estimates also show that the speed of convergence varies over the income distribution and that income is more reliably captured for richer than for poorer households. In general, the results of this chapter indicate that income mobility can be seriously overestimated when not accounting for measurement error. This is important for economists and policy makers that attempt to estimate whether the gap between the poorest and richest households is closing over time and how long it will take for the society to become more equal.

1.4. The effect of neighbourhoods and school quality on education and labour market outcomes

Finally, Chapter 4 adds to the literature on how neighbourhood and school quality affects education and labour market outcomes. Previous studies that analysed the relative importance of family background, neighbourhood and school effects in the US and Northern Europe, mostly found only small neighbourhood or school effects (e.g. Raaum et al., 2006; Lindahl, 2011; Nicoletti and Rabe, 2013; and Schnitzlein, 2014). However, as argued by Case and Yogo (1999), location and schooling quality should matter more in countries like South Africa, where movement was restricted and resources were distributed unevenly between race groups. This study adds to the literature, by using a spatial approach to link a neighbourhood wealth index from the Census 2011 community survey to the household and school information from the National Income Dynamics Study. The more informative school wealth quintiles estimates, prove to be good proxies for school quality and have a higher predictive power in explaining schooling outcomes than existing measures of school quintiles, at least in the metro regions of South Africa. The analysis proceeds to show large differences in education outcomes depending on where a student attends school, which is mostly dependent on socio- economic status. However, this study also demonstrates that even children from the poorest neighbourhood would perform well if they go to one of the richest 20% of schools. Yet, given the limited number of quality schools, the segregated location of quality schools, financial as well as transport constraints, only about 10% of children the poorest 60% actually attend a top quintile schools. In order to achieve more equal education outcomes, the quality of schools in the poor neighbourhoods would need to be very considerably improved. The chapter also provides evidence that the benefits of improved school quality include an improved likelihood of tertiary enrolment. In addition, previous attempts to identify the effect of school quality on labour market wages have been confounded by using available but less informative quality measures like cognitive ability tests (du Rand et al., 2011; van der Berg et al., 2011). However, using the new school wealth index as an instrument for school quality, there seems to be a significant premium for quality education in labour markets earnings regressions, which show the long-term implications of the schooling system.

Chapter 2:

How reliable are self-reported reservation wages?

Evidence from South Africa

2.1. Introduction

Standard search theory posits that reservation wages have an effect on both the duration of unemployment and the realized wage distribution (Mortensen, 1976, 1986; Pissarides, 2000). As a consequence, a host of empirical studies in labour economics have investigated the determinants of the reservation wage, its behaviour over spells of unemployment, and the effect of high reservation wages on key labour market outcomes (Lancaster, 1985; Jones 1988; Shimer and Werning, 2007; Addison et al., 2010, 2013; Brown and Taylor, 2011).

Empirical research on reservation wages largely relies on data from survey questions where individuals are asked to report the lowest wage they would work for.[1] Answers to subjective questions are influenced by a variety of cognitive and non-cognitive factors (Bertrand and Mullainathan, 2001). According to Kingdon and Knight (2001), individuals may provide unreliable answers on reservation wages for three reasons: first, they may have limited knowledge about the labour market. Second, respondents may be inclined to report their expected wages rather than the 'lowest wage they would accept'. Third, individuals may imagine themselves in a bargaining context when asked to state their reservation wage and therefore report higher values. Bertrand and Mullainathan (2001) discuss a number of reasons to doubt whether subjective questions in general elicit meaningful answers: the wording and ordering of questions, scaling problems and forms of social pressure from the interviewer. In addition, attitudes may not "exist" in a coherent way because respondents are reluctant to admit lack of attitude or because they misreport their attitude. In most cases, the factors (both cognitive and non-cognitive) influencing the answers to subjective questions cannot be treated as exogenous. With respect to questions about lowest acceptable wages, this introduces the possibility that empirical models in labour economics using self-reported reservation wages – as either a dependent or explanatory variable – may suffer from measurement error bias.

We contribute to the literature on the role and determinants of reservation wages by using a series of survey questions on explicit job offers as well as traditional questions on the lowest acceptable wage. Both types of

[1] The actual wording of the question varies. Examples are: "What is the lowest amount in take-home pay that you would be prepared to accept from a new job?" (Jones, 1988); "If you were offered a job as a -, would you accept it at -?"(Holzer, 1986b); "What type of work have you been looking for?" followed by "What would the wage (or salary) have to be for you to be willing to take it?" (Holzer, 1986b).

questions were asked in the Cape Area Panel Study (CAPS) in South Africa (2002-2009) with the view of eliciting a more comprehensive picture on the lowest wage respondents would truly accept. Our analysis shows that people may systematically misreport their reservation wages when asked in a traditional way and that this non-classical measurement error can bias the coefficients of common regression models in empirical labour economics. On the other hand, using a "probed" reservation wage from explicit hypothetical job offers, we find a significant negative relationship with the length of the unemployment spell. Reservation wages decline by about 4.8% per year of unemployment duration, which is in line with evidence based on high-quality data from the U.S. (Krueger and Mueller, 2014). We also observe that the probed reservation wage is negatively associated with the probability of employment. We find that a 10% increase in the reservation wage reduces the likelihood of being employed in the next period by about 0.6%. Our findings are in contrast with the evidence from other studies in South Africa based on self-reported reservation wages. The empirical analysis in the paper highlights the importance of investigating whether (and why) individuals tend to provide noisy reports of their reservation wages. This can prove useful in the formulation of appropriate survey questions. In addition, the paper's results are of relevance for the debate on the determinants and effects of reservation wages in high-unemployment (and high-inequality) countries like South Africa.

2.2. Theory and Literature review

According to job search theory, the reservation wage can play an important role in explaining unemployment duration and job acceptance. Under the assumption that wage offers are independent realizations from a known wage offer distribution and jobs, once accepted, are indefinite, Lancaster and Chesher (1983) show that the optimal reservation wage rate rw^r can be written as:

$$rw^r = b + \frac{\lambda}{\rho} \int_{w^r}^{\infty} (w - w^r) \, dF(w) \qquad (1)$$

where the parameter b is the amount of unemployment benefits net of any search costs, λ is the wage arrival rate, ρ is the discount rate of future incomes, and w is the wage offer whose cumulative distribution function is F(w). The model posits that a wage offer greater than or equal to rw^r is accepted, while a smaller offer is declined and the job search continues.

One of the immediate implications of this basic model is that higher reservation wages will decrease the probability of employment in any given period, thus increasing the expected duration of unemployment. Lancaster (1985) showed that the empirical relationship between reservation wages and duration of unemployment in England is indeed positive. That is, higher reservation wages are associated with longer spells of unemployment.[2] Holzer (1986b) finds that differences in reservation wages account for 26 to 42 percent of the black-white gap in unemployment duration in U.S data. Using data from Europe, Brown and Taylor (2013) find that the elasticity of unemployment duration with respect to the reservation wage is positive and significant. One of the empirical challenges in estimating the relationship between reservation wages and employment probabilities is that unobserved characteristics can influence both the job-arrival rate and the reservation wage. Lancaster and Chesher (1983), Jensen and Westergård-Nielsen (1987), and Addison et al. (2010) all find a significant positive correlation between the offer-rate and the reservation wage. For this reason, Bloemen and Stancanelli (2001) estimate the probability of transitioning into employment in a simultaneous-equation system as a product of the job-offer rate and the acceptance probability. Their findings indicate a negative relationship between reservation wage and probability of employment.

If the reservation wage plays an important role in explaining employment outcomes, then understanding employment requires knowing what determines reservation wages. Equation (1) shows that unemployment benefits should increase the reservation wage, as should family support and individual wealth that generates non-labour income. These hypotheses have been tested in a number of empirical studies. Lancaster and Chesher (1983), Feldstein and Poterba (1984) and Addison et al. (2009, 2010) show that the elasticity between unemployment benefits and both reservation wages and re-employment likelihood is negative and significant. Bloemen and Stancanelli (2001) show that household wealth and reservation wages are positively correlated.

While the basic model in Equation (1) assumes that reservation wages are constant over the unemployment spell, several contributions (e.g. Mortensen, 1977; Lancaster, 1985; and Jones, 1988) point out sources of non-stationarity and document a declining trend in reservation wages over the course of joblessness. This result has been confirmed in a recent study of 13 countries in the European Community Household Panel (EHCP) by Addison et al. (2013). The authors apply an econometric framework similar to Lancaster (1985) and find a small but significant negative trend of reservation wages over the unemployment spell. Such a trend can be related to a number of factors: the limited duration of unemployment benefits, liquidity-constrained job search, human capital loss over the spell, initial over-confidence, adverse signalling effects of being long-term unemployed, or learning about the job offer distribution and the availability of jobs (e.g. Burdett and Vishwanath, 1988; Kriechel and Pfann, 2006; and Krueger and Mueller, 2014). Brown et al. (2011) provide evidence of a strong decline in reservation wages over time in a real-time search laboratory experiment. They explain their findings by "non-stationary subjective costs". Using longitudinal data from weekly interviews of unemployed workers in New Jersey, Krueger and Mueller (2014) also find a decline in reservation wages over the unemployment spell. Interestingly, their results imply a rather small impact of

[2] The author uses simultaneous equations models to deal with the reverse causality problem in the relationship between unemployment spells and reservation wages. This is because reservation wages are themselves influenced by the duration of unemployment. This relationship will be discussed further below.

unemployment insurance benefits on reservation wages and a moderate decline of reservation wages over the jobless spell.

2.2.1. Empirical studies in South Africa

The South African literature on reservation wages is surprisingly limited given the extent of the unemployment problem in the country. To our knowledge, only three papers have attempted to estimate the direct effect of reservation wages on the likelihood of employment (Kingdon and Knight, 2001; Nattrass and Walker, 2005; and Rankin and Roberts, 2011). These studies obtain a prediction of the remuneration unemployed respondents could expect given their characteristics and compare it to self-reported reservation wages. Using the PSLSD (1993) and the October Household Survey (1994), Kingdon and Knight (2001) find that most unemployed respondents have higher reservation wages than their 'predicted' wage. The authors draw no further conclusions about this relationship as they consider the answers to the reservation wage question in the surveys to be unreliable. Nattrass and Walker (2005) use a dataset from Cape Town, which was explicitly designed to obtain more reliable reservation wage data, by asking unemployed workers in different ways about their "lowest acceptable wage", "minimum possible wage" and "would you work for R33 a day?". Interestingly they find inconsistencies for many of the answers of the participants and conclude that "either people respond differently to slightly different phrasing of questions about the reservation wages; or they do not have a definite reservation wage." Finally, the authors estimate Heckman-corrected predicted wages and use this information to generate a dummy variable that equals one for individuals whose reservation wage was greater than the predicted wage (and zero otherwise). Using this as a dependent variable in a probit regression, they find a positive correlation between employment likelihood and reservation wages. They conclude that high reservation wages do not result in higher unemployment and that workers have realistic wage expectations.

In a recent study, based on the South African Young Persons Survey (SAYPS), Rankin and Roberts (2011) show instead that young labour market participants overestimate their wage possibilities and that reservation wages decline until the age of 30 and following first labour market experiences. The authors suggest a number of potential factors that could explain high youth reservation wages: transport costs,[3] loss of intra-household transfers, social grants, and hoping for jobs in big rather than small firms. Banerjee et al. (2008) suggest that high expectations following the political transformation could have led to unrealistically high perception of wage possibilities in South Africa.[4] Banerjee et al. (2008) also name a few features of the South African institutional context that could bring reservation wages to levels firms are unwilling to pay. First, although unemployment benefits are only given to a small proportion of the unemployed, state pensions and other government transfers are relatively generous in South Africa (Case and Deaton, 1998). Much attention in the literature has been given to the effect of the old age pension (OAP) on labour market

[3] In South Africa, many unemployed live in locations far away from potential workplaces. This is one of the legacies of Apartheid-era forced removals (Banerjee et al., 2008, 734).

[4] In addition, the introduction of minimum wages in South Africa could have changed the perceptions of fairness. Falk et al. (2006) show in laboratory experiments that minimum wages can cause long-lasting changes in subjects' reservation wages even when they are no longer in place.

outcomes because of its unique combination of high coverage and value.[5] Evaluating the literature on OAP, Leibbrandt et al. (2013) conclude that the pensions can have both a positive and a negative labour market effect. On the one hand, Abel (2013) argues that the income effect of having an OAP recipient in the household leads to an increase in the reservation wage and to a lower probability of employment. This is in line with studies like Bertrand et al (2003), Ranchhod (2006) and Sienaert (2008) that find a decrease in labour market participation among old age pensioners as well as among working age individuals in pension-receiving households. The effect seem to be larger when the eligible pensioner is female, as females seem to share their income more with the rest of the household (Leibbrandt et al. 2013). On the other hand, Posel et al. (2006) find that African women are more likely to become migrant workers after the arrival of a pensioner, since the extra income enables them to go for job search. Similarly Ardington et al. (2016) find that young men from KwaZulu-Natal are more likely to become labour migrants at least if educated enough. However, studies by Klasen and Woolard (2009) and Ardington et al. (2009) show how unemployed family members seem to move to households with pensioners and that this can prolong their unemployment duration.

2.2.2. Measurement error

Reservation wages are difficult to measure in practice. A number of studies acknowledge that self-reports from surveys may be biased for a number of reasons (Hofler and Murphy 1994; Böheim, 2002) and may not reflect what labour economists are interested in measuring (Petterson, 1997). The first potential source of bias is the way the reservation wage question is phrased. Jones (1988) raises the concerns that the question might be posed in a way that it is not meaningful to respondents. Hofler and Murphy (1994, p.962) argue that respondents often "engage in wishful thinking" and state amounts that are higher than their true limit "if confronted with a realistic opportunity". Petterson (1997, p.607) believes that jobless workers respond to such a question in a way that "reflect[s] what they believe they should earn, not necessarily what they can earn". He concludes that reservation wages reflect aspirations and perceptions of self-worth and not the wages at which the respondents are indifferent between working or not. In addition, the perception of fairness might be a function of the received wage, i.e. low-wage earners are less likely to consider their payment as fair than higher-wage earners. This is in line with findings from other studies showing that job-seekers at the lowest end of the wage distribution report reservation wages that are above their previous wage, while those at the top state reservation wages below their previous wages (Holzer, 1986a, 1986b; Jones, 1988; Jensen and Westergård-Nielsen, 1987; Petterson, 1997).

A second source of potential bias is that job characteristics might influence the willingness to accept a job offer (Hofler and Murphy, 1994). Whereas British surveys only ask a single question of the form: "What is the lowest amount in take-home pay you would be prepared to accept for a new job?", American studies

[5] The OAP paid out 1200 Rand in 2012, which is about 175% of the national median per-capita income, more than 90% of black South Africans were eligible to receive the grant once they turned 60 years, and 86% of those eligible take it up (Abel, 2013). Given that only 7 percent of pensioners live without at least one 19-50 year old household member (Sienaert, 2008), the literature has found significant intra-household resource transfers (Ardington et al., 2009).

ask a two-tiered question: (1) "What type of work have you been looking for?" and (2) "What would the wage or salary have to be for you to be willing to take employment in this type of work?". Asking reservation wage questions in the latter fashion might partially reduce the job characteristics bias but could also induce a different bias if the lowest acceptable wage is below acceptable wages in the desired industry.

Third, respondents are likely to give different amounts depending on whether the question is framed as open-ended or closed. Holzer (1986b) compares the answers from the NBER Survey of Inner-City Black Youth with the youth cohort of the National Longitudinal Surveys (NLS). While the NBER survey asks youth whether or not they would accept specific dollar amounts, the NLS asks about the job "sought" and to designate their own dollar figure. Comparing the responses with recent job salaries and other job questions, Holzer finds that the answers in the NBER survey appear to be more consistent than the one from the NLS and that black respondents (especially from the South) exhibit larger inconsistencies. One possible explanation is that open-ended reservation wage questions allow individuals to confuse wage expectations with the reservation wage. Holzer (1986b, p.43) concludes that blacks are either more likely to misinterpret the question than whites or have a "greater degree of expectational error on their part".

Finally, an additional source of inaccuracy in reservation wage data is non-response bias. This might be a serious concern if individuals who do not respond are a non-random sub-sample (Hofler and Murphy, 1994). The data collecting process can also bias the findings on reservation wages. Sestito and Viviano (2011) observe that self-reported reservation wages are relatively higher in the (poorer) South than in the (richer) North of Italy. Their explanation is that the wage distribution of the unemployed sample in their data is affected by a double-selection process.[6] Another interpretation would be that in structurally poorer regions individuals have weaker labour market attachments and therefore provide less reliable responses to open questions about their reservation wage as argued before.

With respect to the direction of measurement error in reservation wages, most studies (e.g. Moylan et al., 1984; Addison et al., 2004; and Brown and Taylor, 2011) find that reported reservation wages are generally higher than accepted wage offers.[7] Krueger and Mueller (2014) find that close to half of their sample accepts wage offers below their stated reservation wage and suggest that this may be due to errors in self-reports. The authors also examine the factors that predict accepting a wage offer below the stated reservation wage and find that long-term unemployed (more than 79 weeks), risk-averse and graduate degree holders are more likely to accept job offers lower than their stated reservation wage. That is, misreporting of the reservation wage is not random.

[6] That is, only people who are looking for work are asked to report their reservation wage. However, people only look for jobs if the utility of searching is greater than the alternatives. If in the North of Italy more people withdraw from the labour market and the question is not asked, reservation wages in the South will appear to be relatively higher.

[7] Conversely, Hofler and Murphy (1994) argue that respondents might understate income questions if the survey is run by a government agency and they think that it could affect their tax payments or credit line. In particular, unemployed job seekers receiving benefits might state low reservation wages to demonstrate their willingness to work.

Concerning South African data, Kingdon and Knight (2001) criticise the survey questions in the 1993 PSLSD survey as open to interpretation and worry that respondents reported wages they find 'fair' rather than their 'true' reservation wage. Furthermore, they point out that respondents living in remote areas with low education, or lack of previous work experience, may have no information about their true 'market worth'. Nattrass (2002) explores inconsistencies in the way people reported reservation wages in another South African dataset—the 2000/2001 Khayelitsha Mitchell's Plain (KMP) survey. In her study, respondents did not give the same answer to the following two questions: (i) "what is the absolute lowest monthly take-home wage that you would accept for any work?" and (ii) "what is the absolute minimum take-home monthly wage below which you would not be prepared to work in any job (taking into account your desired hours of work)?". Nattrass speculates that people may respond differently to slightly different phrasing of reservation wage questions or that they may not have a definite reservation wage. Furthermore, 60% of respondents that stated they would work "in a public works programme nearby (cutting down trees on the sand dunes) for R33 a day", stated a higher reservation wage when answering (i).

Presuming that reservation wages are somehow measured imprecisely, as it appears to be the case both internationally and in South Africa, the question arises of how this will influence the estimated coefficients from standard empirical labour models. Although it is well known that measurement error can bias the parametric estimators of interest, most applied studies assume a classical measurement error–i.e. the error is additive and uncorrelated with the other variables in the model. However, this assumption "reflect(s) convenience rather than conviction" (Bound et al., 1994) and validation studies (e.g. Bound and Krueger, 1991; Bound et al., 1994; Bollinger, 1998; or Bricker and Engelhardt, 2008) show that a variety of labour market outcomes are affected by non-classical measurement error (see also Bingley and Martinello, 2014). As these studies show, the consequences of measurement error on the parameters of interest will depend on whether the error is in the dependent or independent variables and will vary significantly across model specifications.

2.3. Data and descriptive statistics

This paper uses data from the Cape Area Panel Study (CAPS). The CAPS is a longitudinal survey of a representative sample of youth in the Cape Town metropolitan area. The first wave was conducted in 2002, interviewing 4,752 young people between the ages of 14 and 22. Of those 2,140 participants where male. As explained below, our empirical analysis will be restricted to a sample of young males. The sample clusters were taken from the 1996 Census enumeration areas with the aim to achieve equal sub-samples of black and coloured youths.[8] For the empirical analysis, individual weights were used to adjust for over-sampling of the black youth as well as for individual non-responses so as to provide a representative sample of youth

[8] The racial profile in Cape Town is significantly different than in the rest of the country. In the 2001 Census it was 32% black, 48% coloured, 1.5% Indian, and 19% white, while the total South African population was 77% black, 9% coloured, 2.6% Indian and 11% white (Lam et al., 2013).

from the metropolitan Cape Town area. The original sample was followed over five waves during the period 2002 to 2009. Wave 2 of the survey took place in two distinct phases in 2003 and 2004 (Waves 2a and 2b, see Lam et al. 2008). Wave 3 and 4 were conducted in 2005 and 2006, respectively, and the entire young adult sample was re-interviewed. The last round of CAPS (Wave 5) attempted to re-interview the sample and their households for a fifth time in 2009.[9]

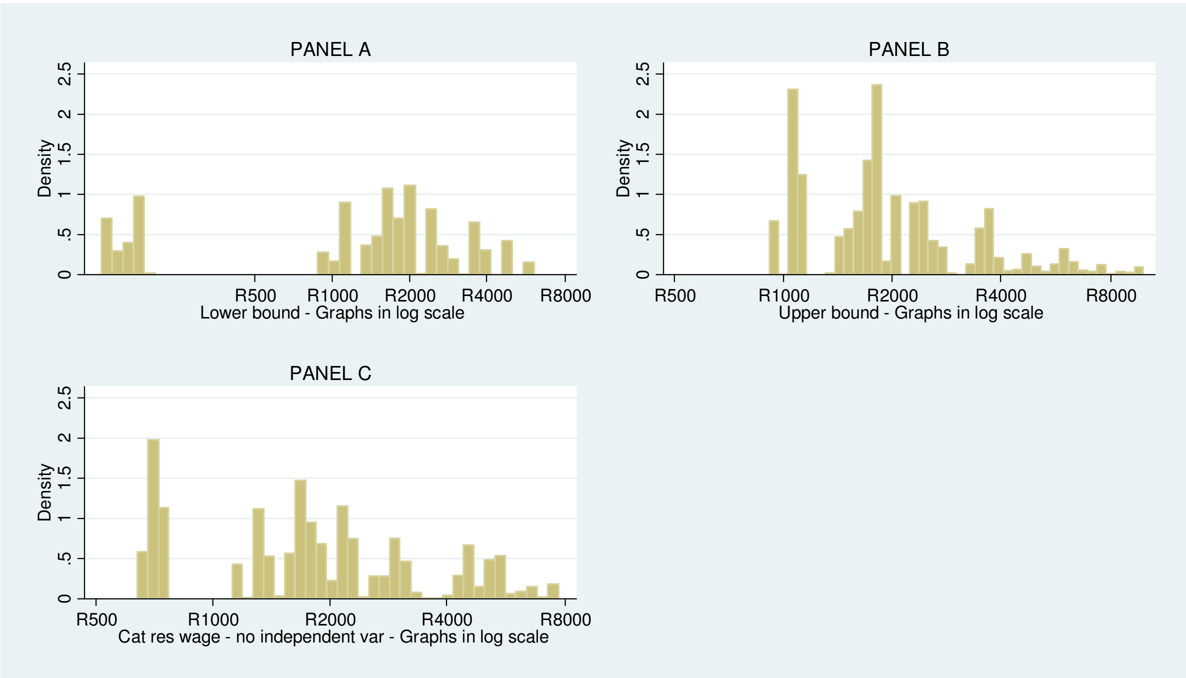
Young men and women in South Africa face very different labour supply decisions due to gendered social norms and fertility decisions, so our empirical analysis of reservation wages is simplified by restricting our sample to young males.[10] The number of male observations and sample characteristics in each wave can be found in Table A1 in the Appendix. Since the set of job offer questions was not asked in Wave 1, most of our analysis will be based on the panel sample from Waves 2 to 5. The total attrition rate between these waves (2003-2009) is 26.3%.[11] Despite non-negligible attrition, we have at least two consecutive panel observations for 85.7% of the sample (see Table A2 in the Appendix). Having longitudinal information for most of our sample allows us to observe transitions into the labour market and variation in reservation wages over time. As shown in Table A1, 25% of the male sampled youth where employed and 58% in school in Wave 2. By the fifth wave, 63% of the sample was working and only 9% was still studying.

As mentioned above, the precise wording of the survey question is important for self-reported reservation wages (Lancaster and Chesher, 1983; Bertrand and Mullainathan, 2001). For CAPS, there are two different ways to infer the lowest wage individuals would work for. From wave 2 onwards both type of question have been asked to all young adults independent on employment status. The first (and traditional) question was: "*What is the absolute lowest take-home wage that you would accept for any permanent, full-time work*?" The second is a series of questions about whether or not respondents would accept a sequence of increasing hypothetical wage offers,[12] e.g. "*Would you accept a job as general worker for a monthly wage of R1438?*", "*Would you accept a job as machine operator for a monthly wage of R1619?*" The set of hypothetical job offers have always been asked after the traditional reservation wage question. Individuals who respond that they would accept a job offer of R1619 but would decline a job offer of R1438 can therefore interpreted to have a reservation wage in the [R1438, R1619] interval (which will be referred to as upper and lower bound). Respondents who would decline all hypothetical job offers (this share ranges between 5% of the sample in wave 4 and 35% in wave 5) have reservation wages that exceed the highest wage offer. When responding to these wage offers, individuals are required to carefully probe their preferences in a way that is not required when providing an answer the traditional reservation wage question. Therefore, in the following this measure will be referred as the "probed" reservation wage measure.

[9] For a more detailed description of the panel study, including information on study design, enumeration areas and sample selection in each wave, see Lam et al. (2013).
[10] In earlier versions of this paper the empirical analysis was performed on both men and women, and the results were largely similar to those obtained using the male sample only.
[11] The overall attrition rate between Waves 1 and 5 (2002-2009) is about 36%.
[12] Across the different waves of CAPS, there have been up to 7 different job options included as questions of the same type, as well as various wage steps attached to these offers (see Appendix Table A3).

In Figure 1, panels A and B display the distribution of the lower and upper bounds of reservation wages as just above.[13] The panels show that a significant number of people would work in the lowest job category (as domestic workers), hence the cluster of observations at the left tail of the distribution in both panels. Panel C, on the other hand, plots the distribution of a reservation wage from the categorical information. Since the exact value is unobservable to us, an interval regression is used to obtain a point estimate (see Table A5 in the Appendix. Typically, an interval regression fits a model of a dependent variable, y_i, where the information on y for each observation is either interval, left-censored, or right-censored data. Panel C shows the probed reservation wages from an interval regression without any explanatory variables. That is, the value for each individual is estimated from an assumed log-normal distribution of reservation wages between upper and lower bound values.[14] For every respondent, the predicted value lies between the last declined and first acceptable hypothetical job offer. For those with no last declined information because they accept the first option the predicted value is at the left of the first acceptable job offer.[15]

Figure 1: Lower/upper bounds and derived reservation wage



(Source: CAPS wave 2-wave5)

[13] All monetary values in the analysis of this chapter have been deflated to reflect 2008 prices. Deflator used from: http://www.statssa.gov.za/publications/P0141/CPIHistory.pdf

[14] Interval regressions are equivalent to an ordered probit model but with fixed cut points (Wooldrige, 2002). The interval regression is preferred (and more efficient) because it not only uses the ordering of the intervals, but also the values at the intervals. It has the added benefit of giving us predicted Y_i values that have economic meaning. Comparing our interval regression model with an ordered probit model shows that the log likelihood as well the z-statistics of the models are similar. Interval regressions have also shown to be more reliable than midpoint estimates in case the income brackets are relatively wide (von Fintel, 2007).

[15] If we ran an interval regression with independent variables (education, age, race and labour market status) predicted values would be less concentrated to single points, as the information from the independent variables allows 'smoother' predictions.

Undoubtedly, having to derive a point estimate of reservation wages from interval information introduces noise in our measure. In the empirical section below, we will show how this needs not be a big concern for (at least some) of the empirical models estimated in the paper. A potentially more serious concern for our empirical analysis stems from the fact that the survey questions attach specific job titles to the various wage offer amounts. That is, we do not know if respondents would consider the same offers as acceptable/not acceptable if associated to other types of jobs. It is possible that individual reservation wages vary across jobs if the nonwage characteristics differ and if these characteristics enter the individual's utility function (Holzer, 1986b). In extreme cases where a higher wage offer is associated with an occupation that some individuals perceive as having very unattractive non-wage attributes, this may lead them to accept wage offers that are lower than other rejected offers. For example, in wave 2 and 3, 3.9% of the sample would work as a domestic worker for R864 but would not work as a security guard for R1300; and 6.8% of males would work as a machine operator for R1619 but not as a cashier for R2000. Restricting our sample to males helps remove the effect of gendered occupation-specific non-wage attributes, but 13% of the sample still provided responses that violated the transitivity property. It is not clear whether these cases reflect a strong preference reversal for the non-wage attributes of adjacent job offers, or response or coding errors. In these cases, the upper bound reservation wage is set at the lower amount. Although this is an important caveat for our analysis, we argue that attaching specific job titles to wage offers should bias upwards our reservation wage derived from these responses. That is, we expect that on average individuals would accept lower wage offers if presented with job descriptions more in line with their preferences/skills (Holzer, 1986b).

2.4. Internal consistency

We start our empirical analysis of the relative reliability of the two reservation wage measures by investigating the internal consistency of these responses. In particular, we document how these reservation wages relate to each other, and to a series of other wage measures (predicted, typical, and subsequently earned wages). We also take advantage of variation in the ordering of questions across survey waves to investigate the relative sensitivity of the reservation wage measures to priming effects.

It is perhaps not surprising to note that in this sample of young South Africans individuals tend to provide much higher values to a question about what they perceive as a "typical wage" for someone like them[16] than the predicted wage for someone who shares their observable productivity characteristics.[17] This difference is about 40% on average for the whole sample and about 64% for those individuals without any work experience. Even though those without any work experience have 30% lower predicted wages than the rest

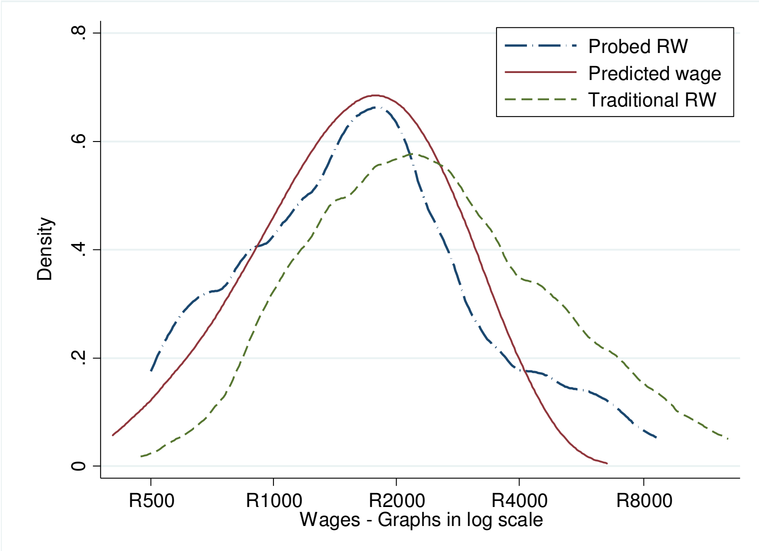
[16] The survey question asks: "What is the typical take-home monthly wage for other people like you (same age, education, and skills) who have full-time jobs?".

[17] As in Rankin and Roberts (2011), the predicted wage is obtained using a Heckman selection estimator with education, experience, race, wave and neighbourhood dummies as the productivity attributes and being household head, number of people working in the household and household size as the determinants of employment that are excluded from the wage equation.

of the sample, their reported typical wages are the same. With prevalent weak attachment to the labour market, many young job-seekers appear to be over-optimistic about the wage offer distribution, possibly as a result of limited market feedback on which they can downwardly adjust their expectations.[18]

Figure 2 overlaps the kernel density estimates for the traditional and probed reservation wage responses, as well as predicted wages obtained from a Heckman selection model. The curves reveal that all three measures appear to be roughly log-normally distributed, and that the traditional reservation wage measure is substantially higher on average than the probed reservation wage or predicted wages.

Figure 2: Kernel density estimates of traditional and probed reservation wages, and predicted wages



(Source: CAPS wave 2-5)

If responses to the traditional reservation wage question are partly anchored to perceptions about fair or desired wages that are only partially adjusted towards the true lowest acceptable wage offer, then we would expect the traditional measure to be an upwardly biased estimate of true reservation wages. This would be particularly true for those with weak labour market attachment since they are less informed about their earning possibilities and would state a traditional reservation wage witch is more correlated to their perception of fair wages. By forcing individuals to consider whether or not they would accept specific wage offers, we would expect the probed measure to be less biased and hence lower than the traditional measure for several individuals. If this hypothesis is correct, then responses to the hypothetical wage offers should frequently contradict the responses to the traditional reservation wage question. Indeed, in our sample, these kinds of contradictions occurred in about 60% of cases. In other words, roughly 60% of the sample reported lowest acceptable wage offers that were higher than one of the hypothetical wage offers that they subsequently stated they would accept. This share is even higher for those with weak labour market

[18] Holzer (1986b) documents that unemployed black youth in the United States have higher wage expectations relative to the labor demand they face.

attachment: 67% for those who have not worked before and 66% for those who reside in high unemployment districts. These apparent contradictions are entirely consistent with our hypothesis.

How would we expect the two reservation wage measures to compare to predicted wages, typical wages and accepted wages? If the traditional measure is partly anchored to over-optimistic wage expectations that are also reflected in responses to typical wages, then we would expect rw_1 to be close to responses about the perceived typical wage, whereas rw_2 should be much lower than that on average. Columns 1 and 2 in Table 1 confirm that this is indeed the case: the traditional measure is almost identical to individual response about typical wages on average, while the probed measure is 61% below this value.

Predicted wages are reflective of what firms are actually paying workers, and since these wages have been accepted by workers, we would expect true reservation wages to be lower than predicted wages on average. Columns 3 and 4 in Table 1 demonstrates that the traditional measure is 22% higher than predicted wages, whereas the probed measure is 33% lower.

Columns 5 and 6 of Table 1 compare the two reservation wage measures for the unemployed sample with accepted wages in the subsequent survey period. Since job-seekers accept work when receiving a wage offer higher or equal to their reservation wage, we would expect most accepted wages to be above previously reported reservation wages. This is consistent with what happens for the probed measure, which is on average 21% below the accepted wage, whereas the traditional measure is 23% higher on average than subsequently accepted wages.[19]

Table 1: Mean/median differences between predicted, accepted, typical wages and reservation wages

	RW – Typical wage		**RW – Predicted wage**		**RW_{t-1} – Accepted wage**	
	(1)	(2)	(3)	(4)	(5)	(6)
	rw_1	rw_2	rw_1	rw_2	rw_1	rw_2
Obs.	1268	1268	1179	1179	555	555
Mean	.002	-.618	.216	-.334	.227	-.208
T-test (t-value)	(0.16)	(-23.52)	(11.99)	(-22.62)	(6.35)	(-7.48)
Median	-.052	-.618	.155	-.330	.090	-.311

Note: Columns (1)-(2) and (3)-(4) show the numbers for the unemployed sample. Columns (5)-(6) those unemployed that found a job next period. (Source: CAPS wave 1-5)

Furthermore, about 52% of the unemployed sample accepts a job offer later in the survey that has lower monthly wages than their self-reported reservation wage in the previous period. In comparison, 30% of reservation wages derived from the explicit job offer questions were larger than the accepted wages. This is

[19] We also observe that 55% of accepted wage offers were lower than the traditional reservation wage in the previous survey, whereas this was a much lower share of 20% for the probed reservation wage. The upward bias in the traditional reservation wage measure may explain why Krueger and Mueller (2015) find that 44 percent of their respondents accepted a lower wage than their (traditional) reservation wage in the previous survey period. This suggests that our findings may be relevant beyond the South African labor market.

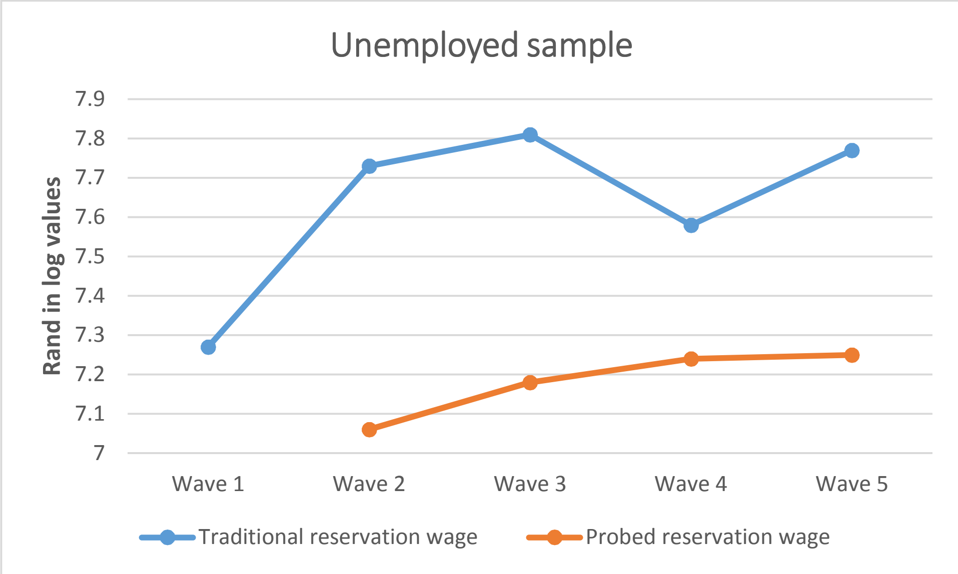
a substantial reduction in the degree of 'over-expectation' in our sample (even more so considering that we believe our derived measure to be upwardly biased).

If responses to the traditional reservation wage question are partly anchored to other memorable wages, then we would expect this measure to be particularly sensitive to seemingly irrelevant priming effects like the questions that precede the reservation wage question in the survey. On the other hand, the probed measure should be more stable across surveys. The different survey structures provide a pseudo-experiment to test this hypothesis. In waves 1, 4 and 5 the traditional reservation wage question followed after questions about the current job search strategy or probability of finding work, whereas in waves 2 and 3 it was preceded by a question that asked about the individual's labour market prospects at the age of thirty. Being asked to imagine oneself in the future is likely to trigger thoughts of wage aspirations and wishful thinking, which – if our hypothesis is correct – should make these wage aspirations more prominent as an anchor for responses to the traditional reservation wage question. The hypothetical wage offer questions were also moved around in the survey. In wave 1 these questions were not asked, whereas in wave 2 they followed a series of questions on government assistance. In waves 3 to 5 these questions followed directly after the traditional reservation wage question. Our hypothesis is that being asked to accept or reject specific wage offers requires individuals to probe their preferences, which would imply that such responses ought to be much less vulnerable to irrelevant priming effects. If true, these responses will be unaffected by the preceding question and relatively stable across survey waves.

The observed responses to both measures across waves are reported in Figure 3. The trend in responses to the traditional reservation wage question over the survey waves shows a visible upward jump for waves 2 and 3, when the traditional question was preceded by question on future wage expectations.[20] On the other hand, the probed reservation wage measure is more stable over time, showing only a moderate concave trajectory that is consistent with the Mincerian life-cycle wage path.

[20] This effect is larger for those with weak labour market attachment (as measured by having ever worked before, or residing in a high unemployment district) than for the sample as a whole.

Figure 3: Mean traditional and probed reservation wage values over the 5 wave period.



(Source: CAPS wave 1-5)

The patterns in Figure 3 provide further evidence that for many job-seekers responses to the traditional reservation wages question might not correspond to the lowest acceptable wage offer. Overall, part 3 and 4 have given descriptive evidence for our hypothesis that the answers to the traditional reservation wage measure may provide over-inflated and not the true reservation wage, while asking individuals about hypothetical wage offers may elicit more reliable responses.

2.5. Regression analysis

To evaluate whether or not the probed measure is truly more reliable than the traditional measure, two regression models are applied. First, we will test which measure is more in line with standard theoretical search models by regressing some standard observed variables on the reservation wages and see which coefficients are more in line with the theoretical predictions.

Secondly, we will observe the correlation between employment status, which should be influenced by true reservation wages, and our two observed measures. If the traditional measure is more influenced by measurement error than the probed measure we would expect a stronger negative correlation between employment and the probed variable.

2.5.1. Measurement error bias and econometric model

Assuming that reservation wages are measured somehow imprecisely, the question arises how this will influence the results of our model. It is well known that measurement error in the data can bias the parametric estimators of interest. In applied studies, the underlying assumption normally is that measurement error is classical, i.e., the error is additive, not correlated with the dependent, independent and all other variables in the model. However, this assumption "reflect convenience rather than conviction" (Bound et al. 1994) and all seminal validation studies e.g. Bound and Krueger (1991), Bound et al. (1994), Bollinger (1998) or Bricker and Engelhardt (2008) show that labour market-related outcomes are affected by non-classical measurement error (Bingley and Martinello 2014).

The consequences of measurement error on the parameters of interest β depends on the type of model, whether the error is in the dependent or independent variable and whether the error is classical or non-classical in nature (Hausman 2001). Given that the error is additive and using ordinary least squares estimators the following model should be estimated:

$$Y^{*} = \beta_0 + \beta_1 X^{*} + \varepsilon$$

However, we do not observe the true Y^{*}or the true X^{*} but rather:

$$Y_i = Y^{*} + v$$

$$X_i = X^{*} + u$$

In case the measurement error is classical and Y_i is observed instead of Y^{*} the β_1 coefficient is still consistent and only the standard error σ_y is larger (Bound et al. 2001). If X^{*}instead of X_i is measured the OLS estimator will give an inconsistent β_1. By textbook the bias will be:

$$\text{plim}\ \hat{\beta} = \beta \frac{\sigma_x^2}{\sigma_x^2+\sigma_u^2} = \lambda\beta$$

λ is normally referred to as to as the reliability ratio and since 0< λ <1 the coefficient $\hat{\beta}$ is biased towards zero and known as the attenuation bias. If there is only one independent variable the proportional bias is just equal to σ_u^2/σ_x^2. Second, Bound et al. (1994) show that even if the error is non-classical, i.e. u is correlated with the true X^{*} the proportional downward bias is equal to the regression coefficient for X_i from a regression of u on the set of measured X_i.

We are interested in correctly estimating the determinants of reservation wages. However, there are good reasons to believe that the assumption $cov(X_i^{*}, v_i) = 0$ is not valid, e.g. people who have never worked before or are less attached to the labour market might give less accurate answers when asked what the lowest amount is that they would work for. This is in line with the findings of Krueger and Mueller (2014) that find less accurate reservation wages for the long term unemployed. This group is susceptible to stating different reservation wages depending on whether they are asked directly or if they are questioned using a set of verbal multiple choice job options. It follows that the measurement error is non-classical and the coefficients in the OLS model will be biased. In case of this non-classical measurement error the error term is mean-reverting: individuals with particularly high reservation wages are more likely to under-report and those with very low reservation wages are more inclined to over-report. This could arise if, for example, individuals with a weaker labour market attachment are simultaneously more desperate about the wages they would accept and more likely to over-report their reservation wage. In this case the measurement error is mean-reverting, and should induce a negative correlation between the measurement error and the true reservation wage, $Cov(v_1, rw^*) < 0$. In this case the measurement error will displace some of the informative variation in the reservation wage measure, which will attenuate the regression coefficients towards zero. This attenuation bias will be larger, the higher the variance of the measurement error term.

The categorical nature of the information obtained from the hypothetical reservation wage also poses certain econometric complications. However, when the reservation wage is the dependent variable, the coefficient estimates from an interval regression[21] are directly comparable to those obtained from an OLS regression on reservation wage point data.

Our hypothesis implies two predictions about the relationship between the determinants of reservation wages and the two observed measures. First, measurement error may have the effect of attenuating the coefficients on the explanatory variables and this attenuation should be stronger for the traditional than for the probed measure. Secondly, we would expect proxies of weak labour market attachment (e.g. whether the individual has worked before or whether the individual resides in a high unemployment region) to be positively associated with the traditional measure, but not with the probed measure. We test these implications in Table 2 below.

2.5.2. Determinants of reservation wages

Standard job-search theory suggests that reservation wages should be higher for those with valuable assets, high non-wage income, who reside in high transportation cost regions, who possess productive attributes that are associated with higher predicted wages, who have been unemployed for a relatively short duration, and who reside in low unemployment regions. International studies have confirmed that these predictions can accurately describe the self-reported reservation wages of job-seekers in countries where these individuals have a stronger attachment to the labour market (e.g. Bloemen and Stancanelli, 2001; Addison et al., 2009; Krueger and Mueller, 2014).

We posit the following basic regression model to estimate the determinants of log of reservation wages:

$$rw_1 = \alpha_1 + \alpha_2 X_{it} + \alpha_3 C_{it} + \varepsilon_{it} \qquad (2)$$

where rw_1 is equal to the log of monthly reservation wages observed by the traditional measure. X_{it} is a vector consisting of the variables of interest: an asset index generated by multiple correspondence analysis (MCA)[22]; log per capita household work income and log per capita household grant income; log of the

[21] Interval regressions are ordered probit estimators in which the cut-offs are specified rather than estimated (Wooldrige, 2002). If the thresholds at which the latent variable produces different discrete values are economically meaningful, then the coefficient vector inherits this property. Furthermore, its magnitudes are directly comparable to those obtained from an estimator that uses continuous point data as the dependent variable.

[22] Although it is more common to use a related technique, the Principal Component Analysis, it has been shown that MCA is preferable for variables that are not continuous or normally distributed (Booysen et al., 2008). The index is based on variables relating to various aspects of the household's living conditions including access to water and sanitation, as well as on household assets like having a car or TV.

cluster transport costs[23]; log of predicted wages, log of unemployment spell in weeks[24], the local unemployment rate and a variable indicating if the young work seeker has ever worked before. C_{it} is a vector for the set of control variables consisting of some education dummies to control for ability and schooling, as well as wave and location dummies. For the probed measure a similar model is estimated, using the same set of variables of interest X_{it} and control variables C_{it} in an interval regressions.

In Table 2, the coefficients from regressions of both reservation wage measures from model (2) are reported. The dependent variable in column (1) is the traditional measure, while in column (2) we report the results from the interval regression. We use panel weights in all specifications to partially correct for non-random attrition and report robust standard errors. Overall, the estimated coefficients in Table 2 confirm that the predictors of true reservation wages affect both measures in the correct direction, but have a weaker partial correlation with the traditional than with the probed measure. This is most notably true for the effect of household assets, household income and transportation costs, all of which have greater magnitude and significance in the specification using the probed measure as the dependent variable. Our findings are in line with recent papers (e.g. Sienaert 2008, Abel 2013) showing a significant positive effect of non-grant and grant per capita household income on reservation wages. It also indicates the potential problem of high transport costs, increasing reservation wages for those living in places far removed from the business centres.

Notably, there seems to be a significant negative correlation between reservation wages and the unemployment spell for both measures. That means an increase of the unemployment spell of one month is associated with a reduction in the reservation wage of 0.4%. The finding of a small decrease of reservation wages over the unemployment spell, is in line with studies like Addison et al. (2009) or Krueger and Mueller (2014). However, while Krueger and Mueller (2014) find that the decrease in reservation wages is mainly driven by older individuals, our results are negative and significant for a sample of youth respondents. We interpret this finding as being consistent with a process of learning about the wage offer distribution by the unemployed youth in South Africa. Ideally we would like to observe the impact of savings or unemployment insurance benefits on reservation wages but this information is not captured in CAPS. In addition, only a minority of the youth has access to the unemployment insurance as only individuals with continuous formal employment become eligible (Abel, 2013).

When an individual resides in a high unemployment district, this ought to decrease the true reservation wage via the decreased wage offer arrival rate (as is the case for the probed measure), but it also pushes up the

[23] Theoretically transport costs should have a significant effect on reservation wages in South Africa, where there is substantial spatial separation between business centres and residential areas as well as a lack of affordable public transportation (Banerjee et al., 2008). Unfortunately, information about transport costs was not available for a high number of respondents. For this reason, we use mean transport costs for each sample cluster instead, as a proxy for potential transport costs of the unemployed in that cluster.

[24] In each wave unemployed are youth looking for work in the last 30 days or those not looking for a job but wanting to work. Unemployment duration can be derived for all unemployed youth since a set of questions in wave 1 enquires about unemployment spells and employment status in each panel month.

wage anchor because the individual will have had fewer opportunities to downwardly adjust their unrealistic wage expectation. The district unemployment rate coefficient in the traditional measure regression shows that these two effects cancel out to leave the reservation wage unaffected by the local unemployment rate. Individuals who have worked before ought to have more realistic and hence lower wage expectations. This can be seen to have a strong negative effect on the traditional measure – which we have argued is often anchored to wage expectations – but not on the probed measure.[25]

Table 2: Determinants of reservation wages

	(1)	(2)
VARIABLES	**OLS** $\log(rw_1)$	**Interval regression** $\log(rw_2)$
Asset index	0.000	0.106***
	(0.026)	(0.028)
Log perc HH work income	0.037	0.045*
	(0.027)	(0.025)
Log perc HH grant income	0.026	0.027
	(0.029)	(0.027)
Ln (local transportation costs)	0.028	0.072*
	(0.062)	(0.041)
Ln predicted wage	0.301***	0.274***
	(0.084)	(0.081)
Log unemployment duration	-0.037***	-0.041***
	(0.014)	(0.012)
Local unemployment rate	0.000	-0.492**
	(0.253)	(0.240)
Ever worked	-0.140***	-0.054
	(0.053)	(0.048)
Constant	5.336***	4.817***
	(0.668)	(0.684)
Observations	991	992
R-squared	0.154	

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Not listed: wave, district and schooling dummies. Sample: Unemployed male with info on both reservation wage measures. (Source: CAPS wave 2-5.)

2.5.3. Employment effects

A common empirical question in labour economics is whether or not high reservation wages impede employability in the labour market. Previous studies on South African data – typically based on the self-reported measure of reservation wages – have found little evidence that high reservation wages are a disincentive to accept low wage offers. It is also well understood that measurement error in the dependent variable can attenuate the coefficients in a linear regression framework (Bound et al., 1994). Non-classical properties in the measurement error, like mean reversion, will typically diminish but not eliminate this attenuation bias. Our hypothesis would therefore imply that both reservation wage measures should have a weaker effect on individual behaviour than is the case for the true reservation wage, but that this attenuation

[25] To control for the robustness of the results, we recreated the analysis of Table 2 after dropping all inconsistent answers to the probed reservation wage. While there were a few small changes, the signs of all coefficient stayed the same and similar significances could be observed. The same is true when using a simple OLS model instead of Interval regressions for the probed model.

bias ought to be more severe for the traditional than for the probed measure. We test this by estimating the effects of both measures on labour market behaviour we can observe for the young job-seekers in CAPS. As mentioned in part 3, to deal with the fact that the probed reservation wage is categorical we use an interval regression to estimate the conditional expectation of the respondent's reservation wage, and use this point estimate as the independent variable in the labour market regressions.

First, to observe the relationship between reservation wages and probability of employment we use fixed effects (FE) and first-differenced (FD) estimators. Allowing for individual fixed effects removes the potential bias due to unobserved time-invariant heterogeneity, such as worker ability or attitude, which may simultaneously determine the probability of finding work and the reservation wage. We also use lagged values of the reservation wage to address reverse causality concerns: finding work may cause workers to upwardly adjust their reservation wages (e.g. Krueger and Mueller, 2014), which could induce an upward bias in the reservation wage coefficient estimate. Table 3 columns (1) and (2) show the regression results for the FE and FD estimators. The coefficients for the lag of the probed measure are -0.041 and -0.061, revealing a negative correlation between the probed reservation wage measure and the probability of being employed, as predicted by economic theory. This effect is significant (0.062) and marginally significant (p-value 0.11) in the FD and FE estimators, respectively. On the other hand, the coefficient for the traditional measure is not significant in either specifications.[26] To check for the robustness of the results and deal with the potential problem of multicollinearity, the FD and FE model shown in Table 3 have also been run without entering both reservation wage measures at the same time. The results given in Table A6 (in the Appendix) show basically the same pattern as before, except that the probed reservation wage measure is now also significant negative in the fixed effect model.

Second, we can observe how accepted wage offers are correlated to the lagged reservation wage measures. Theoretically one would assume that job-seekers who are less desperate about which job offers they would accept should transition into employment less frequently, but do so for higher wages. Therefore, past reservation wages should be positively correlated with accepted wages. Table 3, column (3) shows that there is indeed a highly significant positive coefficient for the probed but not for the traditional measure.

Next, we separately consider the effect of reservation wages on transitioning into and out of employment. Column (4) contains the regression coefficients for a FE employment regression where the sample is restricted to those who were unemployed in the previous period. The coefficients reveal that the probed measure has a large and marginally significant (p-value 0.103) effect on transitioning into employment, whereas the traditional measure has a much smaller and insignificant effect on the probability of finding

26 When estimating the employment regression without fixed effects and using only the contemporary traditional reservation wages measure we find, like Nattrass and Walker (2005), that this measure is positively correlated with the probability of finding work. However, we interpret this as evidence of the importance of addressing the problems of omitted variable bias and reverse causality. It is worth noting that even in this specification, the traditional measure is only significant when included on its own, but turns insignificant once it is included along with the probed measure. It seems that even the endogenous variation in reservation wages is more reliably captured with the probed than with the traditional measure.

work. The coefficients in column (5), where the sample is restricted to those who were employed in the previous period, show that neither reservation wage measure can explain transitions out of employment. Of course, reservation wages may not affect whether or not a worker gets fired, so a more instructive question may be whether reservation wages determine voluntary movements out of employment. Therefore, the hypothesis is that people are more likely to quit their job if their reservation wage is high in comparison to the salary they are receiving. To understand whether people with higher reservation wages are more likely to quit their jobs, we use responses to a question in CAPS that asks the "reasons for losing your last job" (see Table A4 in the Appendix). In column (6) we run a regression comparing youth who quit their job to those who lost their job for other reasons (i.e. fired, contract ended, business bankrupt). It shows that on average people who quit have significantly higher probed reservation wages which again is not the case for the traditional measure.

Table 3: Effects of reservation wages

	(1) FE	(2) FD	(3) OLS	(4) FE	(5) FE	(6) FE	(7) FE
Variable	**Employed**	**ΔEmployed**	**Accepted wage**	**Transitioned into employment**	**Transitioned out of employment**	**Quit**	**Discouraged**
Lagged $\log(rw_1)$	-0.009		0.021	-0.060	-0.004	-0.032	
	(0.020)		(0.047)	(0.056)	(0.029)	(0.021)	
Lagged $\log(rw_2)$	-0.041		0.194***	-0.137	0.001	0.057**	
	(0.026)		(0.066)	(0.083)	(0.031)	(0.025)	
Lagged $\Delta\log(rw_1)$		0.002					
		(0.020)					
Lagged $\Delta\log(rw_2)$		-0.062**					
		(0.027)					
$\log(rw_1)$							-0.044*
							(0.026)
$\log(rw_2)$							-0.146***
							(0.037)
Observations	2,817	1,245	387	614	1,584	1,674	1,765
R-squared	0.208	0.157	0.102	0.192	0.435	0.078	0.075

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Also controlled for years of schooling, experience and wave dummies. (Source: CAPS wave 2-5)[27]

[27] To check for the robustness of the results, in the Appendix Table 3 has been reproduced without entering both reservation wage measures at the same time. In addition, column 4, 5 and 6 have been run using OLS as well. The results show basically the same pattern as before, except that the probed reservation wage measure is now also significant negative in the fixed effect model.

Finally, we examine the difference between the economically inactive and discouraged job-seekers. Both types of individuals reported not looking for work, but the discouraged job-seekers responded they would work if someone would offer them a job. This difference in willingness to work should be reflected in a higher reservation wage for the economically inactive than discouraged job-seekers. On the one hand, economically inactive express no desire to work at what they perceive as current wages, which implies that they must have relatively high reservation wages. On the other hand, discouraged job seekers, do wish to work at current wages, which should lead them to have lower reservation wages than the economically inactive. Or in other words a stronger desire to work should lead to lower reservation wages. Table 3 column (7) provides the regression output of the two measures on the likelihood being discouraged. It shows a larger significant negative log(rw_2) coefficient, which again suggests that the probed reservation wage measure is a more reliable indication of an individual's preferences than the traditional measure.

2.6. Conclusion

This chapter explores whether individuals provide reliable answers when asked to state their reservation wages. By taking advantage of the peculiar way the CAPS data in South Africa was collected we can compare two different approaches in capturing reservation wages. The empirical analysis suggests that the way surveys ask for reservation wages is important. Young work seekers may systematically misreport their reservation wages especially if they are less attached to the labour market. The resulting measurement error can bias the coefficients of common regression models used in empirical labour analysis.

Individuals asked to report their lowest acceptable wage appear to start their thought process by thinking of a desired wage (especially when primed by a question about aspirations or expectations) and then not downwardly adjust sufficiently. Hence, our analysis shows that respondents give higher and less reliable answers to the traditional reservation wage question. On the other hand, our probed measure seems to be more internally consistent and the regression results to be in line with labour market search models. That is, using the probed measure we find significant positive effects of transportation costs and household wealth, as well as household income on reservation wages. Having a unique panel of young job-seekers we can also show a significant negative correlation with unemployment duration. This result is in line with the international literature and it is a novel finding for South Africa. Our results lead us to conclude that there is a distinct possibility of non-classical measurement error in reservation wages - i.e. correlated with at least some of the other independent variables.

We therefore conclude that when questioning about reservation wages in survey designs, the researcher should be aware of ordering and avoid asking about life aspiration and other questions that could cause some form of anchoring. There seems to be evidence that this could trigger some cognitive process where respondents think of desired or fair wages they want to earn and won't report the lowest amount they would accept. The magnitude of this anchoring effect might be an interesting research question for future experimental studies. To obtain a meaningful reservation wage, the survey could first ask: "What is the absolute lowest monthly take home wage you would accept for full employment" giving the respondents a list of rand values to choose from, e.g.: 1000Rand, 1500Rand, 2000R, 3000R, 5000R, 8000Rand, more than 12000R.

With regards to the relationship between reservation wages and employment probabilities, we show that young workers with higher wage expectations are less likely to be employed in the next period. This is in contrast to previous studies on the South African labour market. We argue that our analysis provides more robust evidence because it explicitly considers the effect of measurement error in reservation wages and because it controls for time-invariant unobserved heterogeneity. This new evidence can help provide more reliable evidence of the causes of youth unemployment to policymakers.

More generally, this study shows that the way surveys elicit information on reservation wages is of relevance. We suggest that researchers might overestimate reservation wages when using self-reported amounts. On the other hand, we show that responses based on hypothetical job offers are more in line with expected patterns from economic theory.

Chapter 3:

Estimating income mobility -

A cross country comparison

3.1. Introduction

A better understanding of economic mobility lies at the heart of development economics and is of great concern to many researchers and policy makers. This chapter aims to measure income mobility using "micro growth regression", where income growth is regressed on initial income. Such an approach will provide an estimate of "absolute mobility" and corresponds to the idea that mobility will lower total inequality as well as providing equality of opportunity (Antman and Mckenzie, 2007a). Previous studies using this approach have found large negative coefficients of the initial income variable, which would imply strong "beta-convergence" (e.g. Fields et al., 2003a; Woolard and Klasen, 2005). Translating the convergence coefficient from Fields et al. (2003a) of -0.56 (over 5 years) for South Africa would suggest that half the income gap between any two households in the country should be eliminated within 4.3 years. Although a new affluent black elite and middle class has evolved since the mid-1990s, most of the black population continues to live in poverty with very little income growth (Adato et al., 2006). On the other hand, most whites remain close to the top end of the South African income distribution, which is difficult to reconcile with the rapid speed of convergence suggested by previous studies (e.g. Fields et al., 2003a). One explanation of this apparent paradox could be the mean-regressive effect of classical measurement error (Fields et al., 2014). In particular for developing countries where collecting reliable survey data is difficult, it is known that measurement error could cause income dynamics to be largely overestimated (Antman and Mckenzie, 2007a; Fields, 2008a). Moreover, it has been acknowledged that new approaches to deal with measurement error in micro-mobility have to be found (Fields, 2008b).

This chapter aims to apply a newly developed GMM estimator, proposed by Burger et al. (2016), to quantify the effect of measurement error in income data when estimating economic mobility. This approach has the advantage of being more efficient than the two-stage least squares (2SLS) estimator used in previous studies and it allows testing for the validity of the underlying assumptions of the framework. In addition, the method can be generalised for the case of non-classical measurement error. That is, it allows the income convergence estimates of β and the measurement error statistic α to change with the initial level of household income. This approach will be applied to four large representative national panel studies from the USA, South Africa, Chile and Tanzania. Using the Panel Study of Income Dynamics (PSID) from the USA makes it possible to compare the speed of convergence and the data-reliability statistic from the GMM estimator to the results from previous validation studies. The findings show that the naïve OLS regression overestimate the extent of income mobility by a factor of about 5, while the half-life convergence gap prediction of the GMM model are about 26.7 years and in line with the results from studies like Chetty at al. (2014).

Applying the GMM estimator to data sets from South Africa, Chile and Tanzania provides some additional important insights into the reliability of income mobility estimates using panel date from developing countries. In light of recent interest in a new growing middle class (e.g. Ravallion, 2010; Burger et al., 2015) and understanding poverty dynamics in these countries, the results of this study should also provide evidence on how measurement error might influence vulnerability estimates and poverty analysis. The findings from all three countries, strongly indicate the presence of measurement error and imply that previous studies have overestimated the degree of mobility by a factor of 4-5 for South Africa and Tanzania and 5-6 for Chile. In particular, for Chile and Tanzania between 36% and 45% of the income variance seem to be due to measurement error in the initial income variable instead of actual income changes. Finally, the nonparametric estimates clearly show that the convergence coefficient is larger for poor households in all countries and measurement error is non-classical in nature. That is, income is more reliably captured for richer than for poorer households.

The remainder of this chapter is structured as follows: Section 2 provides an overview of the literature. Section 3 gives an outline of the strategy. Section 4 briefly discusses the data followed by the presentation of the results. Section 6 offers concluding remarks.

3.2. Income mobility and measurement error

There is an extensive literature on the measurement of economic mobility.[28] This chapter is based on the micro growth literature using panel data to compare how income is changing between two points in time. We will restrict our attention to the concept of weak unconditional beta convergence. Accordingly, the log per capita household income y_t^*, is characterised as an autoregressive process of order 1, or AR(1) process:

$$y_t^* = \mu + \rho y_{t-1}^* + u_t$$

In this context, the standard measure of economic mobility is the slope coefficient from a regression of current period earnings on lagged earnings (e.g. Jarvis and Jenkins, 1998; Fields et al., 2003a; Antman and Mckenzie, 2007a; Fields et al., 2014). The simple earnings dynamics model is:

$$\Delta y_t^* = y_t^* - y_{t-1}^* = \mu + \beta y_{t-1}^* + u_t \qquad [1]$$

where $\beta \equiv \rho - 1$ represents the extent of income mobility in the economy. This model is straightforward to interpret and provides a measure of convergence. When $\beta < 0$ incomes are exhibiting conditional convergence, while when $\beta > 0$, conditional divergence takes place. Empirically, the existing literature from

[28] Jäntti and Jenkins (2014) provide a comprehensive review on this topic.

developing countries has mostly found that $\beta < 0$, which implies that poor households on average grow more rapidly than the rich and incomes converges to the conditional mean (e.g. Fields et al., 2003a; Woolard and Klasen 2005; Fields and Puerta, 2010).

While most empirical studies interpret the β simply to determine income convergence, the point estimate of β can also be understood as the speed at which this convergence occurs. In the cross-country growth literature it is common practice to estimate how fast countries converge on their steady states (e.g. Barro and Sala-i-Martin (2004, p. 58)). We follow a similar approach calculating how rapidly the income gap between two arbitrarily households in the country will disappear. Equation [1] and the assumption that income shocks are i.i.d. imply that the expected one-period change in the relative income gap between any two households (denoted A and B) can be expressed as:

$$\frac{\left(y^*_{A,t-1} - y^*_{B,t-1}\right) - E\left(y^*_{A,t} - y^*_{B,t} | y^*_{A,t-1} - y^*_{B,t-1}\right)}{y^*_{A,t-1} - y^*_{B,t-1}} = -\beta$$

Therefore, if $\beta < 0$ then $-\beta$ represents the share of any income gap that we would expect to be eliminated between periods $t - 1$ and t. To calculate the expected half-life of an income gap, i.e. the time it will take for half of any income gap to be eliminated, the simple formula $t \cong \frac{0.69}{\log(1+\beta)}$ periods can be used. For example, Fields et al. (2003a) find convergence coefficients of -0.56 (over 5 years for South Africa), -0.53 (over 4 years for Indonesia), -0.52 (over 1 year for Spain) and -0.64 (over 1 year for Venezuela). These coefficients imply that the expected half-life of the income gap between the richest and poorest households is 4.3 years (South Africa), 3.7 years (Indonesia), 1 year (Spain) and 0.7 years (Venezuela), respectively.

However, a consistent measurement of β is needed to evaluate the true amount of mobility and speed of convergence. In practice, the collection of income and consumption data in household surveys is often not precise and noisy, thus we observe $Y_{i,t} = Y^*_{i,t} + \varepsilon_{i,t}$. This means ΔY_t and Y_{t-1} are both measured with error and the OLS estimate of β is biased. Hence, Antman and McKenzie (2007a) find that the naïve OLS estimates would suggest that 33 percent of the gap in income between two randomly selected households would close within three months, however their pseudo-panel analysis finds that only 1.2 percent of this gap would be eliminated during the same time. Glewwe (2012) using an IV approach with panel data from Vietnam finds that between 15 to 42 percent of estimated economic mobility is due to measurement error. Similarly, when predicting long-term household income in an IV approach, Fields et al. (2014) find signs of rapid convergence using the inconsistent OLS estimator ($\beta = -0.558$), while the results from the 2SLS show much lower convergence ($\beta = -0.090$). The authors explain the difference by apparent short run shocks.[29] While shocks may play a role, they are unlikely to lead to a more than six-fold increase to measured income convergence as in the above estimates.

[29] Mean earnings appear to be converging stronger in the OLS regression because workers who received an initial positive (negative) earnings shock may now adjusting back to their lower (higher) permanent level of earnings.

The degree and direction of the bias will depend on the assumptions one makes about $\varepsilon_{i,t}$. In recent years, many studies have tried to quantify the magnitude and direction of such measurement error in income and earnings data. Most of these studies used validation data from different sources such as administrative data, and defined measurement error as the difference between the survey and the administrative data (e.g. Bound and Krueger 1991; Bound et al. 1994). The only validation study to use data from a developing country is a study by Akee (2011) using data for Micronesia in the Western Pacific Ocean.

To summarize the findings of the validation literature it is useful to think how the error $\varepsilon_{i,t}$ influences the OLS estimator β. Antman and McKenzie (2007a) show that under the assumption that the cross-sectional sample size N increases towards infinity, the θ_{OLS} asymptotic bias can be written as:

$$\theta_{OLS} = \left[Cov(u_{it}, Y_{i,t-1}) + Cov(\varepsilon_{it}, \varepsilon_{i,t-1}) + Cov(\varepsilon_{it}, Y^*_{i,t-1}) - \beta Cov(\varepsilon_{i,t-1}) - \beta Cov(Y^*_{i,t-1}, \varepsilon_{it-1})\right] / Var(Y_{i,t-1}). \quad [2]$$

The first term $Cov(u_{it}, Y_{i,t-1})$ is positive and nonzero in the presence of individual fixed effects in the error term u and if earnings shocks, u are autocorrelated. Findings from the U.S. validation studies show that measurement errors are auto-correlated over time and the second term $Cov(\varepsilon_{it}, \varepsilon_{i,t-1})$ is positive. However, Akee (2011) finds that there is almost no correlation between the error terms over the six year period in his study in the data.

Regarding the covariance between true earnings and the measurement error $\beta\ Cov(Y^*_{i,t-1}, \varepsilon_{it-1})$, studies using U.S. data have found a negative correlation which imply mean reverting effects (e.g. Bound and Kruger, 1991; Gottschalk and Huynh, 2010), which has also been shown by Akee (2011) for Micronesia. It suggests that individuals who have low earnings are more likely to over report and those with high earnings tend to understate their earnings in the survey.

In the special case that there are no fixed effects and the measurement error is classical in nature (ε_{it} is mean zero), the θ_{OLS} asymptotic bias is:

$$\theta_{OLS} = \beta\left[1 - \frac{Var(\varepsilon_{i,t-1})}{Var(Y_{i,t-1})}\right] \quad [3]$$

In the case where the classical measurement error is large this would lead to an attenuation bias toward zero and overstate true mobility and convergence (Fields, 2008a).

Another way of thinking of the problem is to estimate the total variation in the initial income measure that is due to variation in actual initial income, y^*_{t-1}, rather than measurement error, ε_{t-1}. Under the assumption that y_t, suffers from classical measurement error, i.e. $\varepsilon_t \equiv y_t - y^*_t \sim iid(0, \sigma^2_\varepsilon)$, rewriting equation [1] in terms of the observed but noisy income measure gives:

$$\Delta y_t = \mu + \beta y_{t-1} + u_t + \varepsilon_t - (\beta + 1)\varepsilon_{t-1} \quad [4]$$

As revealed in [3], initial income y_{t-1} is negatively correlated with the model error term via the initial period measurement error term ε_{t-1}, which will downwardly bias the OLS estimate of the convergence parameter β. In this context, Burger et al. (2015) show that the expected value of the OLS slope coefficient obtained from regressing Δy_t on y_{t-1} (which we denote as θ_1) can be expressed as:

$$E(\theta_1) = \frac{\text{Cov}(\Delta y_t, y_{t-1})}{\text{Var}(y_{t-1})} = \beta - \frac{-(\beta+1)\sigma_e^2}{\text{Var}(y_{t-1})} = (\beta + 1)\alpha - 1 \qquad [5]$$

where $\alpha \equiv \frac{\text{Var}(y_{t-1}^*)}{\text{Var}(y_{t-1})} = \frac{\text{Var}(y_{t-1}^*)}{\text{Var}(y_{t-1}^*)+\sigma_e^2}$ which represents the reliability of the observed measure of initial income y_{t-1} and is also sometimes referred to as the "reliability statistic" (Gottschalk & Huynh, 2010; Abowd & Stinson, 2013). If $\alpha = 1$ income is measured without error, while $\alpha = 0$ would indicate that the income measure contains no information about the "true" household income y^*.

The structure of the error $\varepsilon_{i,t}$ will not only influence the direction of the bias in β but also the techniques to correct for it. To address measurement error in the absence of administrative data mostly two approach have been used. A pseudo-panel using cross sectional data (e.g. Antman and McKenzie, 2007a, 2007b; Cuesta and Pizzolitto, 2011) and an IV approach using predicted long-term income, where the prediction is based on household or individual characteristics such as age, education, sector of occupation and dwelling characteristics (e.g. Fields et al., 2003a, 2003b; Lee, 2010; Glewwe, 2012; and Fields et al., 2014). However, Cruces et al. (2013), constructing pseudo-panel - from Chilean panel data, conclude that this methodology does not perform well in predicting the income mobility pattern seen in the actual panel data. In another study Newhouse (2005), estimates income dynamics in Indonesia and addresses non-random income measurement error and unobserved household heterogeneity by using several instruments, including rainfall, assets and consumption. In this chapter we will follow a different new approach developed in the paper by Burger et al. (2016). The idea is to use the information provided in a three wave panel set to simultaneously estimate the extent of income mobility and the reliability of the income measure. The GMM estimator is also more efficient than a classical 2SLS approach. The detailed assumptions and technique will be explained in part 4 below.

The problem of potential measurement error in the existing income panel data has been well recognized in the literature concerned with poverty dynamics in South Africa (see for example Agüero et al., 2007; Fields et al. 2003a, 2003b; and Woolard and Klasen, 2005). In fact, it has been argued that measurement error in developing countries could be even larger than found in existing validation studies from industrialized countries such as the US with mostly formal employment (Fields et al. 2003a; Antman and McKenzie 2007a). In the following parts, we shall estimate the extent of measurement error in South African panel data and compare this to Tanzania and Chile data sets. To validate the reliability of our approach we will first compare the convergence estimates obtained from the GMM estimator with some validation studies from the US.

3.3. Methodology

In this part the methodology used to estimate β – the speed of convergence – and α – an approximation for the reliability of observed measures of initial income y_{t-1} – should be explained. To get a benchmark for our results, we will also estimates the slope coefficient β using naïve OLS and IV estimators following the approach by Lechtenfeld and Zoch (2014). Here the second lagged income variable y_{t-2} was used to instrument for basic year income y_{t-1} in a 2SLS procedure.

As mentioned before, for the main analysis the GMM estimator suggested by Burger et al. (2016) will be applied. For this approach, more than 2 waves of panel data is needed and the following assumption have to be valid: First, the income dynamics equation [1] can be generalised as:

$$\Delta y_t^* = \mu_t + \beta_t y_{t-1}^* + u_t \qquad [6]$$

and the intercept as well as the slope of the first-order autoregressive income process are time-varying. Second, the income convergence coefficient β is constant over the period observed in the data, that is $\beta_t = \beta < 0$ and $u_t \sim iid(0, \sigma_u^2)$. This assumption should be unproblematic for small t and can be tested empirically. Third, income measurement error is classical: $\varepsilon_t \equiv y_t - y_t^* \sim nid(0, \sigma_\varepsilon^2)$. Since, validation studies from the US have shown that measurement error is non-classical in the context of earnings (e.g. Bound and Kruger, 1991; Gottschalk and Huynh, 2010), the approach allows to test for the assumption of classical as well as no measurement error. Allowing for this assumption seven regression coefficients can be used to estimate β and α which can be found in column 1 in Table 1. The L(.) represents the linear projection operator, e.g. $L(y_2|y_1)$ represents the linear projection of y_2 on y_1. Column 2 gives the predicted coefficients under the assumption of no measurement error, while column 3 provides the estimated effect of classical measurement error on θ_1-θ_7.

Table 1: Regression coefficients and population moments

Parameter		**Population mean**	
		No measurement error	**Classical measurement error**
θ_1	$L(y_2 - y_1 \mid y_1) = \theta_1 y_1$	β	$(\beta+1)\alpha - 1$
θ_2	$L(y_3 - y_2 \mid y_2) = \theta_2 y_2$	β	$(\beta+1)\alpha - 1$
θ_3	$L(y_3 - y_2 \mid y_1) = \theta_3 y_1$	$\beta(\beta+1)$	$\alpha\beta(\beta+1)$
θ_4	$L(y_3 - y_1 \mid y_1) = \theta_4 y_1$	$\beta(\beta+2)$	$\alpha(\beta+1)^2 - 1$
θ_5	$L(y_3 - y_2 \mid y_1, y_2) = \theta_5 y_1 + \theta_6 y_2$	0	$\dfrac{(\beta+1)^2(\alpha-1)\alpha}{\alpha^2(\beta+1)^2 - 1}$
θ_6	$L(y_3 - y_2 \mid y_1, y_2) = \theta_5 y_1 + \theta_6 y_2$	β	$\dfrac{1-\alpha(\beta+1)+\alpha^2\beta(\beta+1)^2}{\alpha^2(\beta+1)^2 - 1}$
θ_7	$L(y_3 - y_2 \mid y_2 - y_1) = \theta_7 (y_2 - y_1)$	$\frac{1}{2}\beta$	$-\dfrac{1-\alpha+\alpha\beta^2}{2(1-\alpha-\alpha\beta)}$

Source: Table provided by Burger et al. (2016).

The basic rationale of the coefficients are as follows:

θ_1, gives the effect of wave 1 income, y_1, on subsequent income growth between waves 1 and 2, Δy_2. While θ_2 represents the same relationship between wave 2 income, y_2, and Δy_3. These two coefficients should provide the convergence estimator β if income is measured without error but will be biased as follows under the assumption of classical error: $(\beta + 1)\alpha - 1$.

$\hat{\theta}_3$ can be obtained from regressing Δy_2 on y_1 and $\hat{\theta}_4$ from regressing $y_3 - y_1$ on y_1. In the absence of measurement error and a stationary AR(1) process the regression coefficient θ_3 should be $\beta(\beta + 1)$ and θ_4: $\beta(\beta + 2)$, that is a larger proportion of the initial income gap should be eliminated between wave 1 and 3 as between wave 2 and 3. Using the formula provided in column 3 allows to test for the presence of measurement error: it will downwardly bias the coefficients of θ_1 and θ_2, and will upwardly bias θ_3. In the presence of measurement error there should be unexpectedly little additional income convergence between waves 2 and 3, given the income mobility that is observed between waves 1 and 2.

The regression coefficients θ_5 and θ_6 are obtained from regressing y_1 and y_2 simultaneously on Δy_3. If there is no error β: $E(\hat{\theta}_5|\beta, \alpha = 1) = 0$ and $E(\hat{\theta}_6|\beta, \alpha = 1) = \beta$, that is y_1 should have no effect on Δy_3 when simultaneously controlling for y_2. However, this should be different when $\hat{\theta}_5$ and $\hat{\theta}_6$ are affected by measurement error. In this case, y_1 will be correlated with the true value of wave 2 income y_2 which will intensify the attenuation bias in the coefficient on y_2 and θ_6 will be more downwardly biased than θ_2. On the other hand, measurement error should upwardly biased regression coefficient θ_5, which will make an AR(1) process seem like an AR(2) process in which income growth depends negatively on the first lag of income and positively on second lag of income.

Finally, θ_7 is captured by regressing Δy_3 on Δy_2 and has an expected value of $\frac{1}{2}\beta$, since households that experienced faster income growth between waves 1 and 2 should grow slower in the next period. However, in the presence of measurement error this negative correlation should be larger than expected.

Having set up regression coefficients θ_1-θ_7 enables us to test whether or not income is really measured without error: α =1, measurement error is classical and if there is an AR1 income process that produce an internally consistent set of regression coefficients. For example, Burger et al. (2016) show that the regression coefficients θ_1 and θ_3 provide the parameters of interest as:

$$\hat{\beta} = \frac{\hat{\theta}_3}{\hat{\theta}_1 + 1} \text{ and } \hat{\alpha} = \frac{(\hat{\theta}_1 + 1)^2}{\hat{\theta}_3 + \hat{\theta}_1 + 1} \qquad [7]$$

In such way, calculating α will directly test he hypothesis of no measurement error. Hence, using all regression coefficients at the same time would provide a test for the hypothesis that α =1. If income is measured without error and income has the form of an AR1 process, then the predicted and estimated coefficient values should only differ due to sampling variation. In a similar way, calculating the implied values for β from column 3 in Table 1 provides another opportunity to test the hypothesis of no measurement error. Finally, if $\alpha \neq 1$ then the value of α and the formula provided in column 3 of Table 1 could be used to calculate the expected values of β implied by the regression coefficients. If the calculated and observed estimates are within a relatively narrow range for these new estimates obtained using $\alpha \neq 1$, then we can reject the hypothesis of no measurement error but can maintain the assumption of classical measurement error and an AR1 process.

While the informal approach outlined above gives a simple intuitive way to test the measurement error assumptions, the GMM system estimator developed in Burger et al. (2016) offers a more efficient approach to estimating the model parameters and testing the over-identifying restrictions. Maintaining the assumption of classical measurement error, the regression coefficients θ_1-θ_7 provide five linearly independent coefficients that depend on two unknown parameters. Using the relationship between β and α developed in Table 1 allows to construct a vector of sample moments:

$$g\big(y_{it}, \boldsymbol{\theta}(\beta,\alpha)\big) = \begin{bmatrix} (y_{2i} - y_{1i} - \theta_1 y_{1i})y_{1i} \\ (y_{3i} - y_{2i} - \theta_2 y_{1i})y_{1i} \\ (y_{3i} - y_{2i} - \theta_4 y_{2i})y_{2i} \\ (y_{3i} - y_{2i} - \theta_5 y_{1i} - \theta_6 y_{2i})y_{2i} \\ \big(y_{3i} - y_{2i} - \theta_7(y_{2i} - y_{1i})\big)(y_{2i} - y_{1i}) \end{bmatrix}$$

With the identifying assumption $E\big[g\big(y_{it}, \boldsymbol{\theta}(\beta_0, \alpha_0)\big)\big] = 0$ follows directly from the assumptions that both u_t and e_t are i.i.d. processes. The GMM estimator can then be expressed as:

$$(\hat{\beta}, \hat{\alpha}) = \arg\min_{\beta,\alpha} \left(\tfrac{1}{N}\textstyle\sum_{it} g\big(y_{it}, \boldsymbol{\theta}(\beta,\alpha)\big)\right)' \widehat{W} \left(\tfrac{1}{N}\textstyle\sum_{it} g\big(y_{it}, \boldsymbol{\theta}(\beta,\alpha)\big)\right) \qquad [8]$$

where $\widehat{W}$ is the weighting matrix. Even in the case of no measurement error, the GMM estimator should give more efficient estimates of the convergence parameter β than estimates of regression coefficients θ_1 and θ_2. Second, the GMM estimator provides a J-test statistic which allows testing of the validity of the identifying assumptions.

Finally, to deal with the concerns that measurement error is not classical in nature, as income measurement for poor households might be more imprecise than for richer households, a non-parametric approach will be applied as well. This approach will not only relax the assumption of a α being constant, but also allow the convergence parameter to vary by initial income.

In part 5, the methodology outlined above will be applied to four different panel studies. In particular, using the panel data from the USA should allow to compare the estimates of β and α with well-established validation studies and therefore test the functionality of the approach.

3.4. Data

To understand poverty dynamics while controlling for measurement error, household panel data is needed. The four panel studies used in this chapter are the National Income Dynamics Study (NIDS) – South Africa; Panel of Socioeconomic Characterization *"La Encuesta Panel Casen"* (EPC) - Chile; the Tanzanian "Natal Panel Study" (NPS) and The Panel Study of Income Dynamics (PSID) – USA. Since NIDS will not only be used to set the benchmark results for this chapter, but also be part of the main analysis for chapter three, it should be discussed in more detail than the other data sets.

3.4.1. South African Panel Data

The main rationale for using NIDS is its coverage of the entire country. After the release of the new 2014-2015 data set, NIDS now contains a four wave panel spanning a time period of seven years. NIDS is quite large, including 26,776 completed individual interviews in 2008 (wave 1), 28,519 individual observations for 2010-2011 (wave 2) 32,571 successful interviews in 2012 (wave3) and 37,396 individual observations for 2014-15 (wave 4). For now we will concentrate on households observed in wave 1, wave 2 and wave 3 (leaving out wave 4 information for most of the main analysis) to obtain the largest possible amount of observed households. As with all panel studies, there is some attrition between the different waves. Yet, in comparison to the second wave, wave 3 has negative attrition rates (see De Villiers et al. 2013). That means that out of 26,776 core household members, 22,058 have been observed again in wave two and 22,375 in wave three. Attrition among the richest decile is 41.59% and is especially common among the white population (50.31%), which is more than three times higher than attrition among black Africans (13.39%). As richer households drop out at a higher rate, an analysis with the resulting unbalanced sample would incorrectly indicate income convergence towards the mean. To address this issue, the analysis uses a balanced sample and with specific panel weights to adjust for attrition. The balanced sample of individuals that appear in all three waves consist of 18,826 individual observations. However, since we are interested in household income dynamics we will only households, for those the household head stayed the same for all three waves[30]. Limiting the sample to such households will leave us with 2,786 core households.

3.4.2. Chile Panel Data

The EPC is part of the representative national panel (Casen) containing four waves from 2006-2009. In 2006 there were 30,104 successful interviews, of those originally interviewed 23,353 (77.6%) could be found

[30] Arguably this definition of a household is very restrictive. Therefore, to control for the robustness of the results two less strict definitions of a core household were used as well. First, keeping all households where the original household head could be follwed over three periods. Households that moved location or where the household head changed in the survey would still be included. Using these criteria we end up with about 4000 household observations. Second, the sample was further increased by following all households that had three panel observations (wave 1- wave 3) irrespective if the household head could be followed or not. However, we now observe households splitting into more than 1 original household from wave 1. Therefore, in case the original household (w1_hhid) splits into two or three households (w2_hhid) the information from the variable "w2_stayer"/ "w3_stayer" was used. These variables contain the information if members stayed or left the household in-between waves. As Finn et al. (2012) have shown "stayers" seem to experience smaller income changes on average. We therefore followed the household half that reported to be a "stayer". With this procedure 4,644 households could be identified having 3 wave panel information.

again in 2007, 20155 (67%) in 2008 and 18,154 (60.3%) in 2009. As for NIDS we will only use the first 3 waves to maximise the number of households observed in consecutive 3 years. For these three waves 5,560 households can be observed. It is important to note that this is the only study which contains yearly data which will impact the interpretation of the mobility coefficients.

3.4.3. Tanzania Panel Data

For Tanzania the NPS containing information for the years 2008/2009; 2010/2011 and 2012/2013 was used for the analysis. The NPS is another large national representative household survey containing Tanzania mainland and Tanzania Zanzibar. In the first round it contained information on 16,709 individuals living in 3280 households. Even though in the second round, 20,559 individuals in 3,924 households were interviewed, this chapter will only make use of the 3027 households with income information for all three waves. That means attrition for the NPS was very low with only 7.3%. In addition, it is important to note that two-third of the households in the sample are rural. This is the highest share of rural observations for all four countries.

3.4.4. USA Panel Data

Finally, to test the reliability of the GMM approach and to compare the income mobility and measurement error estimates from the three developing country data sets, the PSID from the USA was analysed. The PSID has the advantage of being the longest available national socio-economic panel and numerous studies and working papers on income mobility are available to compare the results from this study. For the analysis the data from the years 2007, 2009 and 2011 were chosen, totalling 6,534 household information for these three waves.

3.5. Results

In this part the results from the empirical analysis are given. The focus will be on obtaining the true estimates of β the speed of convergence and α an estimate of measurement error. The variable of interest is the change in real per capita household income from all sources. Household income and not earnings was chosen to ensure comparability of the results between countries and data sets. E.g. for Tanzania – as for many developing countries -earnings estimates are only available for a very small part of the population and the numbers are less reliable than total household income[31]. Second, total per capita household income might be a better estimate of economic status if households in developing countries are large and only a few household members have regular earnings income.

[31] Household income, still might be influenced by seasonal income especially in Tanzania.

3.5.1. Income convergence in the USA

Table 2 columns (1)-(6) show the regression coefficients which represent θ_1 - θ_7 introduced in part 3. These will then compared to the coefficient from the 2SLS estimator presented in Column (7). Column (1) and (2) give the convergence coefficients that correspond to θ_1 and θ_2. They indicate that 30.2% of the income gap between any two households in the USA would be disappear between 2007 and 2009 and another 25.5% of the income gap would be eliminated between 2009 and 2011. This would imply that the half-life of the income gap in the USA is somewhere between 3.8 and 4.7 years respectively. This is similar to the findings by the study of Fields et al. (2003), giving convergence coefficients that are indicating half-life convergence between 0.7 years for Venezuela, 1 year for Spain and about 4.2 years for South Africa. However, recent findings from Chetty at al. (2014) using administrative data and calculating intergenerational mobility between sons and fathers come up with an intergenerational elasticity of coefficient of 0.45 which can be translated to a convergence coefficient of -0.55 and a half-time gap of about 26 years.[32] This suggests that the naïve convergence estimates are most likely biased downwardly by measurement error in household income as shown in part 2. Not surprisingly IV estimates of θ_2 given in column (7) provides a much lower coefficient of only -0.0543 implying a half-time gap of 24.7 years.

Table 2: Regression coefficients for USA (PSID) income regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS estimation of θ_1 - θ_7						**IV estimation**
	Δy_2	Δy_3	Δy_3	$y_3 - y_1$	Δy_3	Δy_3	Δy_3
y_1	-0.302*		-0.0373	-0.349*	0.286**		
	(0.0351)		(0.0106)	(0.0342)	(0.00890)		
y_2		-0.255*			-0.470**		-0.0543***
		(0.0367)			(0.0281)		(0.00840)
Δy_2						-0.363**	
						(0.0205)	
Constant	3.121*	2.572*	0.362	3.586*	1.876*	0.0140	0.537***
	(0.363)	(0.392)	(0.0917)	(0.371)	(0.216)	(0.0193)	(0.0741)
Observations	6,309	6,321	6,321	6,321	6,321	6,321	6,321
R-squared	0.172	0.120	0.003	0.190	0.200	0.144	0.046

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. (Source: PSID years 2007, 2009, 2011)

To examine the internal consistency of the coefficient estimates provided in Table 2, the predicted values for each θ_k, for the case that θ_1 represents the true convergence parameter β and $\alpha = 1$ (income is measured without error), are given in Table 3 row 2. None of the predicted values seem to be near the observed values θ_k, indicating that the case of β=-0.302 and $\alpha = 1$ does not provide internal consistent regression coefficients. Another way to test if there is no measurement error (implying $\alpha = 1$), is given in row 4 in Table 3. Here the value for β was calculated using the formula of Table 1 column 2 – the case of no measurement error. The estimates of β are reaching from -0.039 to -0.726 rejecting the hypothesis of $\alpha = 1$ for the possibility to observe consistent β. Therefore, the informal method seems to show that

[32] Given that the children in the sample where about 30 years old.

measurement error is needed to explain the income convergence process, although it cannot be formally tested in this way.

Table 3: Regression coefficients and implied parameter values for USA

θ_k	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7
Estimated values of θ_k	-0.302* (0.0351)	-0.255* (0.0367)	-0.0373 (0.0106)	-0.349* (0.0342)	0.286* (0.0089)	-0.470** (0.0281)	-0.363** (0.0205)
Predicted values of θ_k if $\beta = -0.302$ and $\alpha = 1$	-0.302	-0.302	-0.211	-0.513	0.000	-0.302	-0.151
Predicted values of θ_k if $\beta = -0.053$ and $\alpha = 0.737$	-0.302	-0.302	-0.037	-0.339	0.339	-0.539	-0.439
Value of β implied by θ_k (if $\alpha = 1$)	-0.302	-0.255	-0.039	-0.193	NA	-0.470	-0.726
Value of β implied by θ_k (if $\alpha = 0.737$)	-0.053	0.11	-0.057	-0.057	-0.089	-0.112	-0.178

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. (Source: own calculation)

In a next step, allowing for measurement error new values for β and α were estimated from equation [7] obtaining estimates of $\hat{\beta} = -0.053$ and $\hat{\alpha} = 0.737$. These values where then used to predict the values of θ_k in row 3 in Table 3. Even though, these predictions don't fit the observed values in column 1 perfectly, they are still significantly closer than the non-measurement error predictions in shown in column 2. Similarly, allowing for measurement error the implied β from each regression coefficient with an $\alpha = 0.737$, are much smaller and lie in a more narrow range between 0.11 and -0.178. Thus, allowing for measurement error gives much more consistent estimates from the informal approach and significant lower estimates for the convergence coefficient β.

A more efficient and formal approach to estimate β and α is the system GMM estimator. Here the regression coefficients θ_1 - θ_7 from Table 2 are used to solve equation [8] without putting any restriction on the value of α and β. The results in column 1 of Table 4 show a β of -0.053 which implies that 5% of the income gap will be eliminated between each wave. This is very close to results of equation [7] as well as the IV estimates but only a sixth, respectively a fifth of the naïve OLS estimates θ_1 and θ_2. The results are highly significant and have much smaller standard errors than the two other estimators. In addition, the GMM estimator allows the formal testing of the validity of the over-identifying restrictions. The J-statistic of 0 and a p-value of 1 implies that all five linearly independent regression coefficients are solved in a way that is internally consistent.

Given a speed of convergence of 5%, the new resulting half-life gap for the USA, would be about 26.7 years, which overlaps very much with the findings of Chetty at al. (2014) and is in line with other studies from the US using the PSID (Solon, 1999, 2002). Second, the estimated α of 0.785 implies that 21.5% of variation in household income can be explained by measurement error and 78.5% is due to actual variation. A reliability statistic of 0.78 again is in line with other validation literature for the US e.g. Abowd & Stinson (2013) who find that self-reported earnings of US workers have a reliability statistic of 0.7.

Table 4: GMM estimates for US income dynamics

	(1)	(2)
β	-0.0503***	-0.103***
	(0.00799)	(0.00205)
α	0.785***	1
	(0.0339)	-
Observations	6,271	6,271
J-test statistic	0	1.933
p-value	1	0.748

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. (Source PSID years 2007, 2009, 2011)

As a last step, restricting α=1 will estimate the best fit for β assuming no measurement error. Column 2 of Table 4 shows that the estimate of -0.102 is twice the size of the earlier estimate but still much smaller than obtained by the naïve OLS regression. Yet, the J-statistic and p-value of 0.748 rejects the hypothesis of no measurement error since the over-identifying restrictions don't give a consistent estimation of β.

3.5.2. Income convergence in South Africa, Chile and Tanzania

Upward economic mobility is one of the most important indicators for economic development. However, most estimates of economic mobility rely on the assumption of no measurement error of income, an assumption which has to be rejected for most panel sets, as shown for the US in the previous section. As a result, conventional income convergence coefficients and the resulting half-life estimates are overestimated and underestimated respectively. Yet, the results from the last section for the USA have shown that the GMM estimator produces β and α estimations which are more efficient and in line with published validation studies. Therefore, the approach will now be applied for panel studies from South Africa, Chile and Tanzania.

First, the regression coefficients θ_1 - θ_7 for all three countries are produced in Table B1-B3 in the Appendix. For South Africa the estimates of θ_1 and θ_2 obtained from regressing Δy_2 on y_1 and Δy_3 on y_2 are both about -0.25, which would imply that half of the income gap between the richest and poorest South African household ought to be eliminated within 4.8 years. For Chile θ_1 and θ_2 are equal to -0.362 and -0.449. Transferred to half-life estimates this would indicate that after only 1.5 or even 1.2 years half of the income gap disappears. Given these rather different coefficient estimates between the two waves, the maintained assumption of a constant convergence coefficients maintained by our GMM estimator may be problematic. The naïve convergence coefficients for Tanzania are -0.536 and -0.485, indicating half-life estimates 1.8 and 2.1 years. Even though all three countries might experience some income mobility these high mobility estimates may well be symptomatic of the presence of measurement error.

Table 5 reports the β and α estimates from the GMM model for the three countries.[33] All convergence coefficients are much smaller than the conventional estimates suggest. For South Africa the conventional approach over-estimates by more than a factor of 4 in comparison with the GMM estimate of -0.0590. Column 3 shows that Chile's best estimate for β is -0.0747 and that the naïve convergence coefficients might have been inflated by a factor of 5 to 6. Tanzania's β equals to -0.114. These results imply new half-life approximations of 22.7 years for South Africa, 8.9 years for Chile and 11.4 years for Tanzania. All coefficients are significant at the 1% level and estimated highly efficient, as can be observed by the small standard errors.

In terms of data reliability, the GMM model provides estimates between 0.8 for South Africa and 0.634 for Chile and even 0.552 for Tanzania. This indicates that 20% of variation in reported household income is due to measurement error in South Africa, but about 36% in Chile and 45% in Tanzania. This means, while the data reliability for the US, Chile and Tanzania correspond to their levels of economic development. Comparing the results between the two African economies it seem like the data reliability parameter is significantly higher for South Africa than for Tanzania. There are two interpretations for why this might be the case. First, a purely technical reason is that South Africa's income distribution is much more unequal than that of Tanzania. In this case, given the same absolute level of measurement error, the data reliability parameter α would be larger for South Africa, due to the larger variation of y_{t-1}. Hence, at a given level of relative income misreporting, South Africa will therefore have a higher reliability estimator. Second, it might be that Tanzania's household income process is more influenced by shocks due to its larger share of subsistence economic activity.

Table 5: GMM estimates for South African, Chile and Tanzania income dynamics

	South Africa		Chile		Tanzania	
	(1)	(2)	(3)	(4)	(5)	(6)
β	-0.0590***	-0.0886***	-0.0746***	-0.151***	-0.114***	-0.109***
	(0.0174)	(0.00455)	(0.0189)	(0.00401)	(0.0308)	(0.00389)
α	0.801***	1	0.634***	1	0.552***	1
	(0.0195)	-	(0.0181)	-	(0.0267)	-
Observations	2,770	2,770	5,382	5,382	2,888	2,888
J-test statistic	0.249	73.2	16.96	428.6	3.461	142.7
p-value	0.969	0	0.001	0	0.326	0

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. (Source: NIDS, EPC and NPS panels)

Next, the calculated J-statistic allows us to test for the validly of the over-identifying restrictions. For South Africa the J-statistic of 0.249 with a p-value of 0.969 indicates that the GMM estimator can calculate all five linearly independent regression coefficients in a way that is internally highly consistent. Similarly, consistency

[33] For South Africa the GMM estimates using a larger household sample, relaxing the definition of core households can be found in Table B4 (in the Appendix).

can't be rejected for Tanzania given a p-value far above 0.05. Only for Chile the hypothesis that all coefficients can be estimated by the GMM model in an internal consistent way has to be rejected with a p-value of 0.001. However, this might be due to the very different coefficient estimates between the two waves, which violates the constant convergence hypothesis. Finally, column (2), (4) and (6) all highly reject the assumption of no measurement error in the data sets.

To check for the robustness of the results, all regression coefficients have also been calculated for sub-samples of the data, excluding 0.1 % of largest income changes at the top and 0.1% at the bottom, in all countries. As expected the results in Table 6 show that the data-reliability coefficient α increases for all countries when excluding the largest outliers. In particular, for the USA α is 0.911 implying that 91% of the variation in log household income is now due to variation in actual income and only 9% remains due to measurement error. Furthermore, the gap between the naïve and the GMM approach is much smaller with a regression coefficient for θ_1 of -0.187 versus the GMM coefficient of -0.085. Still even this new estimates show that measurement error will inflate the naïve regression coefficients by more than twofold for the US, about three times for South Africa and Tanzania and four times for Chile.

Table 6: Regression coefficients for all countries excluding extreme values

	USA		South Africa		Chili		Tanzania	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Full Sample	No extreme values	Full Sample	No extreme values	Full Sample	No extreme values	Full Sample	No extreme values
	Δy_t	Δy_t	Δy_t	Δy_t	Δy_t	Δy_t	Δy_t	Δy_t
$OLS: y_1$	-0.302*	-0.187**	-0.25***	-0.200***	-0.362***	-0.314***	-0.536***	-0.46***
	(0.0351)	(0.0110)	(0.0247)	(0.0204)	(0.0165)	(0.0120)	(0.0221)	(0.0201)
$GMM: \beta$	-0.05***	-0.085***	-0.059***	-0.071***	-0.075***	-0.078***	-0.114***	-0.14***
	(0.00799)	(0.00521)	(0.0174)	(0.0137)	(0.0189)	(0.0135)	(0.0308)	(0.0261)
$GMM: \alpha$	0.785***	0.911***	0.801***	0.868***	0.634***	0.712***	0.552***	0.640***
	(0.0339)	(0.0118)	(0.0195)	(0.0147)	(0.0181)	(0.0133)	(0.0267)	(0.0245)
J-statistic	0	0	0.249	1.566	16.96	7.476	3.461	4.341
p-value	1	1	0.969	0.667	0.001	0.058	0.326	0.227
Observations	6271	6,057	2,770	2, 669	5,382	5,197	2,888	2,784

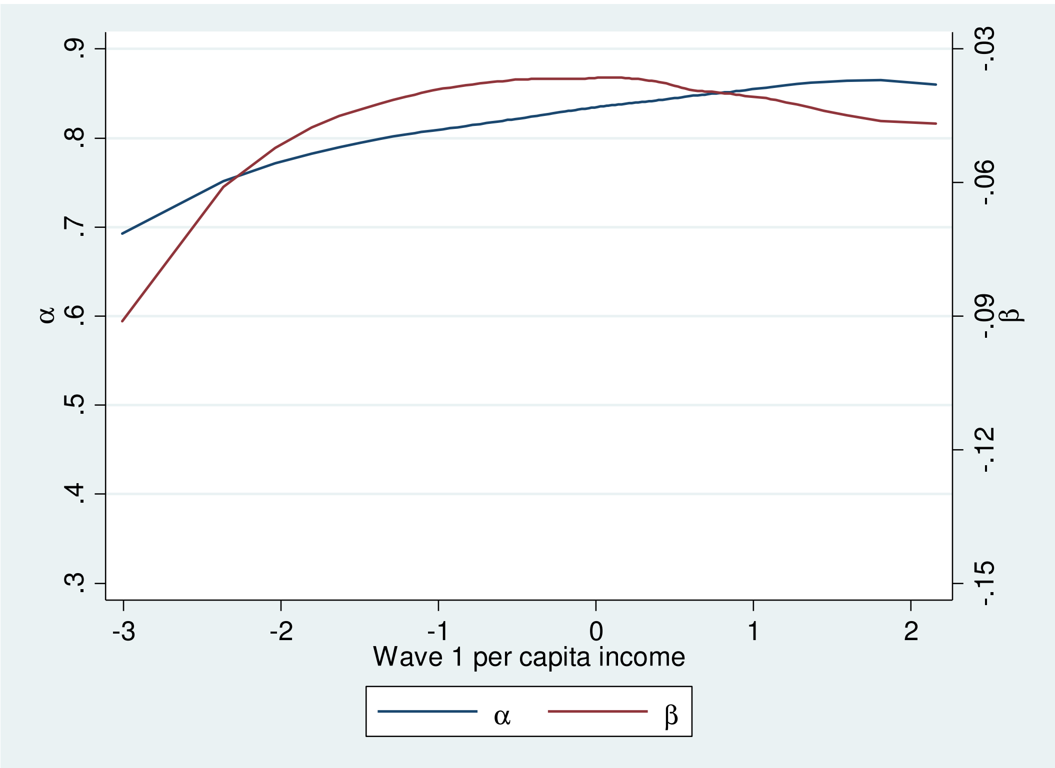
Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. (Source: PSID, NIDS, EPC and NPS panels)

3.5.3. Results from non-parametric approach

In general, the results of the GMM estimator have highlighted the problem of measurement error in income mobility estimates for all countries. They also question if high transition rates in and out of poverty observed for many developing countries (e.g. Baulch and Hoddinott, 2000; Dercon and Krishnan, 2000), are due to true income shocks or simple measurement error. To find out whether or not measurement error affects income observations of poor and rich households in the same way this section makes use of the non-parametric approach which is also less restrictive on the form of the measurement error, allowing for β and α to vary with the income levels.

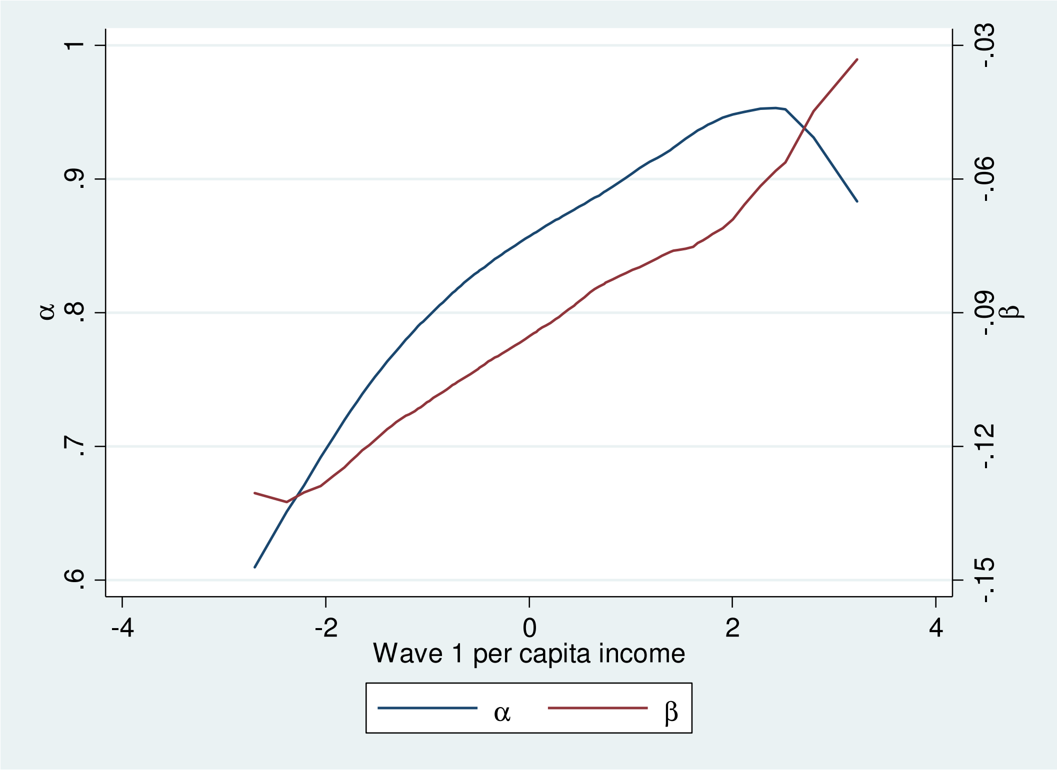
Figures 1-4 show that the non-parametric estimates of β increases with initial income levels for each of the four countries considered. This indicates that poorer households are more income mobile, while the incomes of richer households are more stable. However, while β estimates ranges in relatively narrow ranges of about -0.09 to -0.04 in the USA, and -0.15 to -0.8 in Chile, this rate of convergence is much more varied across income groups in South Africa (-0.14 to -0.03) and particularly in Tanzania (-0.50 for poor households and -0.10 for rich households). This means income convergence strongly depends on initial income in Tanzania. On the other hand, the income reliability statistic α seems to show quite high values (around 0.75) at the top and the bottom and of the income distribution and more measurement error for average households (around 0.65) in Tanzania. For the three other countries measurement error predictions are again larger for poor households and lower with increasing household income. This relationship seems to be particularly steep in South Africa, with a share of measurement error explaining household income variation up to 40% for poor households and only 5% for the richest ones.

Figure 1: Nonparametric estimates of $\beta(y_1)$ and $\alpha(y_1)$ for the USA



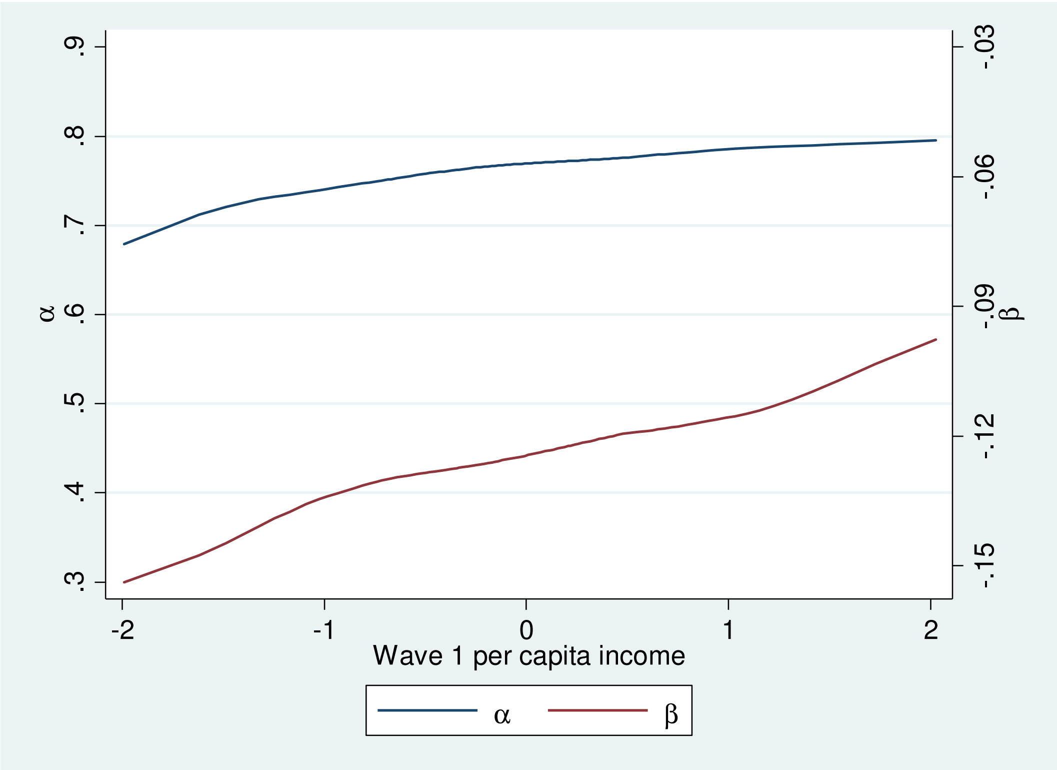
(Source: PSID years 2007, 2009, 2011)

Figure 2: Nonparametric estimates of $\beta(y_1)$ and $\alpha(y_1)$ for South Africa



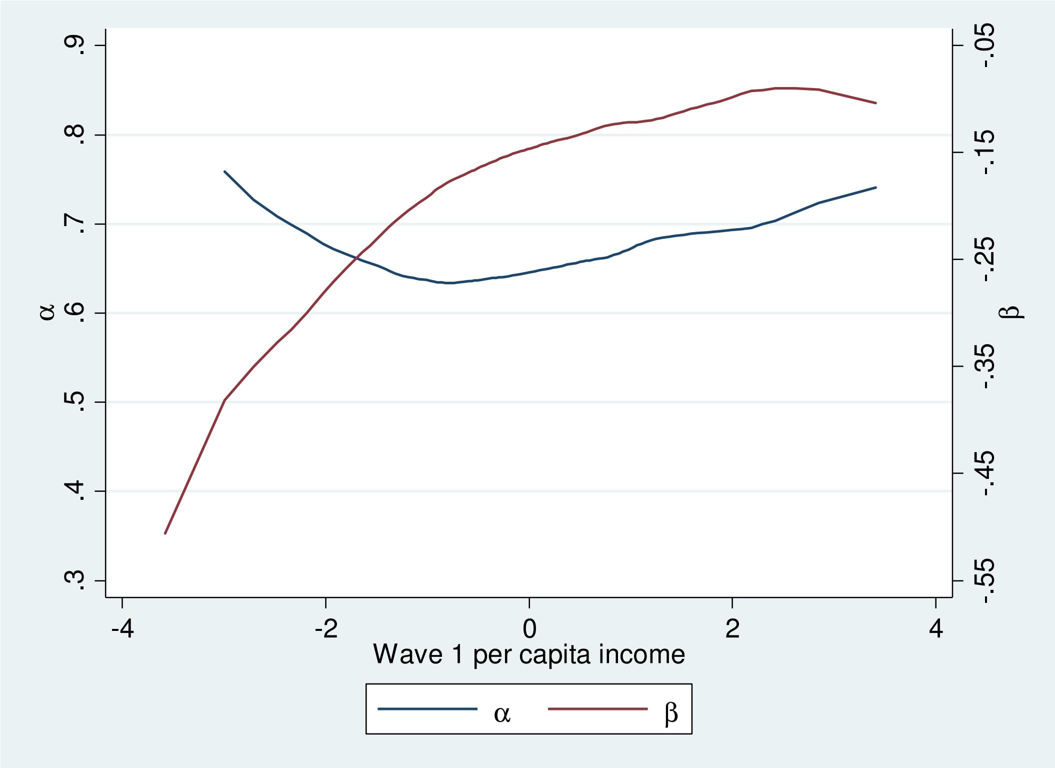
(Source: NIDS waves 1-3)

Figure 3: Nonparametric estimates of $\beta(y_1)$ and $\alpha(y_1)$ for Chile



(Source: EPC waves 1-3)

Figure 4: Nonparametric estimates of $\beta(y_1)$ and $\alpha(y_1)$ for Tanzania



(Source: NPS wave 1-3)

3.6. Conclusion

Upward economic mobility continues to be one of the most important indicators for economic development. However, the results of this chapter have shown that without controlling for measurement error naïve OLS regressions produce convergence coefficients that are unrealistically large and that are inconsistent with other moments in a three wave panel dataset. In order to quantify the effect of measurement error in household income data and to obtain less biased β estimates a newly developed GMM estimator was applied to four large national panel studies. The results for the USA imply that the naïve OLS approach overestimates income mobility by a factor of 5-6, while the GMM model estimates a convergence coefficient that is internally consistent and translates to a half-life income convergence gap of about 26.7 years which is in line with previous validation studies.

We also apply this method to three developing countries: South Africa, Tanzania and Chile. Our results indicate that the hypothesis of no measurement error can be rejected for all the countries we consider and the α estimates imply that the share of income variation that is due to actual income changes is only 55% for Tanzania, 63% for Chile, 79% for the USA and 80% for South Africa. Therefore, while the data reliability for the US, Chile and Tanzania correspond to their levels of economic development, South Africa's data reliability appears to be unexpectedly high. For the convergence coefficient the GMM model provides a new β of -0.059 or a half-life gap of 22.7 years for South Africa, a β of -0.075 or a half-life gap of 8.9 years for Chile and a β of -0.114 or a half-life gap of 11.4 years for Tanzania. These results show that initial estimates were overestimated by a factor 4-6.

Finally, using a nonparametric approach that allows β and α to vary with the initial income levels shows that the convergence coefficient is indeed larger for poorer households in all countries. This is particularly true for Tanzania and – to a lesser extent – for South Africa, the households at the bottom end of the income distribution are much more upwardly mobile than the downward mobility experienced by richer households. In a similar way, the data reliability is lower for the poorer households and better for richer households in all countries.

In general, the results of this chapter indicate that income mobility can be substantially overestimated without controlling for measurement error. Researchers that use panel data to estimate poverty and economic mobility should therefore be concerned to use appropriate estimators that can deal with the problem of measurement error in household income. On a more positive note, even if initial estimates were overly optimistic the results still show significant convergence for all countries, implying that households below the average are growing faster and catching up with the richer households.

Chapter 4:

The effect of neighbourhoods and school quality on educational and labour market outcomes

4.1. Introduction

The extent to which family, neighbourhood and the quality of schooling can explain differences in socio-economic outcomes is an important question that has produced several studies in the past (e.g. Card and Krueger, 1992; Case and Deaton, 1999; Case and Yogo, 1999; Solon et al., 2000; and Altonji and Mansfield, 2011). In light of recent findings by Lam et al. (2011) that movement through secondary schooling is still predominantly along racial lines, the effect of special location and schooling outcomes is a particularly important research question for South Africa. In addition, South Africa provides an interesting setting to observe neighbourhood and school effects, due to its unique history and sharp differences between wealth and school quality in different neighbourhoods, which were shaped under the apartheid regime (Case and Yogo, 1999). Given South Africa's high income inequality levels, making sure that child from poor socio-economic background and former disadvantaged areas also receive quality education is essential to transform the education system into a resource for increased opportunity and racial equity.

Because of the strong correlation between family wealth, school choice, neighbourhood and schooling outcomes, it is very difficult to disentangle the relative importance of each factor. Previous studies analysing school and neighbourhood effects didn't have the necessary data to observe school quality, neighbourhood wealth and household characteristics at the same time. Furthermore, since these factors are highly correlated, imperfect measurement of one will make the other factors seem more important than they really are. The only way in which the relative importance of these education and wage determinants can be accurately measured is to find more reliable measures of each of these factors.

This chapter makes an important contribution to the literature by building a model which explains schooling and labour market outcomes, given more informative measures of students' location, parents' background as well as school quality measures. To do so, the National Income Dynamics Study will be linked to the master list of schools in South Africa, as well as to Census 2011 community data. Spatial linking of the Census 2011 will allow us to construct a new wealth index for 85,000 small areas and identify precisely the neighbourhood wealth of each household and school in the data. It will be shown that this wealth measure for each school is an excellent proxy for school quality, which explains schooling outcomes more accurately than the official school quintiles provided by the department of education (this is true at least in metropolitan regions). Finally, the data allows us to observe whether a child is going to school in the same neighbourhood he/she is living and the actual distance he/she travels to school every day.

Having constructed the data in this unique way, this chapter will try to answer some questions that have important policy implications. How large are family background effects relative to school quality effects when explaining schooling outcomes? What is the effect of sending a child from the poorest neighbourhoods to the best schools in the region? Are there long-run effects of going to a high quality school that determine university enrolment and an earnings premium? While this chapter will show large differences in education outcomes depending on the quality of the school, it can also demonstrate that even children from the poorest neighbourhood would perform well if they went to these high performing schools. However, given the limited number of quality schools in the country and the financial as well as transport constraints faced by the poor, only about 10% of the poorest 60% of children actually manage to get high quality education. Hence, to achieve more equal education outcomes, the quality of schools in the poor neighbourhoods would need to be improved. This would have large effects, as it can be shown that there are indirect effects as well as a direct premium for quality education in the labour market.

The structure of this chapter will be as follows: first, a short background and literature review is given, second, the three main data sets and the merging process with which the datasets were linked will be explained, and descriptive statistics and maps are provided. Lastly, a regression analysis for education outcomes and labour market earnings is performed. Finally, a conclusion about the findings are drawn.

4.2. Background and Literature

One of apartheid's enduring legacies is the large inequality of education opportunity for children from different racial and socio-economic backgrounds. This situation is rooted in the apartheid school system that created separate departments for children of different race groups: White, Indian, Coloured and Black (Hill, 2016). The result was four school systems within South Africa with vastly different resources[34], curricula and learning environments. While transforming the education system to achieve equal opportunity has been an important policy of the post-apartheid agenda, the institutional memory of the former school departments are still causing significant differences between schools along racial lines (e.g. Van der Berg, 2007; Van der Berg et al., 2011; Taylor et al., 2013; and Yamauchi, 2011). In particular former black schools (also called township schools in urban areas) have not seen much racial mixing and are still under-resourced with numerous administrative problems. In the worst case, the situation at these schools can be described as a culture of learning where teaching is almost non-existent (Msila, 2005). By contrast, former white and Indian schools are now much more racially diverse, although not socio-economically diverse, due to their fee-charging structure (Yamauchi, 2011).

Today, because of the *exit* option and due to new laws stating that no child can be excluded from a school for financial reasons (Hunter, 2015), in theory any parent is allowed to send their children to former Model

[34] The situation in the 1960s was that black students received only one-tenth of the school funding as white children (Hunter, 2015).

C schools (formerly white schools). However, during apartheid, different population groups were also segregated in separate residential areas which means good schools are still located in selected areas (Yamauchi, 2011). As a results, de Kadt et al. (2014) find for the "Birth to Twenty cohort study" of 1,428 children in Soweto-Johannesburg that over a third of them travel more than 6km to school, 60% leave the suburb they live in and only 18% attend their nearest school. However, household-level financial constraints caused by schooling fees, additional transport and time costs will prevent the poorest children from going to better schools in practice. A qualitative study by Msila (2005) interviewed parents currently sending their children to historically black schools. She observes that mostly distance and economic reasons prevent parents from sending their children to former Model C schools. In terms of racial patterns, another study by Hill (2015) show that coloured families in Cape Town are 50% more likely to send their children to "better" schools than black families.

Because of the strong correlation between family wealth, school choice, neighbourhood and peer effects, it is very difficult to disentangle the relative importance of each factor in schooling and labour market outcomes. That is, do schools produce different outcomes because they influence student performance or because they were able (or failed) to attract students that would have succeeded regardless of the school chosen? Similarly, do children in rich neighbourhoods perform well at school because of the neighbourhood they live in, the school they go to or the family they come from? The answer to this question is of particular importance due to its policy implications (Antonji and Mansfield, 2011). In the case where most of the variation of learning outcomes can be explained by parents' socio-economic status (SES) and their involvement in students' learning behaviour, increasing school funding of the poorest schools will be unlikely to improve student outcomes. On the other hand, if parents' SES is only significant because it is a proxy for the quality of the school their children go to, then improving school quality is likely to have large effects, in particular for the poorest students (Altonji and Mansfield, 2011).

To solve the correlation and identification problem, the seminal study by Solon et al. (2000) used a variance decomposition approach to bound the proportion of socio-economic outcomes that can be attributed to disparities in family and neighbourhood background by using the correlation between siblings and unrelated neighbours. The neighbourhood correlation captures the pure neighbour effect but also family traits because of the sorting mechanism and therefore is an upper bound[35]. Previous studies adapting the Solon et al. (2000) methodology found relatively high sibling correlations in brother income and education for China, the US, UK and Germany and smaller effects for the Nordic European countries (e.g. Björklund et al., 2004; Raaum et al., 2006; Lindahl, 2011; Eriksson and Zhang, 2012; Nicoletti and Rabe, 2013; or Schnitzlein, 2014). The proportion that can be explained by disparities in neighbourhood background seem to be nearly zero for income and small for education. Other studies for the US that focused on school quality and its

[35] The sorting mechanism means that parents with similar characteristics will move into the same neighbourhoods. Therefore, there is a strong correlation between parents' characteristics and neighbourhood wealth.

importance in reducing disadvantage from one generation to the next, also found very small effects (e.g. Jenning et al., 2015).

This literature is in sharp contrast to studies by Card and Krueger (1992), Case and Deaton (1999), and Case and Yogo (1999) who find for the USA as well as South Africa that school quality measured by pupil/teacher ratios have large and significant effects on the return to schooling for black men. Hanushek, Lavy, and Hitomi (2008) show that children in Egypt were more likely to drop out of low quality primary schools, which is similar to findings by Harbison and Hanushek (1992) who observe a reduction in repetition rates for Brazilian primary students when going to high quality schools. Another study by Glewwe et al. (2014) show that whether or not there are differences in learning between children from different backgrounds at the same school depends on the type of students and the country context.

There are at least two possible reasons for the contrasting results in the literature on the role of school quality. First, while the first bunch studies following the approach of Solon et al. (2000) explicitly try to disentangle the effect of neighbourhoods, family and school quality by using sibling correlations, the later might not efficiently separate family and school quality effects. Hence, those studies finding large schooling effects might be biased due to unobserved heterogeneity due to family effects. The other possibility is that as Case and Yogo (1999) rightly argue: "schooling quality should matter in countries like South Africa, where resources were distributed very unevenly between regions in the past" but these difference are not large enough in Western societies to matter. In the specific case of South Africa where we do know do that there are basically two school systems – those of the former white and those of the former back schools - we should definitely observe some quality differences in schools. Given the close proximity of poor and rich neighbourhoods and the clear distinction between poor and rich schools and children travelling in between these boundaries, this study should share light on the role of parental background, neighbourhood, and school quality. Furthermore, due to data constraints most studies, except for a view like Altonji and Mansfield (2011) or Jenning et al. (2015), focused on short run effects and didn't observe the long-term impact of school quality, family and neighbourhood background. Again, having long-run panel data this study will be able to address this issue.

4.3. Data and descriptive statistics

To analyse the effect of neighbourhood and quality schooling the National Income Dynamics Study (NIDS) is used. In this part, it will be explained how NIDS is merged with Census 2011 community spatial data and the master list of schools in South Africa. Some descriptive information and maps of the merged data will be provided.

4.3.1. Census 2011 data

The Census 2011 data is provided by Statistics South Africa (STATSSA). The primary sampling units (PSUs) were the 103,576 Census enumeration areas (EA) (see STATSA, 2012b). To obtain detailed information on the neighbourhood level the analysis was based on the "Small area layer" (SAL) from the Census 2011 Community Profiles which STATSSA provides in SuperCROSS. SAL are the second lowest geographical areas (after the EA level) in which the country is divided for the Census design. In a second step the information from the SAL were aggregated to weighted averages (using population size) to get the "Sub place" (SP) information. STATSSA (2012a) defines SP as "the second (lowest) level of the place name category, namely a suburb, section or zone of an (apartheid) township, smallholdings, village, sub village, ward or informal settlement". The Census community data provides information for about 85,000 SAL and 22,000 SP in South Africa. However, our analysis showed that there are SALs which do not have sufficient household information given their population size.[36] For those 629 SALs no wealth index was constructed but rather the weighted averages from the SP were given. Using SAL level information therefore has the advantage of identifying data irregularities we would not observe on the aggregate SP level and having smaller and more even distributed area samples to construct a wealth index using principal component analysis (PCA).

An example of SP and SAL maps can be found in the Appendix (Figure C1). For dense areas like the City of Cape Town each SP is divided into about 16 new SALs. Between 0 and 111,937 individuals live in a SP. However the median size is about 5,400 individuals living in 1,500 households for SP and about 580 individuals living in 160 households in a SAL.

One of the largest challenges in this approach comes from the fact that the data is provided on community level rather than household level. Therefore, all variables in SuperCROSS are given as total counts (e.g. the number of households with access to clean water in that particular SAL). These totals were transformed to percentages e.g. the percentage of households with access to clean water in that SAL. The wealth index was then constructed using PCA analysis for a set of variables, namely household income (in brackets), education (for everyone age 25-64), labour market status, household assets and household services. The household income is given in 12 brackets ranging from "R1-4800" to "R2457601 or more" and "no income". The four lowest income categories and the highest five categories were grouped and aggregated together. Similarly education was grouped to "No education", "Some primary education", "Secondary education", "Matric" and "Higher education". In addition only education levels for the age group 25-64 were chosen to observe completed education and not enrolment. For a complete list of all variables used to construct the wealth index, see Table C1 in the Appendix.

Figure C2 in the Appendix shows the distribution of the wealth index. To make the wealth index comparable to the official school quintiles provided by the department of basic education (DBE), quintiles of wealth were generated using the population size of each SAL and SP respectively as weights. Wealth quintile

[36] For more detailed information on the net Census coverage error see (STATSSA, 2012b).

information has been merged with the national school master list and NIDS learner data using GPS coordinates and GIS software as described below.

4.3.2. Master list of schools

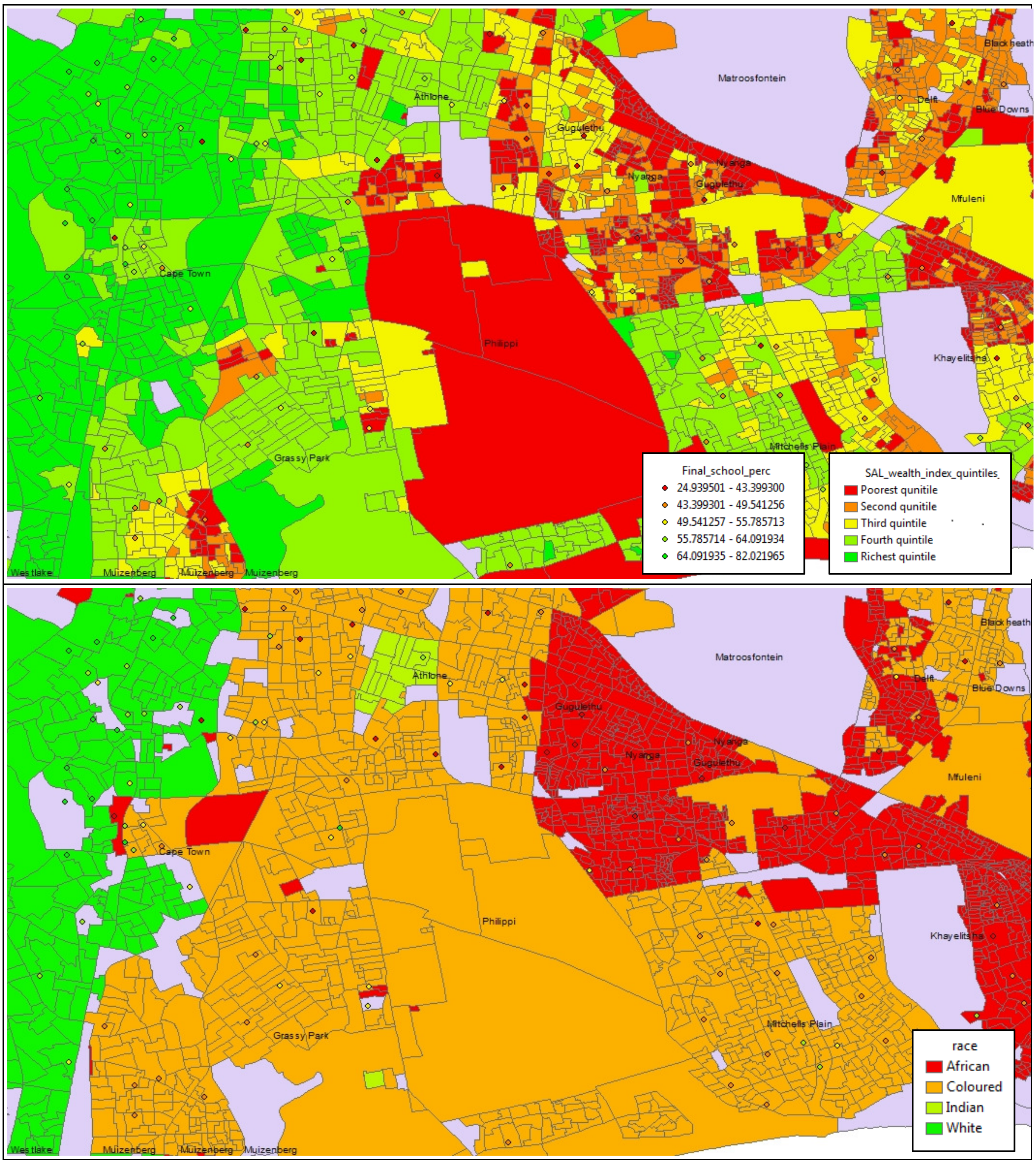
The master list of schools in South Africa provided by the DBE has detailed information for all 25,827 schools of the country, including ex-department, quintile, learner-teacher ratios and GIS information which are used for the matching process[37]. The upper panels of Figure 1 and Figure 2 show the newly calculated neighbourhood wealth quintiles and the average matric examination results for 2014 for the municipalities of Cape Town and Johannesburg.[38]

The neighbourhood wealth quintiles are illustrated by colours: green for the richest and red for the poorest areas. Similarly, the matric examination results for 2014 are also divided into quintiles and coloured in green for best and red for worst performing schools. As mentioned before, due to the legacy of apartheid there is a strong correlation between the neighbourhood a school is based in and the average school results it produces, which clearly can be seen in the maps.

In addition, the bottom panels of Figure 1 and Figure 2 illustrate how racially segregated South Africa still is today. That is, there appears to be some clustering for each race group where their share of the SAL population exceeds more than 50%. Hence, given the institutional memory of the schooling departments from the apartheid regime and the clear racial separation, we would expect some form of omitted variable bias in any schooling model which doesn't sufficiently control for household and schooling location.

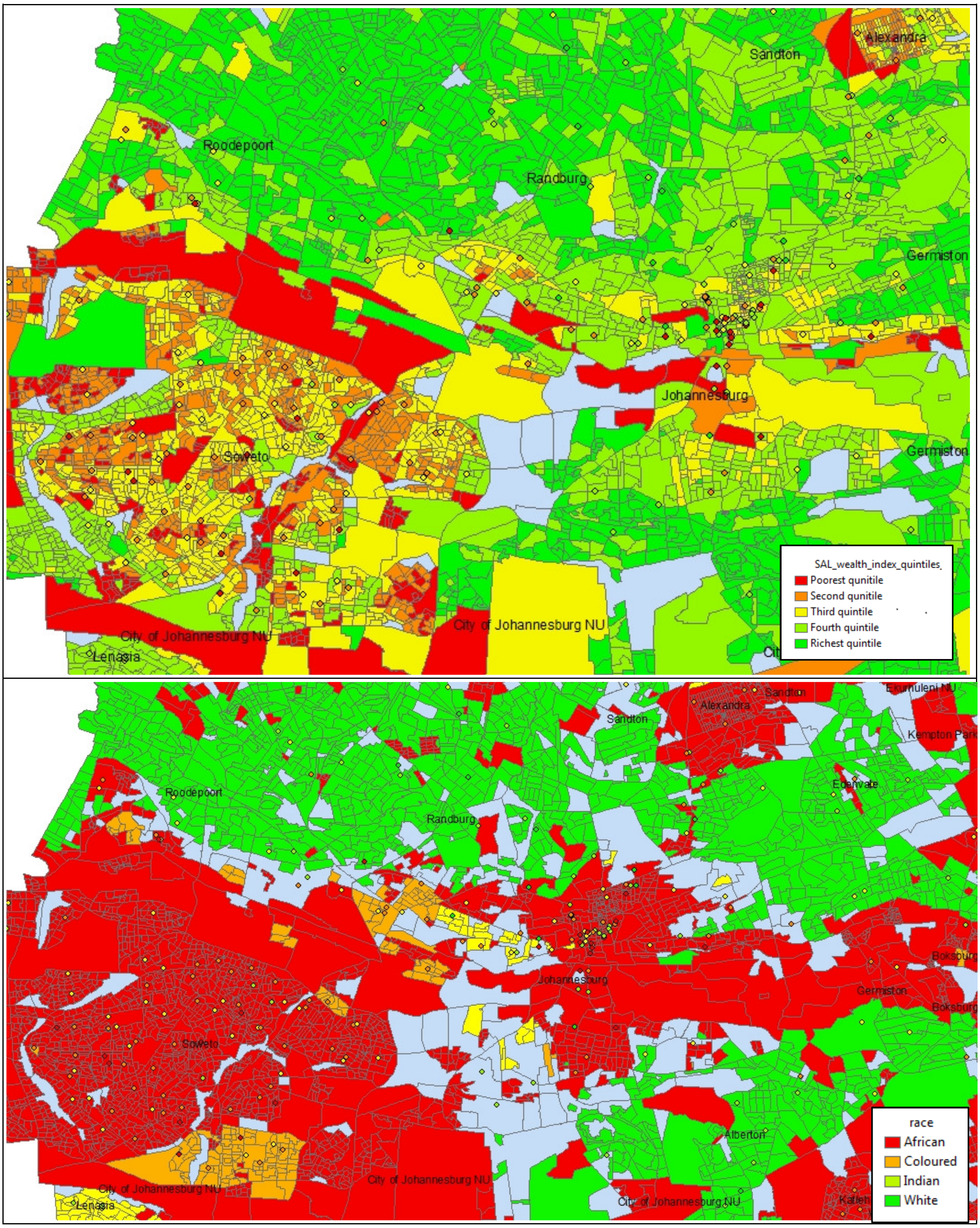
[37] For more information on the master list of schools in South Africa see Van Wyk (2015).
[38] For the overall picture of the metropolitan regions Cape Town; Johannesburg, Tshwane and Ekurhuleni; as well as eThekwini see Figure C2, Figure C3, and Figure C4 in the Appendix.

Figure 1: Neighbourhood wealth quintiles, geographic race distribution and matric 2014 examination results in metropolitan Cape Town



Note: The colours in the bottom panel indicate a share larger than 50% for a particular race group in that SAL in the Census 2011 data. In addition, the average matric 2014 examination results per school are displayed.
(Source: Census 2011 and DBE school data)

Figure 2: Neighbourhood wealth quintiles, geographic race distribution and matric 2014 examination results in metropolitan Johannesburg



Note: The colours in the bottom panel indicate a share larger than 50% for a particular race group in that SAL in the Census 2011 data. In addition, the average matric 2014 examination results per school are displayed.
(Source: Census 2011 and DBE school data)

4.3.3. NIDS

Besides the publicly available data, NIDS also provides secured data on GPS household location, as well as the names of all the schools children are going to or the last school a respondent went to before leaving the school system. Using the name of the school, as well as the household location, it was possible to merge in further information (containing GPS data for each school) from the master list of schools using fuzzy matching.[39] Having GPS coordinates for all households as well as schools makes it possible to link the new neighbourhood wealth index from the Census 2011 community data to the household and school location using the SAL and SP maps described in the last section. This data also allows us to calculate the distance to the closest school, as well as determining the actual school students attended.[40]

The maps shown in Figure 1 and Figure 2 (and Figure C3, C4 and C5 in the Appendix) all paint the picture of municipalities where poor and rich communities lie in close proximity to each other. They also show the historic placement of well-funded former model C (white) schools in the formal areas close to the town centres. Given the findings of previous studies that most parents would like to send their children to former model C schools (e.g. de Kadt et al., 2014; Msila, 2015) we should observe at least some children in NIDS from the poorest neighbourhoods commuting to richer former model C schools in town. Indeed, Table 1 shows that a substantial part of the student population from the poorest 40% of neighbourhoods are traveling to schools that lie in wealthier areas than their own household location. In addition, Table 2 and Table 3 give the wealth of the school and student location by race group. While black students predominantly live and go to school in the poorest two neighbourhood quintiles, white children live in the top two quintiles and almost always attend the richest schools. Interestingly, it seems that Indian and to some extent coloured students usually go to schools based in richer areas than their own.[41] These descriptive findings can be formally tested in a simple OLS model which regresses the difference in the revised school quintile to the learner wealth quintile on race, age and a couple of household characteristics (as seen in Table C3 in the appendix). It shows that Indian families are more likely to send their children to high quality schools. The same is true for parents with more years of education (column 1), parents with matric (column 2) and households with higher per capita income (column 3). Finally, it appears to be easier for families to send their children to richer schools when they reside in urban areas. These results seem to be in line with other studies on school choice behaviour for South Africa (e.g. Msila, 2005; de Kadt et al.; 2014; Hill, 2015; and Hunter, 2015), providing confidence that the wealth quintiles provide sensible information for further analysis.

[39] Fuzzy matching was done using the user written Stata command "reclink". That means in case no perfect match between the key fields in the two datasets exist, the best match, ranked by a matching score, was manually reviewed.
[40] This data was accessed through the DataFirst Secure Research Data Centre at the University of Cape Town.
[41] Looking at provincial mobility, Table C2 in the Appendix show that, students from the rural province of the Eastern Cape are slightly less mobile in terms of moving between richer and poorer neighbourhoods than children in the Western Cape.

Table 1: Movement between school and student neighbourhoods for poorest 40%

		Frequency	Percentage
School wealth quintile – Student wealth quintile	-1	449739	6.20%
	0	3224914	47.50%
	1	1829967	27.00%
	2	492821	7.30%
	3	634828	9.40%
	4	158415	2.30%

This table uses the neighbourhood wealth quintile of the school and of the students' location from the Census 2011 community data (Source: Census 2011 and NIDS wave 1-3: age 14-18).

Table 2: Students neighbourhood quintile and race

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Black	50.81	24.37	12.44	9.45	2.93
Coloured	6.73	22.03	30.87	31.13	9.23
Indian	32.79	0	1.64	32.79	32.79
White	0.95	3.81	2.86	15.24	77.14

(Source: NIDS wave 1-3)

Table 3: School quintile and race

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Black	39.08	24.73	17.24	10.46	8.49
Coloured	1.8	6.1	22.62	33.21	36.27
Indian	0	2.08	2.08	12.5	83.33
White	1.33	0	1.33	5.33	92

(Source: NIDS wave 1-3)

As stated earlier, the data allows us to calculate the actual distance students travel to school (see Table C4 and Table C5 in the appendix). White and black learners travel the furthest on average. As expected, given the higher number of primary schools, children travel about twice as far in secondary than in primary school phase. Interestingly, while rural children have to travel further in the primary school phase, urban children have to travel further in the secondary school phase to go to school. Those finding are in line with previous studies (e.g. Hunter, 2015) showing that middle class white students are actually driving relatively far to go to prestigious good schools in the city, while black students have to travel all the way from the townships to the city centre to access quality education.

4.4. Regression analysis

In this part, the relative importance of family background, neighbourhood effects and school quality will be tested. In a first step, it should be determined how well the school neighbourhood performs as a proxy for school quality in comparison to the official school quintiles from the DBE. Second, using this proxy for school quality and wealth, different regressions on education and labour market outcomes are run on the NIDS panel data.

4.4.1. Revised versus official DBE school quintiles

Using a very simple model, we compare the appropriateness of the two school wealth measures predicting the average matric 2014 school results. This should give a first indication on how well the revised school wealth measure performs in comparison to the official DBE school quintiles.[42] The results in Table 4 column (1) and column (3) show that for the full sample containing all the schools in the country, both measures seem to do similarly well in explaining the variation in average school results with an R^2 of 34% and 33% respectively. On average, the higher the school quintile the better the school performs. It is worth noting that both measures predict that children enrolled in the top quintile school have 11 percentage points higher matric examination results than those in the poorest quintile.

In column (2) and (4) the same specifications are estimated for the metropolitan samples of the cities of Cape Town, eThekwini (Durban), Johannesburg and Pretoria. For schools based in these metropolitan regions, it appears that the current neighbourhood wealth quintile of a school (column 4) has a significantly higher model fit (R^2) than the official school quintiles (column 2). In addition, there seem to be larger effects of going to quintile 4 and 5 for the revised measure than the official one. Of course, by not controlling for household wealth or parent's education there is a large omitted variable bias in this simple model, so one should be cautious of interpreting the coefficient estimates as the causal effects of school resources. That is, children living in wealthy neighbourhoods have richer and better educated parents, which should also influence a child's matric performance. However, it is interesting that the coefficient for the school neighbourhood wealth index decreases by about two-thirds after controlling for the share of SGB-teachers[43], which should give a first indication that this measure is a good proxy for the funding and quality of the school which may be less vulnerable to measurement error induced attenuation bias.

[42] The official DBE school quintiles are taken from the master list of schools. The quintiles were developed by the DBE in the late 1990s using information from the first Census 1996 to rank the schools based on the wealth of their community (van Wyk, 2015). These quintiles are still highly relevant today since they determine the financial support a school is entitled to by the government. Second, in the absence of other measures they have been used as a proxy for school quality in many studies (e.g. van der Berg, 2008; Spaull and Kotze, 2015). Our revised school quintile measure has the advantage that it is based on much newer Census 2011 data, which provides more recent income information regarding current wealth in a neighbourhood, particularly in the metropolitan regions.

[43] SGB-teachers are employed by the school governing body (SGB) and not by the government. SGB-teachers might be more motivated to teach than public teachers since they have different contracts and can potentially lose their job. In addition, they show that the school has funding capacities to employ these extra teachers.

Table 4: Matric 2014 average school results and school quintiles

	(1) Full sample	(2) Metro	(3) Full sample	(4) Metro	(5) Metro	(6) Metro
VARIABLES	Matric average school percentage	Matric average school percentage	Matric average school percentage	Matric average school percentage	Matric average school percentage	Matric average school percentage
2. DBE school quintile	1.361***	-0.411				
	(0.229)	(1.041)				
3. DBE school quintile	1.486***	-1.389				
	(0.218)	(0.921)				
4. DBE school quintile	4.325***	0.420				
	(0.254)	(0.907)				
5. DBE school quintile	11.10***	7.811***				
	(0.231)	(0.874)				
2. School neighbourhood quintile			1.740***	0.942		
			(0.207)	(0.685)		
3. School neighbourhood quintile			2.066***	1.095*		
			(0.229)	(0.635)		
4. School neighbourhood quintile			4.353***	4.755***		
			(0.223)	(0.630)		
5. School neighbourhood quintile			11.45***	12.19***		
			(0.221)	(0.642)		
School neighbourhood wealth index					1.405***	0.427***
					(0.0559)	(0.0643)
SGB teacher share						31.58***
						(1.456)
Constant	45.56***	48.32***	45.56***	46.91***	47.79***	47.44***
	(0.167)	(0.823)	(0.140)	(0.524)	(0.227)	(0.187)
Observations	5,996	989	5,996	989	989	986
R-squared	0.342	0.323	0.334	0.401	0.390	0.588

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1;
Number of matric students used as weights (Source: DBE school data)

4.4.2. Short-run education effects

The advantage of using NIDS is that it not only provides the possibility to test long-term outcomes of quality education but also that it is representative on national level. Table 5 reports the pooled OLS regression results from NIDS wave 1-3, with years of education reached as the dependent variable for the sub-group of 15-18 year olds.[44] This age group is of particular interest since most grade repetition and school drop-outs occur in grades 9-12 (van Wyk, 2015). All regressions use population weights to adjust for attrition and clustered standard errors. In case the youth were observed when they had already left the school system, the information of the last school before the dropout has been linked to the individual.

[44] Table C6 in the Appendix shows the average household characteristics for age 15-18. Note the high percentage of fathers missing, the low mean per capita income and average years of education for parents.

Table 5: Pooled OLS regression: reached years of education by age 15-18

VARIABLES	(1) Years of education	(2) Years of education	(3) Years of education	(4) Years of education
Age	0.720***	0.708***	0.729***	0.725***
	(0.0225)	(0.0210)	(0.0208)	(0.0205)
Male		-0.520***	-0.517***	-0.519***
		(0.0790)	(0.0779)	(0.0788)
White		0.299*	-0.178	-0.327*
		(0.176)	(0.166)	(0.171)
Indian		0.848***	0.425**	0.297
		(0.214)	(0.187)	(0.190)
Coloured		0.266**	0.152	0.0839
		(0.117)	(0.108)	(0.113)
Mother's years of education			0.0659***	0.0604***
			(0.00860)	(0.00865)
Mother not in the household			-0.161**	-0.173**
			(0.0791)	(0.0786)
Father's years of education			0.0455***	0.0380***
			(0.00787)	(0.00786)
Father not in the household			-0.133**	-0.110**
			(0.0514)	(0.0500)
Ln(per capita income)				0.119***
				(0.0262)
2. School neighbourhood quintile	0.255**	0.175	0.0786	0.0745
	(0.126)	(0.111)	(0.106)	(0.106)
3. School neighbourhood quintile	0.425***	0.323***	0.149	0.139
	(0.131)	(0.101)	(0.0985)	(0.0985)
4. School neighbourhood quintile	0.618***	0.414***	0.166	0.137
	(0.125)	(0.120)	(0.113)	(0.113)
5. School neighbourhood quintile	1.029***	0.715***	0.312***	0.258**
	(0.108)	(0.115)	(0.119)	(0.121)
Constant	-3.598***	-3.173***	-3.980***	-4.524***
	(0.360)	(0.368)	(0.360)	(0.385)
Observations	7,254	7,254	7,247	7,245
R-squared	0.267	0.314	0.357	0.360

Not reported province dummies. Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
(Source NIDS wave 1-3)

Column (1) of Table 5 shows the simple model with just age and school neighbourhood quintiles as explanatory variables. In this specification children going to the richest school quintile 5 have attained roughly one more year of education than the children from the poorest schools at the same age. As we start controlling for race and gender in column (2), parental education in column (3) and household income in column (4) the coefficient drops to about one-quarter but remains sizable and statistically significant. There is a high correlation between parental education, household income and school wealth quintiles.

If we estimate the same specifications but replacing the current school quintiles with the DBE school quintiles (see Table C7 in the Appendix), the coefficients for school quintiles dummies become insignificant and negligible in magnitude once we control for parental education and family income. Hence, it seems that the revised school wealth quintiles provide a more informative measure of school quality than the official DBE school quintile measure.

One obvious issue with using years of education obtained is that this does not necessarily imply actual learning but merely years of schooling completed. The true learning gap between the poorest and richest schools may be even larger.[45]

It is also noteworthy that there appears to be a large and significant gender effect, causing boys to obtain about half a year less education at the same age as girls. Interestingly, there seem to be no remaining substantial race effect after controlling for household income, parental education and school characteristics. This means black children have the same grade progression if they have similar socio-economic backgrounds and attend the same schools as children from the other races. Yet, the average black child has a mother and father with about 7 years of education and lives in a household with a mean per capita income of about R900, whereas the average white child has parents with about 13 years of education and R7000 mean household income. Taking all of this into consideration, an average black child has reached about 1.5 years less education at the same age than the average white child. In addition, the outcome variable only measures school attainment despite differences in the quality of education the average black and white child receives. The finding that South Africa essentially has two very different education systems – one for the poor black (and coloured) students going to formerly disadvantaged schools and one for well-off students – that produce vastly different schooling results, confirms what have been shown by many previous studies using different data sources, e.g. Van der Berg (2007, 2008), Reddy (2011) and Spaull and Kotze (2015). Yet, it remains to be persuasively determined how important the neighbourhood effect is relative to school quality effect.

[45] Estimating the same model on an administrative data set for learners from Cape Town and using Western Cape Systemic Tests of learner performance shows that the learning gap between the poorest and richest children, going to the best and worst schools of Cape Town, are about 1.7 standard deviations which translates to approximately 5 years' worth of learning. This demonstrates that the school neighbourhood effect is perhaps even more important in determining learning than its effect on schooling attainment. This data is not yet in the public domain and hence these regressions were omitted from this thesis.

Table 6: Pooled OLS and cluster FE regressions: reached years of education by age 15-18

	(1) Full sample	(2) Full sample	(3) Urban sample	(4) Urban sample	(5) Rural sample
VARIABLES	**OLS**	**Cluster FE**	**Cluster FE**	**Cluster FE**	**Cluster FE**
Age	0.705***	0.699***	0.747***	0.748***	0.651***
	(0.0223)	(0.0227)	(0.0349)	(0.0352)	(0.0293)
Male	-0.493***	-0.487***	-0.270**	-0.268**	-0.727***
	(0.0767)	(0.0806)	(0.108)	(0.109)	(0.0790)
White	-0.403**	0.0253	-0.0220	-0.0246	-0.654
	(0.177)	(0.249)	(0.262)	(0.261)	(0.809)
Indian	0.218	0.252	0.124	0.144	0.895***
	(0.195)	(0.253)	(0.300)	(0.311)	(0.168)
Coloured	0.154	0.356	0.389	0.383	-1.158
	(0.111)	(0.267)	(0.260)	(0.260)	(0.809)
Ln (School distance)	0.0374*	0.0684**	0.0217	0.0144	0.148***
	(0.0226)	(0.0272)	(0.0388)	(0.0380)	(0.0345)
Ln(per capita income)	0.107***	0.0784**	0.0811*	0.0810*	0.0895**
	(0.0318)	(0.0306)	(0.0479)	(0.0480)	(0.0422)
Mother's years of education	0.0628***	0.0586***	0.0533***	0.0545***	0.0649***
	(0.00934)	(0.00987)	(0.0189)	(0.0197)	(0.0105)
Mother not in the household	-0.150**	-0.132*	-0.371***	-0.375***	0.0563
	(0.0743)	(0.0732)	(0.130)	(0.130)	(0.0791)
Father's years of education	0.0366***	0.0293***	0.0307**	0.0298**	0.0303***
	(0.00853)	(0.00801)	(0.0146)	(0.0149)	(0.00983)
Father not in the household	-0.110**	-0.0965**	-0.0236	-0.0254	-0.161***
	(0.0486)	(0.0487)	(0.0812)	(0.0799)	(0.0540)
Household neighbourhood wealth quintile	YES	YES	YES	YES	YES
2. School neighbourhood quintile	0.0810	-0.117	0.222		
	(0.113)	(0.134)	(0.222)		
3. School neighbourhood quintile	0.158	0.219	0.527**		
	(0.112)	(0.168)	(0.237)		
4. School neighbourhood quintile	0.127	0.189	0.544**		
	(0.125)	(0.171)	(0.222)		
5. School neighbourhood quintile	0.277*	0.354**	0.689***		
	(0.141)	(0.169)	(0.205)		
Difference school to household quintiles				0.155***	-0.0681
				(0.0519)	(0.0753)
Constant	-4.213***	-4.400***	-5.512***	-5.425***	-2.913***
	(0.396)	(0.468)	(0.606)	(0.629)	(0.974)
Observations	7,131	7,131	2,716	2,714	4,415
R-squared	0.354	0.453	0.515	0.514	0.405

Standard erros in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Not reported province dummies.
(Source NIDS wave 1-3)

Column (1) of Table 6 controls for the neighbourhood wealth quintile and the distance children travel to school. While the neighbourhood the child lives in does not seem to have an additional significant effect, the distance to school is a positive and significant determinant of schooling attainment. This could be either because parents' who send their children to schools further away are also more likely to support their offspring with their education in other ways, or because parents are aware of differences in school quality that are unobservable to the econometrician.

To control for other sources of unobserved heterogeneity between neighbourhoods, column (2)-(5) adds cluster fixed effects to the model. In column (2) the coefficient for the highest school quintile does increase to 0.35 from 0.28 before. The same specification is estimated for the urban sample of NIDS youth in column (3). As seen before in Table 5, the school wealth quintile variable seem to explain education outcomes particularly well in South African cities. This could be because the difference in schools based in former white and Indian neighbourhoods is much larger in comparison to the township schools of former black and coloured neighbourhoods within the same city. However, the significant positive coefficient for the "difference in school to household quintile" in column (4) implies that students living in the poorest quintile neighbourhood but going to a quintile 5 wealth school, would reach about 0.6 year more education than a students from the same neighbourhood who are going to the poorest 20% of schools.

For the rural sample in column (5) no positive effect for such movement between neighbourhoods can be observed. Yet, here there appears to be large gains to attending schools further away from home. Finally, in Table C8 in the appendix, some additional school quality variables like the share of SGB-teachers at a school, dummies for private-or mixed funded schools or the student-teacher ratio are added to the model. The coefficient of the school wealth quintile variable stays robustly positive and significant.

4.4.3. Long-run socio-economic effects

We now consider the effect of quality schooling on university enrolment and labour market earnings. In Table 7 university enrolment is regressed on race, age and household characteristics as well as school and household neighbourhood quintiles for the age group 18-23. The counterfactual group for those enrolled at university are youth directly starting to work after school / unemployed or economic not-active. The NIDS panel element was used, to get the actual information from the last school attended and the neighbourhood the youth was living at whilst still going to school. In this model, household wealth and coming from the richest neighbourhood quintile seem to substantially increase the chances of a youth enrolling in university. This finding suggests some kind of credit constraints for poorer students going to university, which is not surprising given the high direct costs of university fees. There also seem to be a significantly positive effect from the school quality measure, particularly for those whose last school was from the highest quintile.

Black youth seem to be more likely to start some form of tertiary education after controlling for social-economic status. Limiting the sample to urban black youth, the advantage for the richest quintile remains stable, while living in the richest neighbourhoods becomes even more important for black youth.

Table 7: Pooled OLS regression: youth enrolled in university age 18-23

	(1)	(2)	(3)	(4)	(5)
	Full sample	Full sample	Full sample	Urban sample	Black sample
VARIABLES	University enrolment	University enrolment	University enrolment	University enrolment	University enrolment
Age	0.0234***	0.0267***	0.0292***	0.0438***	0.0273***
	(0.00584)	(0.00551)	(0.00583)	(0.00875)	(0.00608)
Male	-0.0463***	-0.0553***	-0.0652***	-0.0697***	-0.0680***
	(0.0135)	(0.0128)	(0.0136)	(0.0202)	(0.0150)
White	0.293***	-0.0333	-0.189***	-0.205***	
	(0.0301)	(0.0326)	(0.0388)	(0.0464)	
Indian	-0.0223	-0.283***	-0.400***	-0.480***	
	(0.0495)	(0.0485)	(0.0506)	(0.0707)	
Coloured	-0.0385	-0.140***	-0.197***	-0.210***	
	(0.0285)	(0.0274)	(0.0299)	(0.0357)	
Ln(per capita income)		0.0928***	0.0831***	0.114***	0.0672***
		(0.00729)	(0.00802)	(0.0124)	(0.00873)
Mother's years of education		0.0147***	0.0134***	0.0165***	0.0112***
		(0.00189)	(0.00205)	(0.00326)	(0.00217)
Father's years of education		0.00962***	0.00701***	0.00226	0.00629***
		(0.00203)	(0.00220)	(0.00344)	(0.00231)
2.Ex School neighbourhood quintile			-0.0114	0.0740*	-0.00981
			(0.0214)	(0.0431)	(0.0217)
3. Ex School neighbourhood quintile			-0.0101	-0.0179	-0.0220
			(0.0257)	(0.0444)	(0.0266)
4. Ex School neighbourhood quintile			0.0403	0.0437	0.0232
			(0.0263)	(0.0436)	(0.0286)
5. Ex School neighbourhood quintile			0.123***	0.115***	0.112***
			(0.0270)	(0.0428)	(0.0293)
2.Ex household neighbourhood quintile			-0.0392*	-0.0264	-0.0300
			(0.0208)	(0.0388)	(0.0213)
3. Ex household neighbourhood quintile			-0.0367	0.0103	-0.0209
			(0.0263)	(0.0428)	(0.0276)
4. Ex household neighbourhood quintile			-0.0391	-2.43e-05	-0.00982
			(0.0267)	(0.0424)	(0.0293)
5. Ex household neighbourhood quintile			0.124***	0.121**	0.209***
			(0.0351)	(0.0503)	(0.0425)
Constant	-0.358***	-1.164***	-1.076***	-1.575***	-0.959***
	(0.122)	(0.123)	(0.130)	(0.204)	(0.139)
Observations	3,443	3,440	3,095	1,539	2,602
R-squared	0.068	0.176	0.202	0.263	0.159

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1; (Source: NIDS wave 1-4)

To identify the effect of school quality on earnings, we estimate several earnings functions using OLS in Table 8. The dependent variable is log earnings and the control variables include gender, race and age[46]. The variables of interest are years of education, years of education squared and the wealth index of the school the young worker went to as a measure of school quality. The information for the school were taken from the retrospective question of NIDS that asked which school a respondent last attended. Only the subgroup

[46] Even though mincer wage regressions normally assume non-linear returns to age, for the small age-period at hand, the assumption of linearity for age seems to be sufficient.

of workers aged 20-30 were chosen, since there are concerns around the quality of the merge on school names for older cohorts. In addition, since the school quintiles seem to explain the variation in education outcomes best in urban settings, the analysis was further limited to this subgroup.

To quantify the impact of education quality on earnings, it is important to distinguish between the direct and indirect effects: the indirect effect allows a student from a better school to reach more years of education (as seen in Table 4, 5 and 6), whereas the direct effect is the benefit of after controlling for years of education. The channel this would have to work through is higher ability, better writing, math or other skills that can be observed by their employer. Lastly there is the potential problem of unobserved household heterogeneity, since children that went to richer schools might also have higher ability, richer parents and other unobserved factors that are financially remunerated in the labour market. In an attempt to control for this household effects, mother and father education are included in the earnings model.

Column (1) shows that school wealth does have a positive and significant effect on earnings, indicating some direct positive effect of quality schooling on earnings. Returns to education seem to be convex given the significant and negative coefficient for education and positive and significant coefficient for education squared. There seem to be a wage premium for being male, Indian[47] and white. Next to observe if there is a premium for each year higher quality education received, an interaction term between the school wealth index and years of education reached is entered in column (2). Entering this interaction term, the coefficient of education squared marginally decreases and the coefficient for the school wealth index turns negative. Given the significant and positive interaction term this signals that there is a wage premium for higher quality education only when a certain combination of quality and years of education is reached. This is best illustrated in a graph as shown in Figure 3. The graph shows the returns of education for low, middle and high quality schools. For all schools the returns to education turn positive around 8 years.[48] The more years a student reached the higher the premium for quality education they received. That means that having matric from a high quality school would increase earnings by about 50 percentage points on top of the normal returns to education.

[47] The particularly large coefficient for the Indian dummy might be explained by the small number of Indian youth in the sample.
[48] Given that only 4% of workers in our sample have fewer than 8 years of schooling, the model fit at these lower schooling values is not of great practical importance.

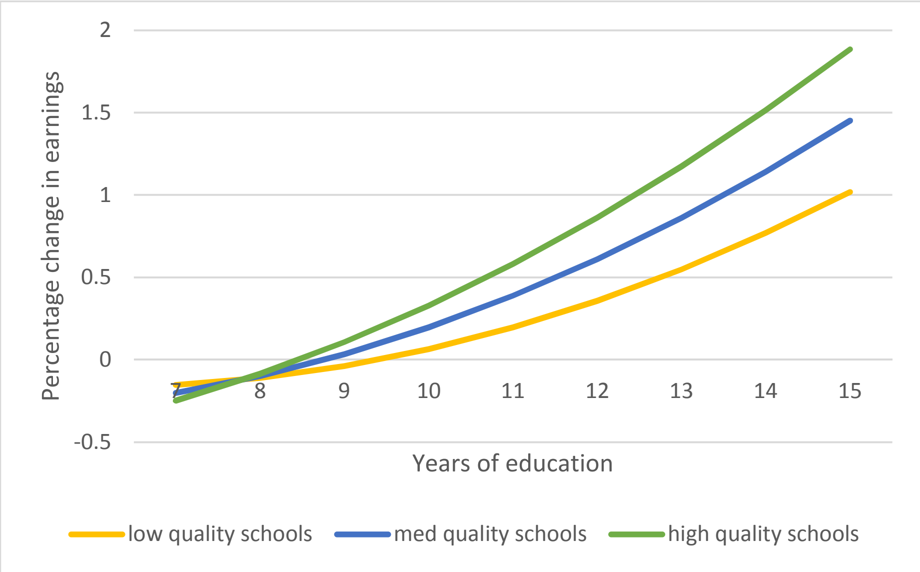
Table 8: Pooled OLS regression: ln(earnings) of age 21-30 in urban sample

	(1) Full sample	(2) Full sample	(3) Black sample	(4) Black sample
VARIABLES	**Ln(earnings)**	**Ln(earnings)**	**Ln(earnings)**	**Ln(earnings)**
White	0.286***	0.277***		
	(0.104)	(0.104)		
Coloured	0.0648	0.0711		
	(0.0541)	(0.0541)		
Indian	0.652***	0.623***		
	(0.162)	(0.163)		
Age	0.0575***	0.0579***	0.0564***	0.0568***
	(0.00647)	(0.00647)	(0.00757)	(0.00755)
Male	0.264***	0.265***	0.257***	0.259***
	(0.0323)	(0.0323)	(0.0376)	(0.0375)
Mother having matric	0.185***	0.175***	0.234***	0.216***
	(0.0530)	(0.0531)	(0.0616)	(0.0616)
Father having matric	0.173***	0.166***	0.168**	0.165**
	(0.0554)	(0.0554)	(0.0673)	(0.0671)
Education	-0.143***	-0.131***	-0.166***	-0.151***
	(0.0390)	(0.0393)	(0.0439)	(0.0440)
Education^2	0.0157***	0.0148***	0.0169***	0.0161***
	(0.00186)	(0.00190)	(0.00213)	(0.00214)
Std. (school index)	0.0429**	-0.187**	0.0616***	-0.394***
	(0.0181)	(0.0951)	(0.0203)	(0.122)
Std. (school index)* education		0.0204**		0.0403***
		(0.00827)		(0.0106)
Constant	5.318***	5.277***	5.489***	5.373***
	(0.277)	(0.277)	(0.314)	(0.315)
Observations	3,003	3,003	2,254	2,254
R-squared	0.254	0.256	0.236	0.241

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. (Source: NIDS wave 1-4)

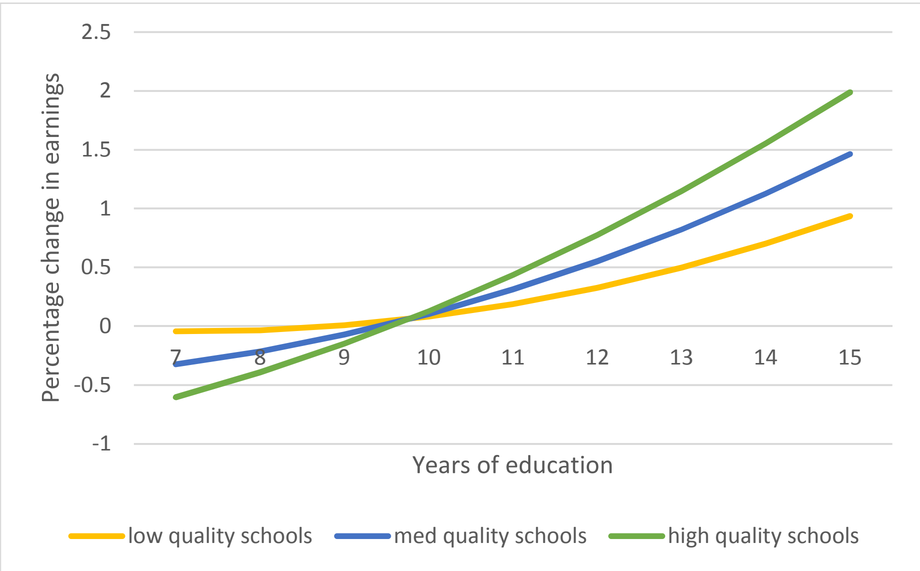
As a robustness check, the same set of specifications are also run only on the black sample. It appears that for black youth in urban areas the direct effect of going to a high quality school is even larger than the overall average. This is demonstrated in Figure 4 that has even larger returns to high levels of education for black workers that go to the highest quality schools.

Figure 3: Returns to education given different school quality



(Source: own calculation from earnings regression seen in Table 8)

Figure 4: Returns to education given different school quality for black youth



(Source: own calculation from earnings regression seen in Table 8)

4.5. Conclusion

This paper evaluated the relative importance of family, neighbourhood and school quality in explaining the variation in school and labour market outcomes for South Africa. To do so, the Census 2011 community data was used to develop a new wealth index for about 85,000 neighbourhoods that can be linked to school and student location from the National Income Dynamics Study. Revised school wealth quintiles were estimated. They prove to be good measures for school quality and are more accurate than the previously used school quintiles in explaining schooling outcomes, in particular for urban areas.

The results of this study have shown that children going to the richest 20% of schools in urban areas have on average roughly 0.6 more years of education than children going to the poorest 20% of schools. A student living in the poorest quintile neighbourhood who attends the richest school in town would perform significantly better than a student from the same neighbourhood who is going to the poorest school. Sadly, due to financial and transport constraints only 10% of children living in the poorest 60% neighbourhoods are manging to go to the richest school quintile. Hence, the study provides further evidence of how segregated the South African school system is: providing quality education to a few and leaving behind a large share of poor children. It highlights the importance bringing high quality schooling into the townships and rural areas where a majority of the poor live. This chapter is amongst the first studies to provide evidence of the importance of quality education in explaining who is enrolling in university (after reaching matric) and determining subsequent earnings. Both of these outcomes are of great concern for decreasing inequality in South Africa. From the earnings model it can be concluded that there are increasing returns to education for going to a high quality school. This means that children receive a 50% premium for receiving matric from a high quality school in comparison to a child with matric from a low quality school.

This study has shown that the revised school wealth quintiles are an important means of identifying schools that need special attention, since it predicts individual learner and school outcomes more accurately than the official school quintiles. Such an instrument could also be used by the department of education to target poor schools and reform a funding system to achieve more equal school outcomes in South Africa in the long run. Moving children from the poorest schools to the richest quintile schools have shown large significant effects. Hence, finding mechanisms to ensure that all children have access to high quality schools – whether through vouchers, bussing or massively improving the schools in the townships and rural areas – is an important step towards equalising education opportunity in South Africa.

Chapter 5:

Conclusion

Viewed broadly, this thesis aims to measure economic and social mobility in South Africa. It has been shown that various problems with survey responses can produce misleading impression of the South African labour market and of income dynamics. This highlights the importance of measuring variables of interest accurately and to carefully consider the ways in which unreliable responses can bias the results of conventional estimators. It was demonstrated that even the most appropriate estimator and identification strategy can fail to yield unbiased estimates if important measurement issues are ignored. To address these shortcomings new approaches remedy well-known survey data reliability concerns have been applied and adapted in this thesis. In the following, the most important findings of the three chapters are outlined and based on these findings, we proceed to discuss a few policy recommendations and further needed research.

5.1. Findings from chapter 2

Given the scope and scale of the youth unemployment problem in South Africa, estimating labour market behaviour and understanding the school to work transition is of the utmost importance. Previous studies that have looked at the direct effect of reservation wages on employment probability haven't found any adverse effects. However, answers to the traditional question on reservation wages may fail to provide meaningful answers. Chapter 2 has revisited this issue by comparing the answers to the traditional reservation wage question to a series of questions intended to elicit a more accurate response to questions asking the respondent about the lowest wage for which he or she would work. The results presented in chapter 2 have shown that the survey design and wording of questions matter for reliable reservation wage measurement. In line with findings in psychology and behavioural economics, we find that different formulations and ordering of the reservation wage question can trigger different cognitive processes in the respondent that elicit different answers. Interestingly, youth that have little or no job experience and are not attached to the labour market appear to be the most sensitive to such framing and priming effects. This is expected given that they have had little labour market exposure and experience to shape their expectations and the observed fluctuations may thus be due to a greater deal of uncertainty about their own perceptions.

On the other hand, the new probed reservation wage measure seems to be more internally consistent and the regression results to be in line with labour market search models. That is, using the probed measure we find significant positive effects of transportation costs and household wealth, as well as household income on reservation wages. Having a unique panel of young job-seekers we can also show a significant negative correlation with unemployment duration. This result is in line with the international literature and it is a novel finding for South Africa. In conclusion, this chapter has shown that in the context of high unemployment and weak labour market attachment which characterise the reality for many South African youth, approaches commonly used in developed countries cannot be applied in a high unemployment context without more research, including experimenting with a variety of question formulations that trigger distinct cognitive processes to gauge whether these provide consistent responses.

5.2. Findings from chapter 3

Upward economic mobility continues to be one of the most important indicators for economic development and also the equity of a society. However, the results of this chapter have shown that estimating the speed of convergence between the poorest and richest households using micro growth regressions without controlling for measurement error would overestimate income mobility significantly. Therefore, a newly developed GMM estimator was applied to four large national panel studies to obtain less biased β estimates. The findings of four large representative national panel studies from the USA, South Africa, Chile and Tanzania show that naïve OLS regression coefficients would overestimate the extent of income mobility by a factor of about 4-6. Translating the GMM β coefficient for the US into the expected half-life of income gaps shows that the estimates are in line with the intergenerational literature which are less vulnerable to measurement error. In addition, the J-statistic indicates that the GMM estimator can calculate all linearly independent regression coefficients in a way that is internally highly consistent. The results provide support for the assumption that the GMM estimator produces realistic estimates for the speed of convergence within a country. The hypothesis of no measurement error can be rejected for all the countries observed and the α estimates imply that the share of income variation that is due to actual income changes is only 55% for Tanzania, 63% for Chile, 79% for the USA and 80% for South Africa. Therefore, while the data reliability for the US, Chile and Tanzania correspond to their levels of economic development, South Africa's data reliability appears to be unexpectedly high. The nonparametric estimates also show that the speed of convergence varies over the income distribution and that income is more reliably captured for richer than for poorer households. This research cautions against the overestimation of income mobility when not accounting for measurement error, which is relevant for all concerned with reliably capturing the extent of upward mobility including policy makers.

5.3. Findings from chapter 4

Chapter 4 tried to evaluate the relative importance of family, neighbourhood and school quality in explaining socio-economic outcomes for South Africa. While previous studies in the US and Northern Europe, mostly found only small neighbourhood or school effects, location and schooling quality should matter more in South Africa, where movement was restricted and resources were distributed unevenly between the different school systems of the past. Using spatial merging techniques to combine different data sets, new school wealth quintiles have been created that predict individual learner and school outcomes more accurately than the old school quintiles. The results of this chapter have shown how children going to the richest 20% of schools reach significant more years of education than children going to the poorest 20% of schools. However, this study also demonstrates that even children from the poorest neighbourhood would perform well if they go to one of the richest 20% of schools. Yet, given the limited number of quality schools, the segregated location of quality school, financial as well as transport constraints, only about 10% of children the poorest 60% actually attend a top quintile schools. In order to achieve more equal and more acceptable education outcomes, the quality of schools in the poor neighbourhoods would need to be drastically improved. The thesis provides evidence of the importance of quality education in explaining university enrolments. In addition, there seem to be a significant premium for quality education in labour markets earnings regressions, which show that schooling has an enduring and long-lasting impact. In general, finding mechanisms to ensure that poor children have access to high quality schools – whether through vouchers, bussing or massively improving the township schools – is a crucial step towards equalising education opportunity and creating a post-apartheid society that is more open and fairer.

5.4. Final conclusion

The overall findings and implications of this thesis are manifold. New available panel data makes it possible to answer questions on economic and social mobility in South Africa. Many studies using panel estimators have been published in recent years. However, these new panel estimators are vulnerable to measurement error: Hence new techniques to control for, and get unbiased results, are important. I.e. Chapter 1 shows that the formulation of survey questions can be instrumental in getting unbiased results, in particular in the setting of week labour market-attachment. In future research, an experimental study could be designed to estimate the magnitude of the anchoring and miss-reporting in survey responses. The second Chapter shows that income mobility estimates using panel estimates are highly sensitive to measurement error. In a next step, consequences for other indicators like poverty dynamics, can be observed to show the relevance of these results. The final question is: What have we learnt from this thesis about social and income mobility in South Africa? Given the current developments in South Africa, this is a very relevant question and its answer has important policy relevance. As chapter 2 shows, income convergence in South Africa is slower than in other developing countries. The time half the income gaps between any two households in the country should be eliminated within 23 years, or equivalent to another generation. This thesis can identify at least two reasons for this slow convergence: Firstly, there is the severe youth unemployment problem resulting from a mismatch between a shortage in qualified job entries and the demand for high paying jobs. In this situation, financial household support or and unreasonably high wage expectations due to week labour market attachment, might further increase the unemployment duration of the youth. Secondly, as demonstrated in chapter 3, South Africa has a highly unequal school system, of which the outcomes are still very much dependent on birthplace and parental background. Hence, to improve social mobility and decrease income inequality, improving the quality of schools in poor neighbourhoods and regions is highly recommended. Other options might be vocational training or other forms of skills transfer to improve the chances of youths on the labour market, especially those from poor backgrounds. The current major student protests about issues regarding free quality education may offer an opportunity to reach the public and convince the government to make the necessary transformation.

Bibliography

AGUIRREGABIRIA, V., MIRA, P., (2010). Dynamic Discrete Choice Structural Models: A Survey. *Journal of Econometric,* 156(1), 38-67.

Abel, M. (2013). Unintended labour supply effects of cash transfer programs: Evidence from South Africa's old age pension. SALDRU working paper Number 114, UCT.

Abowd, J. M., & Stinson, M. H. (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *Review of Economics and Statistics*, 95(5), 1451-1467.

Addison, J. T., Centeno, M., & Portugal, P. (2004). Reservation wages, search duration, and accepted wages in Europe. IZA Discussion Papers 350, Institute for the Study of Labor (IZA).

Addison, J. T., Centeno, M., & Portugal, P. (2009). Do reservation wages really decline? Some international evidence on the determinants of reservation wages. *Journal of labor research*, 30(1), 1-8.

Addison, J. T., Centeno, M., & Portugal, P. (2010). Unemployment Benefits and Reservation Wages: Key Elasticities from a Stripped-Down Job Search Approach. *Economica*, 77(305), 46-59.

Addison, J., Machado, J., & Portugal, P. (2013). The Reservation Wage Unemployment Duration Nexus. *Oxford Bulletin of Economics and Statistics*, 75(6), 980–987.

Agüero, J., Carter, M. R., & May, J. (2007). Poverty and inequality in the first decade of South Africa's democracy: What can be learnt from panel data from KwaZulu-Natal? *Journal of African Economies*, 16(5), 782-812.

Akee, R. (2011). Errors in self-reported earnings: The role of previous earnings volatility and individual characteristics. *Journal of Development Economics*, 96(2), 409-421.

Allanson, P., Atkins, J. P., & Hinks, T. (2002). No end to the racial wage hierarchy in South Africa?. *Review of Development Economics*, 6(3), 442-459.

Altonji, J. G., & Mansfield, R. K. (2011). The role of family, school, and community characteristics in inequality in education and labor market outcomes. *Whither opportunity*, 339-358.

Antman, F., & McKenzie, D. (2007a). Poverty traps and nonlinear income dynamics with measurement error and individual heterogeneity. *The journal of development studies*, 43(6), 1057-1083.

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

Ardington, C., Bärnighausen, T., Case, A., & Menendez, A. (2016). Social protection and labor market outcomes of youth in South Africa. *ILR Review*, 69(2), 455-470.

Banerjee, A., Galiani, S., Levinsohn, J., McLaren, Z., & Woolard, I. (2008). Why has unemployment risen in the new South Africa? 1. *Economics of Transition*, 16(4), 715-740.

Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy*, 223-251.

Barro, R. J., & Sala-i-Martin, X. (2004). Economic Growth (2nd ed.). New York: McGraw-Hill.

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

Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? Implications for subjective survey data. *American Economic Review*, 67-72.

Bertrand, M., Mullainathan, S., & Miller, D. (2003). Public policy and extended families: Evidence from pensions in South Africa. *The World Bank Economic Review*, 17(1), 27-50.

Bingley, P., & Martinello, A. (2014). Measurement error in income and schooling, and the bias of linear estimators.

Björklund A, Jäntti M, Raaum O, Österbacka E, Eriksson T (2004). Family structure and labor market success: the influence of siblings and birth order on the earnings of young adults in Norway, Finland, and Sweden. In: Corak M (ed) Generational mobility in North America and Europe. Cambridge University Press, Cambridge, pp 208–225.

Bloemen, H. G., & Stancanelli, E. G. (2001). Individual wealth, reservation wages, and transitions into employment. *Journal of Labor Economics*, 19(2), 400-439.

Böheim, R. (2002). The association between reported and calculated reservation wages. ISER Working Paper Series, No. 2002-07.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement Error in Survey Data. In J. J. Heckman & E. Leamer (Eds.), Handbook of Econometrics, Vol 5. New York: Elsevier Science B.V.

Bound, J., & Krueger, A. B. (1991). The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right? *Journal of Labor Economics*, 1-24.

Booysen, F., van der Berg, S., Burger, R., Maltitz, M. V., & Rand, G. D. (2008). Using an asset index to assess trends in poverty in seven Sub-Saharan African countries. *World Development*, 36(6), 1113-1130.

Branson, N., Hofmeyr, C., & Lam, D. (2014). Progress through school and the determinants of school dropout in South Africa. *Development Southern Africa*, 31(1), 106-126.

Brown, M., Flinn, C.J., & Schotter, A. (2011). Real-Time Search in the Laboratory and the Market. *American Economic Review*, 101, 948–974.

Brown, S., & Taylor, K. (2011). Reservation wages, market wages and unemployment: Analysis of individual level panel data. *Economic Modelling*, 28(3), 1317-1327.

Brown, S., & Taylor, K. (2013). Reservation wages, expected wages and unemployment. *Economics Letters*, 119(3), 276-279.

Branson, N., & Wittenberg, M. (2007). The measurement of employment status in South Africa using cohort analysis, 1994‐2004. *South African Journal of Economics*, 75(2), 313-326.

Burdett, K., & Vishwanath, T. (1988). Declining Reservation Wages and Learning. *Review of Economic Studies*, 55, 655–665.

Burger, R., van der Berg, S., & Fintel, D. (2015). The unintended consequences of education policies on South African participation and unemployment. *South African Journal of Economics*, 83(1), 74-100.

Burger, R., Steenekamp, C. L., van der Berg, S., & Zoch, A. (2015). The emergent middle class in contemporary South Africa: Examining and comparing rival approaches. *Development Southern Africa*, 32(1), 25-40.

Card, D., & Krueger, A. (1990). Does school quality matter? Returns to education and the characteristics of public schools in the United States (No. w3358). National Bureau of Economic Research.

Case, A., & Deaton, A. (1999). School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics,* 1047-1084.

Case, A., & Yogo, M. (1999). Does school quality matter? Returns to education and the characteristics of schools in South Africa (No. w7399). National Bureau of Economic Research.

Cruces, G., Fields, G., & Viollaz, M. (2013). Can the limitations of Panel datasets be overcome by using Pseudo-panels to estimate income mobility. Technical report, CEDLAS, Universidad Nacional de La Plata.

Cuesta, J., Ñopo, H., & Pizzolitto, G. (2011). Using Pseudo‐Panels To Measure Income Mobility In Latin America. *Review of Income and Wealth*, 57(2), 224-246.

De Kadt, J., Norris, S. A., Fleisch, B., Richter, L., & Alvanides, S. (2014). Children's daily travel to school in Johannesburg-Soweto, South Africa: geography and school choice in the Birth to Twenty cohort study. *Children's Geographies*, 12(2), 170-188.

Dercon, S. & Krishnan, P. 2000. Vulnerability, Seasonality and Poverty in Ethiopia. *The Journal of Development Studies*, 36(6): 25-53.

Du Rand, G., van Broekhuizen, H., & von Fintel, D. (2011). The Role of Levels of Numeric Competence and School Quality in the South African Wage Function.

Duflo, E. (2003). Grandmothers and Granddaughters: Old‐Age Pensions and Intrahousehold Allocation in South Africa. *The World Bank Economic Review*, 17(1), 1-25.

Eriksson T, & Zhang, Y. (2012). The role of family background for earnings in rural China. *Front Econ China*, 7(3):465–477.

Falk, A., Fehr, E., & Zehnder, C. (2006). Fairness perceptions and reservation wages—the behavioral effects of minimum wage laws. *The Quarterly Journal of Economics*, 121(4), 1347-1381.

Feldstein, M., & Poterba, J. (1984). Unemployment insurance and reservation wages. *Journal of Public Economics*, 23(1), 141-167.

Fields, G. S. (2008a). A brief review of the literature on earnings mobility in developing countries ILR Working Paper 2-20-2008. Ithaca: Cornell University.

Fields, G. S. (2008b). Income mobility. In L. Blume & S. N. Durlauf (Eds.), The new Palgrave dictionary of economics. New York, NY: Palgrave Macmillan.

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

Fields, G. S., Cichello, P. L., Freije, S., Menéndez, M., & Newhouse, D. (2003b). Household income dynamics: a four-country story. *The journal of development studies*, 40(2), 30-54.

Fields, G. S., Duval-Hernández, R., Freije, S., & Puerta, M. L. S. (2014). Earnings mobility, inequality, and economic growth in Argentina, Mexico, and Venezuela. *The Journal of Economic Inequality*, 1-26.

Fields, G. S., & Viollaz, M. (2013). Can the Limitations of Panel Datasets be Overcome by Using Pseudo-Panels to Estimate Income Mobility? Universidad Cornell-CEDLAS.

Finn, A., Leibbrandt, M., & Levinsohn, J. (2012). Income mobility in South Africa: Evidence from the first two waves of the National Income Dynamics Study. Southern Africa Labour and Development Research Unit, University of Cape Town.

Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. Handbook of the Economics of Education, 2, 945-1017.

Glewwe, P. (2012). How much of observed economic mobility is measurement error? IV methods to reduce measurement error bias, with an application to Vietnam. *The World Bank Economic Review*, 26(2), 236-264.

Gottschalk, P., & Huynh, M. (2010). Are earnings inequality and mobility overstated? The impact of nonclassical measurement error. *The Review of Economics and Statistics*, 92(2), 302-315.

Hanushek, E. A., Lavy, V., & Hitomi, K. (2008). Do students care about school quality? Developing countries. *Journal of Human Capital*, 2(1),69–105.

Harbison, R. W., & Hanushek, E. A. (1992). Educational performance of the poor: Lessons from rural northeast Brazil. Oxford University Press.

Hill, L. D. (2016). Race, school choice and transfers to opportunity: implications for educational stratification in South Africa. *British Journal of Sociology of Education*, 37(4), 520-547.

Hofler, R. A., & Murphy, K. J. (1994). Estimating reservation wages of employed workers using a stochastic frontier. *Southern Economic Journal*, 961-976.

Holzer, H. J. (1986a). Reservation wages and their labor market effects for black and white male youth. *Journal of Human Resources*, 157-177.

Holzer, H. J. (1986b). Black youth nonemployment: Duration and job search. In The black youth employment crisis (pp. 23-73). University of Chicago Press.

Hunter, M. (2015). Schooling choice in South Africa: The limits of qualifications and the politics of race, class and symbolic power. *International Journal of Educational Development*, 43, 41-50.

Jäntti, M., & Jenkins, S. P. (2015). Chapter 10 - Income Mobility. In B. A. Anthony & B. François (Eds.), Handbook of Income Distribution (Vol. Volume 2, pp. 807-935): Elsevier.

Jarvis, S., & Jenkins, S. (1998). How much income mobility is there in Britain? *The Economic Journal*, 108(447), 428-443.

Jensen, P., & Westergård-Nielsen, N. C. (1987). A search model applied to the transition from education to work. *The Review of Economic Studies*, 54(3), 461-472.

Jones, Stephen R G, (1988). The Relationship between Unemployment Spells and Reservation Wages as a Test of Search Theory. *The Quarterly Journal of Economics*, MIT Press, 103(4), pp. 741-65.

Kingdon, G., & Knight, J. (2001). What have we learnt about unemployment from microdatasets in South Africa?. *Social Dynamics*, 27(1), 79-95.

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

Klasen, S., & Woolard, I. (2009). Surviving unemployment without state support: unemployment and household formation in South Africa. *Journal of African economies*, 18(1), 1-51.

Kriechel, B., & Pfann, G. A. (2006). Learning to update your reservation wage while looking for a new job. *Portuguese Economic Journal*, 5(2), 135-148.

Krueger, A., & Mueller, A. (2014). A Contribution to the Empirics of Reservation Wages (No. w19870). National Bureau of Economic Research.

Lam, D., Leibbrandt, M. & Mlatsheni, C., (2008). Education and Youth Unemployment. SALDRU Working Paper, UCT.

Lam, D., Leibbrandt, M., & Mlatsheni, C. (2010). Human capital, job search, and unemployment among young people in South Africa. Paper presented at 2010 meeting of Population Association of America, Dallas, TX.

Lam, D., Ardington, C., & Leibbrandt, M. (2011). Schooling as a lottery: Racial differences in school advancement in urban South Africa. *Journal of Development Economics*, 95(2), 121-136.

Lam, D., Ardington, C., Branson, N., Case, A., Leibbrandt, M., Maughan-Brown, B., Menendez, A., Seekings, J., Sparks, M (2013). The Cape Area Panel Study - Overview and Technical Documentation Waves 1-2-3-4-5 (2002-2009). UCT.

Lancaster, T. (1985). Simultaneous equations models in applied search theory, *Journal of Econometrics*, 28, 113–126.

Lancaster, T., & Chesher, A. (1983). An econometric analysis of reservation wages. *Econometrica: Journal of the Econometric Society*, 1661-1676.

Lechtenfeld, T., & Zoch, A. (2014). Income Convergence in South Africa: Fact or Measurement Error?

Lee, N. (2009). Measurement error and its impact on estimates of income and consumption dynamics. Available at SSRN 1299330.

Leibbrandt, M., Lilenstein, K., Shenker, C., & Woolard, I. (2013). The influence of social transfers on labour supply: A South African and international review. SALDRU working paper 2013-10.

Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income—evidence from Sweden. *The Journal of Economic Inequality*, 9(2), 207-226.

Msila, V. (2005). The education exodus: The flight from township schools. *Africa Education Review*, 2 (2): 173–188.

Mortensen, D. (1976). Unemployment insurance and job search decisions. *Indus. & Lab. Rel. Rev.*, 30, 505.

Mortensen, D. (1986) Job Search and Labor Market Analysis, in O. Ashenfelter and R. Layard (eds.) Handbook of Labor Economics, Volume 2, North-Holland.

Moylan, S., Millar, J., & Davies, R. (1984). For richer, for poorer?: DHSS cohort study of unemployed men. HM Stationery Office.

Nattrass, N. (2002). Unemployment, employment and labour force participation in Khayelitsha/Mitchell's Plain. CSSR Working Paper, 133.

Nattrass, N., & Walker, R. (2005). Unemployment and reservation wages in working-class Cape Town. *South African Journal of Economics*, 73(3), 498-509.

Newhouse, D. (2005). The persistence of income shocks: Evidence from rural Indonesia. *Review of development Economics*, 9(3), 415-433.

Nicoletti, C., & Rabe, B. (2013). Inequality in Pupils' Test Scores: How Much do Family, Sibling Type and Neighbourhood Matter?. *Economica*, 80(318), 197-218.

Orthofer A. (2016). Wealth Inequality in South Africa: Insights from Survey and Tax Data. REDI3x3 Working Paper 15.

Petterson, S. (1997). Are young Black men really less willing to work?. *American Sociological Review*, 605-613.

Pissarides, C. (2000) Equilibrium Unemployment Theory, Second Edition. MIT Press, Cambridge, MA.

Posel, D., Fairburn, J. A., & Lund, F. (2006). Labour migration and households: A reconsideration of the effects of the social pension on labour supply in South Africa. *Economic modelling*, 23(5), 836-853.

Raaum, O., Salvanes, K. G., & Sørensen, E. Ø. (2006). The neighbourhood is not what it used to be*. *The Economic Journal*, 116(508), 200-222.

Ranchhod, V. (2006). The effect of the South African old age pension on labour supply of the elderly. *South African Journal of Economics*, 74(4), 725-744.

Rankin, N. A., & Roberts, G. (2011). Youth unemployment, firm size and reservation wages in South Africa. *South African Journal of Economics*, 79(2), 128-145.

Ravallion, M. (2010). The developing world's bulging (but vulnerable) middle class. *World Development*, 38(4), 445-454.

Reddy, V. (2012). Towards Equity and Excellence: Highlights from TIMSS 2011. The south african perspective.

Schnitzlein, D. D. (2014). How important is the family? Evidence from sibling correlations in permanent earnings in the USA, Germany, and Denmark. *Journal of Population Economics*, 27(1), 69-89.

Sestito, P. & Viviano, E. (2011). Reservation Wages: Explaining Some Puzzling Regional Patterns. *Labour*, 25 (1), 63–88.

Sienaert, A. (2008). The Labour Supply Effects of the South African State Old Age Pension: Theory, Evidence and Implications. A SALDRU Working Paper 20, Cape Town.

Shimer, R., & Werning, I. (2007). Reservation wages and unemployment insurance. *The Quarterly Journal of Economics*, 122(3), 1145-1185.

Solon, G., Page, M. E., & Duncan, G. J. (2000). Correlations between neighboring children in their subsequent educational attainment. *Review of Economics and Statistics*, 82(3), 383-392.

Solon, G. (1999). Intergenerational mobility in the labor market. Handbook of labor economics, 3, 1761-1800.

Solon, G. (2002). Cross-country differences in intergenerational earnings mobility. *The Journal of Economic Perspectives*, 16(3), 59-66.

Spaull, N., & Kotze, J. (2015). Starting behind and staying behind in South Africa: The case of insurmountable learning deficits in mathematics. International Journal of Educational Development, 41, 13-24.

Statistics South Africa, (2012a). Census 2011 Metadata. Statistics South Africa. Pretoria.

Statistics South Africa, (2012b). Census 2011 Statistical release – P0301.4. Statistics South Africa. Pretoria.

Statistics South Africa, (2014). The South African MPI: Creating a multidimensional poverty index using census data. Statistics South Africa. Pretoria.

Taylor, N., van der Berg, S., & Mabogoane, T. (Eds.). (2013). Creating effective schools. Cape Town: Pearson.

Taylor, S., van der Berg, S., Reddy, V., & Janse van Rensburg, D. (2011). How well do South African schools convert grade 8 achievement into matric outcomes? (No. 13/2011). Stellenbosch University, Department of Economics.

Van der Berg, S., Burger, R., Burger, R., Louw, M., & Yu, D. (2005). Trends in poverty and inequality since the political transition. Stellenbosch Working Paper 1/2005. Stellenbosch University, Department of Economics.

Van der Berg, S., Burger, C., Burger, R., de Vos, M., du Rand, G., Gustafsson, M., ... & van Broekhuizen, H. (2011). Low quality education as a poverty trap. Stellenbosch: Stellenbosch University.

Van der Berg, S. (2007). Apartheid's enduring legacy: Inequalities in education. *Journal of African Economies*, 16(5), 849-880.

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

Van der Berg, S. (2009). The persistence of inequalities in education. In B. Kahn & G. Kingdon (Eds.), South African economic policy under democracy (pp. 327–354). Oxford: Oxford University Press.

Van Wyk, C. (2015). An overview of education data in South Africa: An inventory approach'. South African *Journal of Childhood Education* (Forthcoming).

Von Fintel, D. (2007). Dealing with Earnings Bracket Responses in Household Surveys–How Sharp are Midpoint Imputations?. South African Journal of Economics, 75(2), 293-312.

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

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

Woolard, I., & Klasen, S. (2005). Determinants of income mobility and household poverty dynamics in South Africa. *Journal of Development Studies*, 41(5), 865-897.

World Bank (2012). South Africa, Focus on Inequality of Opportunity. Economic Update: 71553.

Appendix A:

Additional Tables (Chapter 2)

Table A.1: Number of observations and youth characteristics in CAPS

	Wave 1	Wave 2	Wave 3	Wave 4	Wave5
Number of observation	2140	1787	1620	1558	1317
Employed	25%	38%	52%	59%	63%
In school	58%	43%	30%	21%	9%
Age	17.7	19.5	20.8	21.7	24.6
Black	28%	26%	27%	26%	26%
Coloured	53%	56%	55%	55%	54%
White	18%	17%	17%	19%	19%

Note: Descriptive statistics use sample weights. The weighted distributions are within two percentage point of the population group distribution in Cape Town in the 1996 Census (see Lam et al., 2013). Due to the small number of observations from the Indian population (11), this group was omitted from the analysis. (Source: CAPS wave 1-5)

Table A.2: Attrition pattern for CAPS wave 2-5

Frequency	Percent	Wave observed			
		2	3	4	5
1028	51.95	yes	yes	yes	yes
255	12.89	yes	yes	yes	
156	7.88	yes			
107	5.41	yes	yes		
91	4.60	yes	yes		yes
67	3.39	yes		yes	yes
59	2.98		yes	yes	yes
59	2.98	yes		yes	
37	1.87		yes	yes	
120	6.06		other pattern		

(Source: CAPS wave 2-5)

Table A.3: List of hypothetical job offers in CAPS

Job description	Rand amount (in 2002 value)	Wave
Domestic worker	846	2 to 5
Security guard	1300	2 to 5
General worker	1438	2 to 5
Machine operator	1619	2 to 5
Cashier at retail store	2000	2 to 5
Bookkeeper	3000	2 to 4
Accept job for R3000	3000	5
Production manager	5000	4

(Source: CAPS wave 2-5)

Table A.4: Reason for stopping last job

Still employed	817
Study	155
New job	137
Quit	692
Family	118
Dismissed/bankrupt/Stopped	735

(Source: CAPS wave 4)

Table A.5: Interval regression

VARIABLES	(1) model	(2) lnsigma
Employed	0.213***	
	(0.0142)	
In school	0.238***	
	(0.0184)	
Number household members work	0.0154***	
	(0.00550)	
Age	-0.0554**	
	(0.0223)	
Age2	0.00155***	
	(0.000490)	
Ever worked	-0.0684***	
	(0.0194)	
Prim education	0.0823*	
	(0.0460)	
Matric	0.229***	
	(0.0474)	
Tertiary	0.346***	
	(0.0518)	
Numeracy score (%)	0.00294***	
	(0.000309)	
Male	0.0870***	
	(0.0119)	
Asset index	0.155***	
	(0.00888)	
Coloured	0.150***	
	(0.0202)	
White	0.265***	
	(0.0325)	
HH head	0.0758***	
	(0.0233)	
HH size	-0.0126***	
	(0.00230)	
People gave daily self- reported RW	-0.204***	
	(0.0330)	
People gave weekly self-reported RE	-0.212***	
	(0.0175)	
Constant	7.630***	-0.726***
	(0.254)	(0.0110)
Observations	11,852	11,852

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Not reported wave and location dummies. (Source: CAPS wave 2-5)

```
Log-Lik Intercept Only:    -2.789e+06     Log-Lik Full Model:         -2.429e+06
D(11819):                 4857324.281     LR(30):                     721086.366
                                          Prob > LR:                       0.000
McFadden's R2:                  0.129     McFadden's Adj R2:               0.129
```

Table A.6: Effects of reservation wages

	(1A) FE	(1B) FE	(2A) FD	(2B) FD	(4) OLS	(5) OLS	(6) OLS
VARIABLES	Employed	Employed	ΔEmployed	ΔEmployed	Transitioned into employment	Transitioned out of employment	Quit
Lagged $\log(rw_1)$	-0.018				-0.011	-0.012	-0.022*
	(0.019)				(0.038)	(0.020)	(0.012)
Lagged $\log(rw_2)$		-0.044*			0.013	0.022	-0.002
		(0.025)			(0.049)	(0.022)	(0.014)
Lagged $\Delta\log(rw_1)$			-0.011				
			(0.019)				
Lagged $\Delta\log(rw_2)$				-0.061**			
				(0.025)			
Constant	0.606*	0.801**	-0.305***	-0.298***	0.339	0.391*	0.417***
	(0.351)	(0.385)	(0.024)	(0.024)	(0.372)	(0.201)	(0.109)
Observations	2,817	2,817	1,245	1,245	614	1,584	1,674
R-squared	0.206	0.208	0.153	0.157	0.159	0.198	0.045

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Also controlled for years of schooling, experience and wave dummies. (Source: CAPS wave 2-5)

Appendix B:

Additional Tables (Chapter 3)

Table B.1: Regression coefficients for South African income regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS estimation of θ_1 - θ_7						**IV estimation**
	Δy_2	Δy_3	Δy_3	$y_3 - y_1$	Δy_3	Δy_3	
y_1	-0.249***		-0.0427**	-0.292***	0.329***		
	(0.0251)		(0.0196)	(0.0254)	(0.0295)		
y_2		-0.243***			-0.495***		-0.0523**
		(0.0227)			(0.0267)		(0.0247)
Δy_2						-0.409***	
						(0.0280)	
Constant	1.825***	1.911***	0.471***	2.296***	1.375***	0.189***	0.542***
	(0.174)	(0.156)	(0.139)	(0.176)	(0.134)	(0.0211)	(0.176)
Observations	2,770	2,770	2,770	2,770	2,770	2,770	2,770
R-squared	0.129	0.141	0.004	0.170	0.252	0.194	0.054

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1
(Source: NIDS wave 1-3)

Table B.2: Regression coefficients for Chile income regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS estimation of θ_1 - θ_7						**IV estimation**
	Δy_2	Δy_3	Δy_3	$y_3 - y_1$	Δy_3	Δy_3	Δy_3
y_1	-0.362***		-0.077***	-0.443***	0.331***		
	(0.0165)		(0.0153)	(0.0164)	(0.0191)		
y_2		-0.449***			-0.640***		-0.121***
		(0.0151)			(0.0185)		(0.0227)
Δy_2						-0.507***	
						(0.0202)	
Constant	4.186***	5.175***	0.892***	5.127***	3.574***	0.0202**	1.400***
	(0.178)	(0.176)	(0.176)	(0.189)	(0.173)	(0.00995)	(0.261)
Observations	5,462	5,396	5,382	5,455	5,382	5,382	5,382
R-squared	0.157	0.218	0.006	0.189	0.287	0.209	0.102

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1
(Source: EPC wave 1-3)

Table B.3: Regression coefficients for Tanzania income regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS estimation of θ_1 - θ_7						**IV estimation**
	Δy_2	Δy_3	Δy_3	$y_3 - y_1$	Δy_3	Δy_3	
y_1	-0.536***		-0.0316	-0.563***	0.252***		
	(0.0221)		(0.0199)	(0.0236)	(0.0209)		
y_2		-0.485***			-0.605***		-0.0674
		(0.0224)			(0.0221)		(0.0414)
Δy_2						-0.427***	
						(0.0208)	
Constant	5.511***	4.918***	0.337*	5.810***	3.648***	0.130***	0.706*
	(0.216)	(0.226)	(0.205)	(0.232)	(0.244)	(0.0296)	(0.424)
Observations	2,922	2,934	2,888	2,924	2,888	2,888	2,888
R-squared	0.269	0.265	0.001	0.307	0.322	0.218	0.069

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1
(Source: NPS wave 1-3)

Table B.4: Robustness analysis GMM estimates for South African

	Original sample		Wave 1 HH head can be followed in all 3 waves		Wave 1 household can be followed in all 3 waves	
	(1)	(2)	(3)	(4)		
β	-0.0590***	-0.0886***	-0.0472***	-0.0943***	-0.0294	-0.0823***
	(0.0174)	(0.00455)	(0.0139)	(0.00388)	(0.0482)	(0.00662)
α	0.801***	1	0.782***	1	0.698***	
	(0.0195)	-	(0.0171)	-	(0.0368)	
Observations	2,770	2,770	3,979	3,979	4,644	4,644
J-test statistic	0.249	73.2	0.454	109.3	3.155	42.98
p-value	0.969	0	0.929	0	0.368	0

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1. Source (NIDS wave 1-3)

Appendix C:

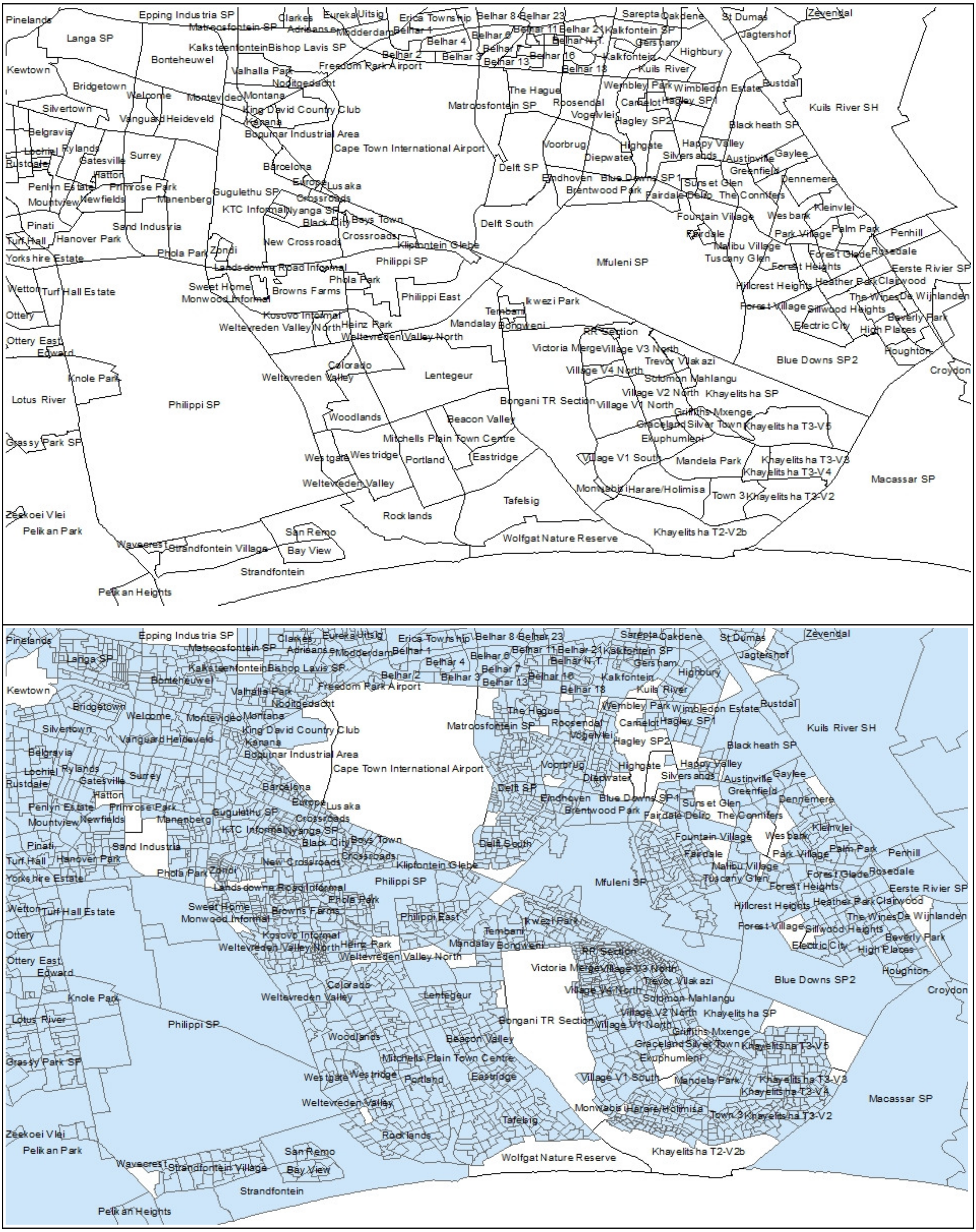
Additional Figures and Tables (Chapter 4)

Table C.1: Variables used to generate the wealth index using PCA

Category	Variables
Labour market status:	Employed, unemployed, discouraged work seeker, not economically active
Education:	No education, some primary education, secondary education, matric, higher/further education
Household income:	No income, low income (1 – 38200 rand) , middle income (38201 - 153800 rand), high income (153800 – 2457601 rand or more), Unspecified
Household assets:	Cell phone, computer, motor car, refrigerator, satellite TV, stove, TV, washing machine
Household services:	Water source –regional water scheme, Waste removal –local authority weekly, Internet access - from home; from cell phone; from work, from elsewhere; no access to internet, Toilet: flush toilet; flush toilet septic tank
Household size:	Household size 1 - Household size 10+

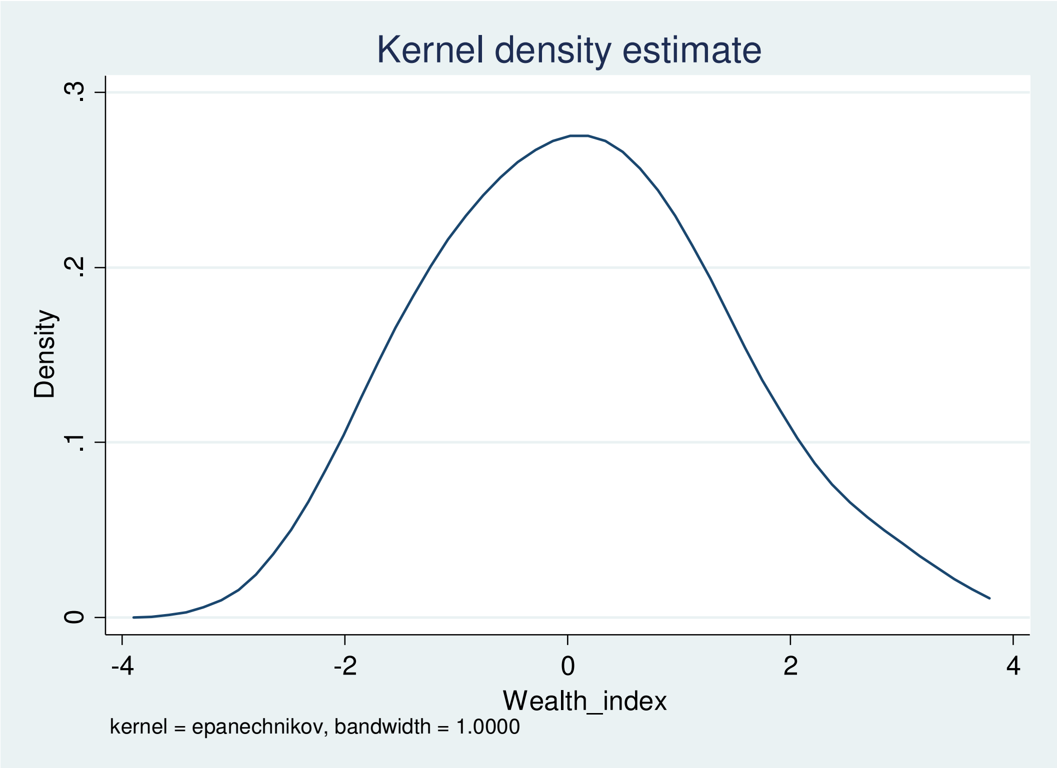
(Source: Census 2011)

Figure C.1: Example of SP and SAL maps in metropolitan Cape Town



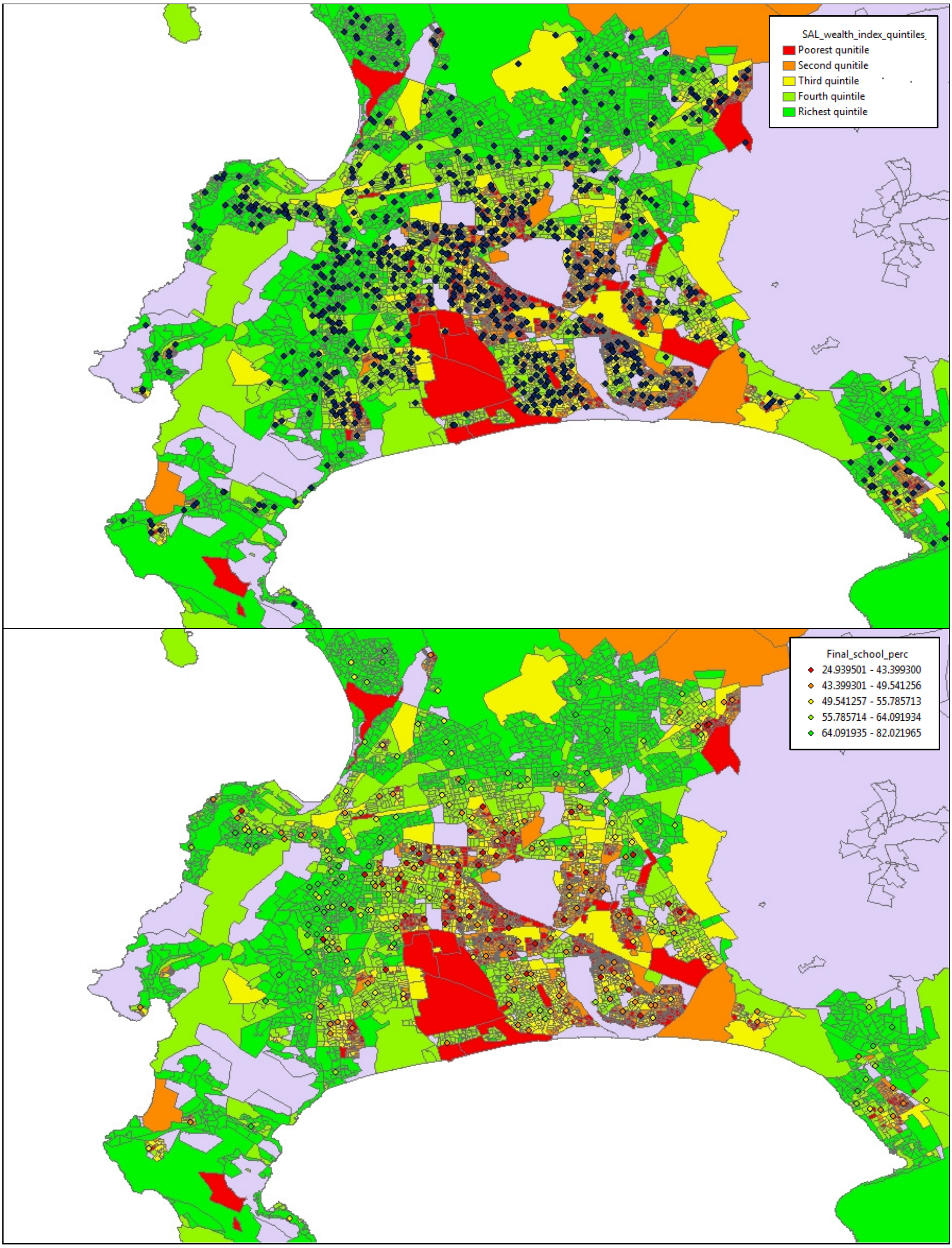
Note: The first map on top shows SP and the second map at the bottom the SAL layers. (Source: Census 2011 data)

Figure C.2: Kernel density of PCA wealth index



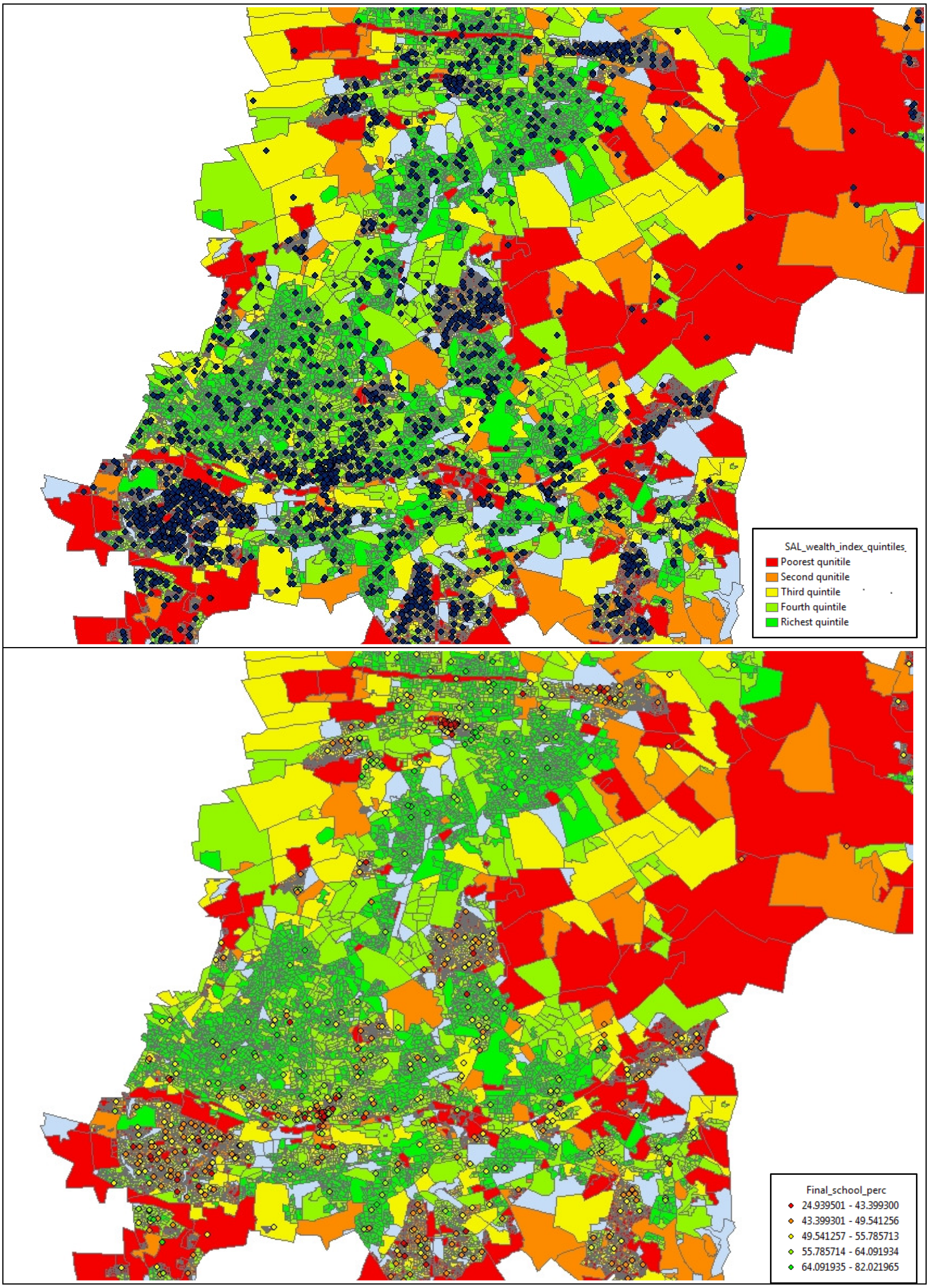
(Source: Census 2011)

Figure C.3: Neighbourhood wealth quintiles, geographic school distribution and matric 2014 examination results in Cape Town metropolitan municipality



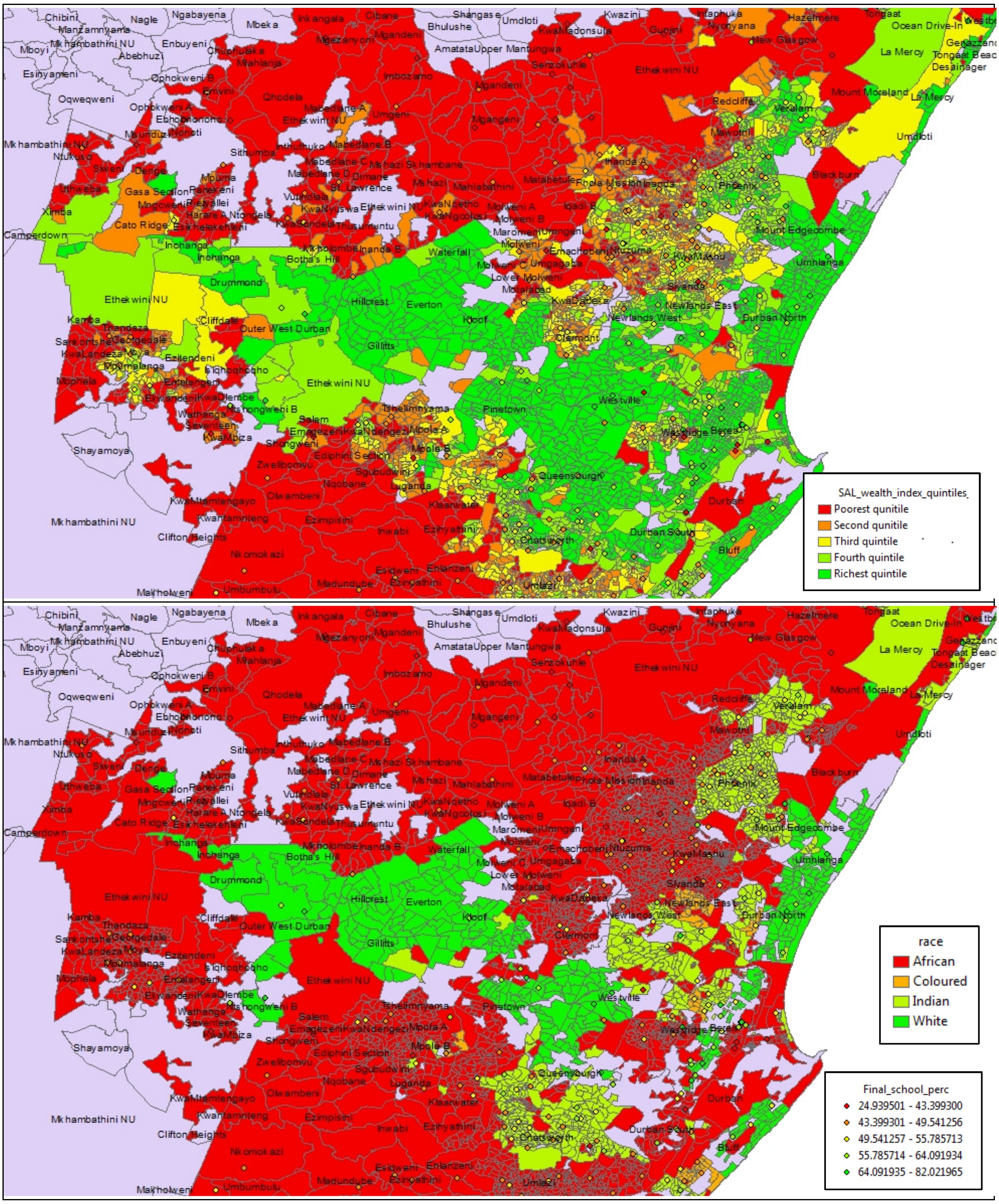
Note: In the first map the entire geographic distribution of primary and secondary schools is shown. In the second map the average matric 2014 examination results per school are displayed.
(Source: Census 2011 and DBE school data)

Figure C.4: Neighbourhood wealth quintiles, geographic school distribution and matric 2014 examination results in Johannesburg, Tshwane and Ekurhuleni metropolitan municipalities



Note: In the first map the entire geographic distribution of primary and secondary schools is shown. In the second map the average matric 2014 examination results per school are displayed. (Source: Census 2011 and DBE school data)

Figure C.5: Neighbourhood wealth quintiles, geographic race distribution and matric 2014 examination results in eThekwini metropolitan municipality



Note: The colours in the bottom map indicate a share larger than 50% for a particular race group in that SAL in the Census 2011 data. In addition, the average matric 2014 examination results per school are displayed.
(Source: Census 2011 and DBE school data).

Table C.2: Difference in school and learner neighbourhood wealth quintiles

		Western Cape		Eastern Cape	
				11818	0.61%
	-3	5562	0.70%	80753	4.26%
School wealth quintile – student wealth quintile	-2	5537	0.70%	20864	1.10%
	-1	37597	4.74%	171369	9.03%
	0	427224	53.89%	1110119	58.50%
	1	178489	22.51%	199933	10.54%
	2	60356	7.60%	114254	6.02%
	3	78126	9.85%	171579	9.04%
				16789	0.88%

(Source: NIDS and Census 2011)

Table C.3: OLS regression – Difference in school and learner neighbourhood wealth quintiles

VARIABLES	(1) Difference school and neighbourhood quintiles	(2)	(3)
Male	-0.0352	-0.0247	-0.0274
	(0.0255)	(0.0255)	(0.0255)
Age	-0.00321	-0.00789	-0.00722
	(0.00938)	(0.00933)	(0.00932)
White	1.922***	1.907***	1.839***
	(0.532)	(0.531)	(0.531)
Coloured	0.975***	0.975***	0.942***
	(0.114)	(0.114)	(0.114)
Indian	3.238***	3.341***	3.283***
	(0.124)	(0.124)	(0.124)
Mother education	0.0174***		
	(0.00368)		
Father education	0.0236***		
	(0.00397)		
Ln(distance to closest school)	0.135***	0.135***	0.132***
	(0.0157)	(0.0157)	(0.0157)
Urban	1.073***	1.095***	1.080***
	(0.0363)	(0.0361)	(0.0362)
Father matric		0.308***	0.287***
		(0.0468)	(0.0470)
Mother matric		0.256***	0.223***
		(0.0431)	(0.0439)
Ln(per capita income)			0.0614***
			(0.0158)
Constant	0.177	0.451**	0.0789
	(0.193)	(0.189)	(0.211)
Observations	5,393	5,393	5,392
R-squared	0.348	0.348	0.350

Not reported province dummies (Source: NIDS wave 1-3)

Table C4 and Table C5 provide the average distance to school for each student in NIDS. The columns to the left provide the values for the new school distance variable which were obtained from the matching process of NIDS with the master lists of schools using fuzzy matching on school name and location. Given the significant differences to the match provided in the NIDS data (distance to school NIDS), in all the analysis of this chapter only the school information for wave 1-3 from the own matching was used.

Table C.4: Distance to school

			Distance to school new			Distance to school from NIDS		
			km	Std.	Number	km	Std.	Number
Secondary age	Urban	Africa	18.68116	107.9719	3178	50.8819	158.7966	3191
		Coloured	10.43689	47.35696	863	91.61249	249.3532	822
		Indian	6.672051	6.829059	56	29.53672	163.9698	58
		White	27.06995	101.6134	139	77.66129	224.9168	14
	Rural	Africa	10.99625	36.3324	6898	28.64904	109.5429	6822
		Coloured	21.93374	45.39698	169	73.19331	227.7002	160
		Indian	7.298066	1.012356	24	25.892	89.19453	25
		White	44.2478	42.08362	9	396.2678	567.6691	9

(Source: NIDS wave 1-4)

Table C.5: Distance to school

			Distance to school new			Distance to school from NIDS		
		Year	km	Std.	Number	km	Std.	Number
Primary age	Urban	2008	5.355357	23.17251	1837	39.87425	164.4093	1717
		2010	7.199118	45.95541	1775	45.70813	191.2405	1704
		2012	6.942235	35.83913	2239	25.70149	119.8447	1988
		2014				37.1221	153.8529	2739
	Rural	2008	9.102381	41.46133	2877	26.61582	116.4689	2681
		2010	7.182278	22.9229	2832	25.495	108.0987	2729
		2012	7.046202	25.88818	3536	20.7728	100.6671	3226
		2014				34.58738	238.4152	3817
Secondary age	Urban	2008	14.15244	68.49511	1305	58.58729	179.4829	1383
		2010	20.04269	137.484	1310	60.26372	181.1181	1275
		2012	17.14205	76.90393	1621	59.57468	188.3	1560
		2014				61.75101	179.7867	2014
	Rural	2008	11.82851	38.67698	2080	31.22783	120.053	2094
		2010	11.14124	37.08875	2281	29.84224	112.0515	2237
		2012	10.9952	34.43937	2739	29.50474	116.0339	2685
		2014				53.94867	492.6208	2952

(Source: NIDS wave 1-4)

Table C.6: Household characteristics for age 15-18

	Mean	Std.	Number
Per capita income	1520.6	2043.81	7245
Black	83.42%	0.31	7245
Coloured	8.48%	0.33	7245
Indian	2.30%	0.10	7245
White	5.80%	0.13	7245
Mother education	7.81	4.06	7245
Mother not in HH	0.16	0.38	7245
Father education	7.24	3.84	7245
Father not in HH	0.38	0.49	7245

Table uses population weights to obtain national representative results (Source: NIDS wave 1-3)

Table C.7: Pooled OLS regression: reached years of education by age 15-18 with DBE quintiles

VARIABLES	(1) Years of education	(2) Years of education	(3) Years of education	(4) Years of education
Age	0.723***	0.709***	0.730***	0.726***
	(0.0222)	(0.0209)	(0.0207)	(0.0204)
Male		-0.513***	-0.517***	-0.521***
		(0.0805)	(0.0786)	(0.0792)
White		0.260	-0.130	-0.256
		(0.178)	(0.168)	(0.168)
Indian		0.873***	0.475**	0.357*
		(0.196)	(0.186)	(0.191)
Coloured		0.241*	0.174	0.116
		(0.127)	(0.118)	(0.125)
Mother education			0.0683***	0.0628***
			(0.00869)	(0.00875)
Mother not in the household			-0.165**	-0.178**
			(0.0793)	(0.0788)
Father education			0.0471***	0.0394***
			(0.00783)	(0.00783)
Father not in the household			-0.130**	-0.106**
			(0.0514)	(0.0499)
Ln(per capita income)				0.128***
				(0.0265)
2. DBE school quintile	0.119	0.0513	-0.0634	-0.0601
	(0.134)	(0.115)	(0.0994)	(0.0993)
3. DBE school quintile	0.251*	0.203*	0.0326	0.0114
	(0.130)	(0.118)	(0.101)	(0.102)
4. DBE school quintile	0.667***	0.470***	0.173	0.151
	(0.122)	(0.123)	(0.110)	(0.109)
5. DBE school quintile	1.020***	0.608***	0.114	0.0136
	(0.130)	(0.141)	(0.131)	(0.134)
Constant	-3.547***	-3.053***	-3.916***	-4.514***
	(0.364)	(0.374)	(0.368)	(0.401)
Observations	7,254	7,254	7,247	7,245
R-squared	0.263	0.311	0.356	0.359

(Source NIDS wave 1-3)

Table C.8: Pooled OLS regression: reached years of education by age 15-18 with additional controls

VARIABLES	(1) Years of education	(2) Years of education	(3) Years of education
Age	0.705***	0.695***	0.696***
	(0.0223)	(0.0218)	(0.0218)
Male	-0.493***	-0.498***	-0.495***
	(0.0767)	(0.0773)	(0.0767)
White	-0.403**	-0.449**	-0.469**
	(0.177)	(0.178)	(0.182)
Indian	0.218	0.177	0.168
	(0.195)	(0.196)	(0.196)
Coloured	0.154	0.124	0.119
	(0.111)	(0.113)	(0.117)
Ln(per capita income)	0.107***	0.123***	0.120***
	(0.0318)	(0.0288)	(0.0290)
Mother education	0.0628***	0.0596***	0.0588***
	(0.00934)	(0.00886)	(0.00881)
Mother not in the household	-0.150**	-0.155**	-0.160**
	(0.0743)	(0.0747)	(0.0752)
Father education	0.0366***	0.0344***	0.0344***
	(0.00853)	(0.00774)	(0.00774)
Father not in the household	-0.110**	-0.101**	-0.0981**
	(0.0486)	(0.0497)	(0.0495)
Ln(School distance new)	0.0374*	0.0357	0.0330
	(0.0226)	(0.0226)	(0.0228)
Neighbourhood wealth quintiles	YES	YES	YES
2. School neighbourhood quintile	0.0810	0.103	0.106
	(0.113)	(0.112)	(0.113)
3. School neighbourhood quintile	0.158	0.182*	0.179
	(0.112)	(0.110)	(0.111)
4. School neighbourhood quintile	0.127	0.144	0.133
	(0.125)	(0.125)	(0.125)
5. School neighbourhood quintile	0.277*	0.275*	0.251*
	(0.141)	(0.141)	(0.138)
Learner-teacher ratio		0.920	0.568
		(2.224)	(2.121)
SGB teacher share		0.0391	
		(0.208)	
Private school			0.134
			(0.215)
Mixed funded school			0.0881
			(0.101)
Constant	-4.213***	-4.113***	-4.089***
	(0.396)	(0.408)	(0.409)
Observations	7,131	7,050	7,057
R-squared	0.354	0.356	0.356

(Source NIDS wave 1-3)