THE UNIVERSITY OF
SYDNEY

TRANSITIONS TO ADULTHOOD: EDUCATION CHOICES, JOB SEARCH AND LABOUR MARKET OUTCOMES

A Thesis

Presented to the Faculty of Agriculture and Environment

of The University of Sydney

in Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Darian Naidoo

March 2014

Declaration

I hereby certify that this work has not been previously submitted for a degree or diploma in any University and that it contains no material previously published or written by another person except where due reference is made in the thesis itself.

Darian Naidoo,

March 2014

TRANSITIONS TO ADULTHOOD: EDUCATION CHOICES, JOB SEARCH AND LABOUR MARKET OUTCOMES

Darian Naidoo, Ph.D.

The University of Sydney 2014

Imperfect information about the costs and benefits of education or about job opportunities may imply that youth will not always make the right investments in education or be matched with the right jobs. This thesis investigates two research questions centred on the transition to adulthood: how to inform youth about the returns to education and what are the wage effects of using social networks to find jobs. These questions are considered in the context of South Africa, using data from the Cape Area Panel Study (CAPS), a longitudinal study of youth.

Regarding the first question, existing research on the impact of information interventions has largely ignored the importance of the sequential nature of education investment and heterogeneity in the returns to education, two factors that may affect the impact of information interventions. This thesis addresses this gap in the literature by using a sequential decision making model of investment in education, to investigate the importance of the (in)accuracy of expectations of the return to grade 12 in South Africa. This model allows us to consider how agents may update their expectations in response to new information, when making sequential investment decisions. We find that even if agents respond rationally and are provided with correct information, the accuracy of their expectations may be reduced possibly leading to sub-optimal educational choices.

Although it is well known that social networks are productive, in the sense that they increase the probability of finding a job, much less is known about the quality of the match between job seeker and employee that is promoted through their use and, in particular, the impacts on wages, resulting from their use. This thesis estimates the impact of social networks on wages for youth in South Africa. It concludes that the use of social networks is associated with wage discounts and that decreasing social distance (the use of relatives rather than friends), is associated with greater wage penalties.

These conclusions have important policy implications. Firstly, we conclude that information interventions should not be conducted without *ex ante* consideration of the issues of heterogeneity, sequential investment and the process of updating of expectations. Such analysis is likely to be less costly than conducting field experiments and provides the benefit of anticipating negative impacts before they occur. Secondly, the negative impact on wages of the use of social networks to find jobs suggests that in the context of South Africa, there may be a role for interventions that assist youth to find employment.

ACKNOWLEDGEMENTS

This work would not have been possible without contributions and support from many people to whom I am deeply grateful. First and foremost, I sincerely thank Dr. Paulo Santos for his excellent supervision of my work. His patience, effort and the enthusiasm he has expressed in my research have all been invaluable. My research was also supervised by Dr. Greg Hertzler to whom I am deeply grateful, particularly for his guidance regarding the research presented in Chapter 2.

Chapter 3 of this paper has benefited from the comments of many people. I thank Christopher Barrett, Simon Quinn, Uttam Sharma, Paul Frijters, Russell Toth as well as participants at the 2012 PacDev Conference (UC Davis), the NSF-LSU Workshop on Networks and Development (Baton Rouge), the 8th Australasian Development Economics Workshop (Monash) and the 2012 Australasian Econometric Society Meeting (Melbourne) for comments that greatly improved this chapter.

I would like to thank all the staff and postgraduate students in the Department of Agricultural and Resource Economics, Faculty of Agriculture and Environment at The University of Sydney, who have been very helpful in many ways throughout my candidature. In particular, I would like to thank my colleagues Elizabeth Bernold, Evarn Ooi, Emrullah Ozden, and David Rogers for their support and encouragement. I also thank my friends Elizabeth Bernold, Evarn Ooi, Ellie Wong and Caitlin Naidoo for help with editing.

I thank Truman Packard, Sonya Woo and David Newhouse for allowing me to fit my employment with The Bank, around the demands of my research work.

Finally, I would like to thank all of my family and friends who have supported me over the years. A special thanks to the Chetty family and other extended

family, for accommodating and assisting me in South Africa. I am also particularly grateful to my parents, Lily and Yogs. Without their valuable support, I would not have been able to undertake a PhD.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Economic transitions to adulthood

According to the *2007 World Development Report* (World Bank, 2006), there are two central and intimately linked economic decisions in the transition into adulthood. The first is progression through education in preparation for work. The second is entry into the labour market. Together, they allow youth to transition from being dependents who are building human capital with economic support from their families and others, to economically independent agents that use their human capital productively in work.

Across countries, the nature and timing of these decisions varies considerably because of differences in education systems and labour market institutions. However, some basic characteristics of these transitions are almost universally true, with two characteristics being particularly relevant to this thesis. Firstly, decisions regarding education are intimately linked to future labour market outcomes. Because education has well-documented returns in the labour market (Psacharopoulos and Patrinos, 2004), expected labour market returns may also have a direct impact on education investment decisions (Manski, 1993). Secondly, every job must be found through some type of search method. It is possible that different types of job search methods lead to differences in the quality of the match between youth and the jobs they find (Loury, 2006). Both decisions are characterised by risk and imperfect information, with important welfare consequences in both the short and

the long term.

Lack of knowledge about future labour market returns associated with different levels and types of education or capacity to make such an investment (Heckman, Lochner, and Todd, 2006, Stange, 2012) may lead to suboptimal decisions, both in terms of private and public costs. For example, a young student who spends a year in education and fails, will necessarily incur some direct costs of education and the opportunity cost of his/her time, and these costs may not have a payoff if only completed grades/degrees are rewarded in the labour market. Perhaps of even greater concern, a student dropping out of school earlier than they should carries long term consequences for both the youth (through reduced lifetime earnings) and society (if underinvestment in education leads to lower overall labour productivity).

Having decided to enter the labour market, youth must decide how much time and resources to spend on job search. The way that jobs are searched for has important implications regarding the costs and potential benefits of job search, in terms of the characteristics of the job eventually found. While the use of friends and relatives is usually perceived to have lower costs than formal search methods Holzer (1987), it may also have consequences in terms of labour market outcomes, measured by the probability of finding a job and the characteristics of the job found. Although much is known about the fact that the use of social contacts is productive (in the sense that it increases the probability of receiving a job offer), the impact of informal search on wages is less clear: finding employment in relatively unsuitable jobs may be associated with wage penalties (Loury, 2006), which may have to be weighted against the costs associated with unemployment, which may be long lasting and "scarring" (Heckman and Borjas, 1980, Arulampalam, Gregg, and Gregory, 2001).

This thesis addresses research questions that are suggested by these decisions. The first question, which is the focus of chapter 2, is how to inform youth about the returns to education. The second, addressed in chapter 3, is whether the use of social networks have important wage effects. A preview of the approaches and main results of these two chapters is presented in Sections 1.3 and 1.4, after briefly introducing issues of education investment and job search in the South African context (Section 1.2), which is the focus of the analysis of this thesis. We conclude in Chapter 4, with a discussion of the policy implications of this analysis, in particular in terms of interventions designed to assist youth with information that may help youth make better choices.

1.2 Education investment and job search in South Africa

In South Africa, there are marked inequalities in educational attainment between race groups, at least in part due to the legacy of apartheid. Particularly, while there are high rates of enrolment, many youth repeat grades in primary and high school, especially African youth and to a lesser extent Coloured youth (Pugatch, 2012, Lam, Ardington, and Leibbrandt, 2006). In addition, Whites complete on average more education than Coloureds who in turn complete more education than Africans. [1]

Educational inequality across race groups is related to economic inequality across races groups as there are increasing returns to investment in education

[1] I adopt the labels used in CAPS and by Seekings and Nattrass (2005) and use the word 'African' to refer to peoples classified under apartheid as "native", "bantu" or "black", the word 'White' to refer to people classified as European or white, and the word 'Coloured' to refer to the people that don't fit the first two categories, including descendants of the indigenous Khoi and San, the descendants of "Malay" slaves from Indonesia and those with mixed race ancestry.

and, for all levels of education, average earnings diverge by race, with Whites earning more than Coloureds who in turn earn more than Africans (Keswell and Poswell, 2004). The economic returns to education are a central motivation in post high school education choices, with two of the top three motivations for pursuing higher education being expected increased employment probability and expected increased income, following higher education completion (Cosser and du Toit, 2002).

However, there has been no research that assesses the accuracy of earnings expectations in South Africa and how they may differ by race. Beutel and Anderson (2007) consider the amount of education youth and their parents expect to complete and find it to be generally high across races, but this doesn't mean that youth are adequately aware of the economic returns to education. Even if imperfect information about such returns only plays a small role in explaining underinvestment in education in South Africa, the cost effectiveness of information interventions designed to address such gaps (Nguyen, 2008) may make them worthwhile .

The definition of the unemployment rate and its relation with job search behaviour is one prominent and controversial policy debate in South Africa, centred on the relative merits of the broad and narrow definitions of unemployment. The narrow definition of unemployment (which is the official definition used by Statistics South Africa) only includes those that have searched for work in the past week, while the broad definition also includes those that want work but are not actively searching for it. It has been argued by Kingdon and Knight (2006) that the broad definition should be preferred over the official definition as the non-searching unemployed display key characteristics similar to that of the formally unemployed. In particular, they are economically more deprived than the searching unemployed

while not being any happier. Kingdon and Knight (2006) also suggest that high search costs and high rates of local unemployment are associated with job seeker discouragement. This implies that the common labour market information problem of workers and firms not being able to find each other in appropriate matches, may be particularly acute in South Africa where searching for work is a costly and often unproductive activity.

Although receiving much less attention than the formally unemployed, discouraged job seekers in South Africa numbered over 1 million people during the first quarter of 2008 (Statistics South Africa, 2008). This represents 3.8% of the working age population and is equivalent to 28% of those formally unemployed. Interestingly, there is an over representation of youth in both the measure of formal unemployment and the measure of the non-searching unemployed (Kingdon and Knight, 2007). While there may be many youth that want work but do not actively search for work, this does not mean that such youth will not be able find employment. It is likely that many of the non-searching unemployed find work through family and friends (Burns, Godlonton, and Keswell, 2010). However, the consequences on labour market outcomes of using different types of job search methods (including finding work through networks when not actively searching) remain unknown. Thus, research on the impact of using social network to find jobs is particularly relevant in the South African context.

Chapters 2 and 3 research the accuracy of earnings expectations and the impact of networks on wages in South Africa, respectively, using data from the Cape Area Panel Study (CAPS), described in Lam, Seekings, and Sparks (2008). CAPS is a longitudinal study of a random sample of 4758 youth aged 14-22 in 2002 living in the Cape Town Metropolitan Area. The study focuses on a wide range of impor-

tant and related outcomes for youth during the transition to adulthood, including schooling and labour market participation. All youth were initially interviewed in 2002 (wave 1) and then in 2003-2004 (wave 2), 2005 (wave 3) and 2006 (wave 4). As CAPS contains data on observed and expected earnings, job search methods, and an extensive range of demographic and background characteristics, it is an ideal source of data for the empirical analysis in both Chapters 2 and 3.

1.3 Informing youth about the returns to education

The discrepancy between youths' expectations of the returns to education and observed market returns may create opportunities for low cost information interventions to improve the investment decisions of youth. For example, if students drop out of high school because they are unaware of its economic returns, simply giving students information about these returns may change their investment decisions.

For any given context however, the exact types of information that should be provided and the likely impacts of different types of information, should carefully be considered prior to conducting any intervention. As there is no theoretical research that provides optimal designs for information interventions, an empirical literature that favours field experiments has developed as a means of evaluating different types of intervention (e.g. Nguyen (2008)). In chapter 2, we argue that this is a problematic approach to address the question of how to inform youth about the returns to education, given that such information interventions may unintentionally lead to some youth having less accurate expectations and, if effective in terms of changing behaviour, lead them to making inferior investment choices. This potential for harm has gone unnoticed because two key factors that impact

on the likely effectiveness of information interventions, have been inadequately considered.

Firstly, previous research on interventions has largely ignored the sequential nature of decisions to invest in education and in particular, the significance of the range of higher education options available to decisions made during high-school. [2] Secondly, heterogeneity in the returns to education has also largely been ignored. There is no empirical or theoretical research on the implications of these two factors and how they impact on the effectiveness of information interventions. Chapter 2 addresses this gap in the literature. Firstly, we use a sequential decision making model of investment in education, that we call Real Options Pathways for Education (ROPE), to investigate the importance of the (in)accuracy of expectations of the return to grade 12 in South Africa. This model allows us to consider how agents may update their expectations when making sequential investment decisions. While the way in which expectations of the returns to education are updated has been previously researched (e.g. Wiswall and Zafar (2011)), it has not been researched in consideration of the sequential nature of education investment.

Secondly, using ROPE, we simulate the possible effects on the accuracy of expectations, of four designs of information interventions about the returns to education. The first design is simply to provide information on average wages conditional on the completion of grades 11 and 12. The second design is to provide information on average wages conditional on grades 11, 12 and for each level

[2] In the context I consider, South Africa, there are higher education options that are only accessible after the completion of grade 12. While it is possible to leave school before grade 12 completion to enter vocational education in the form of national trade certificates (NTCs), these are terminal and do not lead to higher education (i.e. NTC 1, 2 and 3 are considered equivalent to grades 10, 11 and 12 but do not lead to university.

of higher education. The third design is like the first except that it accounts for heterogeneity by providing average wages by race. The fourth design provides 'full information', that is, average wages by race for all levels of education. For all the types of interventions under consideration, our results suggest that each intervention would be harmful to some youth. While the first and simplest intervention may be harmful to some youth, the other three intervention designs do not offer an unambiguous improvement over the simple intervention, even though they could be considered as providing 'better' information. Results for the second intervention show that even if the sequential nature of investment is fully accounted for, and youth fully update their expectations, ignoring heterogeneity in the returns to education means that this intervention may be harmful. In contrast, results for the third intervention show that if heterogeneity is accounted for and youth fully update their expectations, the sequential nature of investment means that the intervention may still be harmful. Lastly, the fourth intervention design demonstrates that even providing perfect information to rational agents could lead them to have less accurate expectations and potentially, induce underinvestment in education, depending on *how the information is used*.

Chapter 2 concludes that information interventions should not be conducted without *ex ante* consideration of heterogeneity, real options, and the way in which youth update their expectations. The unique contribution of this chapter is to show that while ignoring *any one of these factors* is dangerous, the alternative of *ex ante* consideration of these factors is relatively easy and insightful. While we do not provide a definitive answer to the question regarding how best to inform youth about the returns to education, we do provide theory and empirical evidence that suggests ways in which youth should *not* be informed about the returns to

education.

1.4 Social networks and wages in South Africa

The way in which a job is found may have a direct impact on the quality of the worker-job match. While it is a stylised fact that social networks play an important role in job search, their impact on labour market outcomes, namely wages, is still disputed. The issue is important for two reasons. Firstly, if social networks do have any bearing on labour market outcomes, this implies that labour economists should not stick rigidly to the assumptions of neoclassical economics that ignore issues of social structure (Mouw, 2003). Secondly, if social networks are leading to a large proportion of poor quality job matches, this may be an argument in favour of job search assistance programs as it indicates a failure of existing labour market institutions to appropriately match workers with jobs in an efficient manner. Particularly for youth, who often find work through social contacts and suffer disproportionately from unemployment, understanding the impact of social networks is an important empirical question as there is no clear agreement in the theoretical literature regarding the overall benefits and costs of network use (Loury, 2006).

There are two competing hypotheses that summarise the existing literature: The "good matches" hypothesis suggests that social networks are used to overcome information asymmetries between workers and firms resulting in wage premiums while the "limited choices" hypothesis suggests that networks are used as a low cost but less effective alternative to finding jobs through formal search, resulting in wage discounts. The predictions of these hypotheses have not been tested in South Africa, where social network use to find jobs is known to be high among youth

(Magruder, 2010, Burns, Godlonton, and Keswell, 2010). Chapter 3 addresses this gap in the literature.

Its first major contribution is to show that the use of social networks is generally associated with substantial wage penalties, rejecting the good matches hypothesis. In addition, this chapter considers the role played by social distance (the use of relatives versus friends) on wages. We show that the use of relatives is associated with significant penalties for all races, suggesting that firms may discount the value of the information provided by referees if they are close to job seekers. This effect is especially large for Africans, the group among whom unemployment is highest.

Chapter 2

Dangerous Information: How Should Youth be Informed About the Returns to Education?

2.1 Introduction

Consider a student that has just completed grade 11 and is considering whether to enrol in grade 12 or enter the labour market. In deciding between these two options, the student takes into account the expected lifetime earnings associated with each option and the costs of education. The value of enrolling in grade 12 depends on what they would earn if they entered the labour market following grade 12 *and* on the value of the option of going to university following grade 12. Suppose that this student expects to earn (100, 150, 300) conditional on successfully completing year 11, year 12 and university, respectively. The true returns for someone with her characteristics are (100, 150, 360), and these are double the national average (50, 75, 180). This student underestimates the returns to university and the value of grade 12 (which is influenced by the university option open with the completion of grade 12).

We are interested in the following question: how could the provision of information help such a student to make the correct choice about whether to enrol in grade 12? Suppose that we design an intervention that provides information

on average earnings conditional on different qualifications. If the student adjusts her expectations and replaces them with the information provided she will, post-intervention, expect to earn (50, 75, 180). Because these values are lower than her initial expectations, the intervention led to a greater underestimation of the returns to education. This example illustrates a key conclusion of this paper, that more information may not always be beneficial. In fact, correct information may even cause harm if it reduces the accuracy of expectations and leads to sub-optimal choices, where 'harm' is defined as a reduction in the accuracy of expectations of an individual that may result in an inefficient choice being made.

In this chapter, we argue that unless such potential for harm from information interventions is considered *ex ante* in a sequential choice framework, this potential for harm may go undetected. While we know of no paper that explores such ideas regarding information interventions for the returns to education, the idea of correct information being harmful is not without precedent in the literature on the economics of information: Daley and Green (2012), for example, conclude that "higher quality news can lead to more inefficient outcomes" in a market for assets where stochastic information (news) about the assets is progressively revealed to buyers.

The policy relevance of interventions intended to change behaviour through the provision of information rests on three assumptions. Firstly, that expectations regarding the returns to education are inaccurate. Secondly, that the information provided will affect expectations in a predictable way. Finally, that changes in expectations translate into different educational investment decisions (implying that inaccurate expectations, and not other constraints, are an important determinant of educational choices).

Jensen (2010) is a leading analysis of this type of intervention, as it investigates the effectiveness of an intervention designed to provide information on returns to education to eighth graders in the Dominican Republic. The intervention was motivated by the puzzling fact that, despite very high completion rates for primary school (around 90%) and substantial returns to secondary school (mean earnings approximately 40% higher than for those with primary education), less than a third of individuals completed secondary school. A survey of eight graders showed that the *perceived returns* to education were quite low (supporting the first assumption mentioned above). The provision of information on average earnings conditional on the completion of primary, secondary and university education, to eighth graders in randomly chosen schools, was followed, four years later, by an analysis of schooling outcomes of recipients of this intervention. Compared with a control group, the provision of information resulted in an average increase of 0.20-0.35 years by the students in the treatment schools, although for the *very poorest* students the change in perceptions was not accompanied by a significant effect on schooling. In addition, this increased investment in education is linked directly to changes in expectations: when interviewed four to six months after the intervention, expectations of the return to secondary school had increased for 54% of the treatment group, compared to only 27% of the control group. These findings regarding updated expectations and increased investment for the treatment group support the second and third assumptions listed in the preceding paragraph. The policy importance of Jensen's study come from its cost-effectiveness: it costs very little to provide the information while the impacts on schooling are substantial. Nguyen (2008, p.4), for example, concludes that this type of interventions could increase schooling by one year at a cost of US$2.4 per person.

This paper is a discussion of the effect of some critical, but under-researched, aspects of the design of this type of intervention. We start by developing an analytical approach that addresses two potentially important limitations of interventions such as the one just mentioned. The first is to fully incorporate the sequential nature of education investment, that is, the possibility that completing a certain level of education is valuable because it creates the possibility of additional investments in education (that are valuable in themselves). The second is the possibility of heterogeneous returns to education. The practical question we ultimately address is whether the additional complexities that are introduced matter, in the sense that they lead to different results from simpler approaches, such as the one followed in Jensen (2010). As such, this paper should be understood as a cautionary note regarding expanding or scaling up these types of interventions.

The paper proceeds as follows. In section 2.2, we briefly review previous theoretical and empirical work on sequential investment and option values in education and on the importance of expectations in decision making, in particular with respect to the role of subjective expectations of the returns to education. One contribution of this paper is to link these bodies of work through an evaluation of the importance of differences between subjective expectations and observed values in a sequential decision making framework.

We then present in section 2.3, a simple model of the role of sequential investment in education, which we refer to as Real Options Pathways for Education (ROPE). In section 2.4 we adapt this model to fit the context and the data we use, the Cape Area Panel Study (CAPS) in South Africa. This data set is particularly suited for the purpose of this study as it contains extensive information on individuals' expectations on earnings across the years of the transition to adulthood.

Using the ROPE framework, we do not identify a large mismatch between expectations and reality when focusing on the overall net value of completing grade 12 for Coloured and White youth (who form the most part of our sample). [1] However, for African youth, the average underestimation of returns to education is 32%. If we compare expected and observed earnings unconditional on race, there is on average, an overestimation of about 11% of the returns to grade 12, hiding the underestimation of Africans. We also show that measures of the accuracy of expectations are sensitive to the inclusion of the options of further study after grade 12, that is, the value of being able to enrol in tertiary education (the pathway value of this grade). If we were to ignore the pathway value of grade 12, then on average, Africans and Coloureds in our sample overestimate the returns to grade 12 (by about 48% and 16%, respectively), while Whites underestimate the returns to grade 12 by about 8%. Taken together, these results illustrate the importance of considering both option values and heterogeneity in returns to education when considering information interventions.

Four possible interventions are defined and analysed in Section 2.5. We find that the provision of information on outcomes unconditional on race or on a subset of education pathways, would lead some youth to under-invest or over-invest in education. In particular an intervention that takes into account option values but ignores heterogeneity (mirroring Jensen (2010)) could increase the underestimation of returns by Africans and Whites (from about 32% to 35% for Africans and

[1] We adopt the labels used in CAPS and by Seekings and Nattrass (2005) and use the word 'African' to refer to peoples classified under apartheid as "native", "bantu" or "black", the word 'White' to refer to people classified as European or white, and the word 'Coloured' to refer to the people that don't fit the first two categories, including descendants of the indigenous Khoi and San, the descendants of "Malay" slaves from Indonesia and those with mixed race ancestry.

from about 0% to 18% for Whites), while leading to a substantial overestimation of these returns among Coloureds (up from about 3% to 52%). We also find that, independently of the issues of heterogeneity and option values, the *ex ante* analysis of the effectiveness of information interventions is critically dependent on how expectations are updated. We find that unless the information is used to completely replace all prior expectations, it may result in less accurate expectations, even if the information provided accounts for heterogeneity in returns and option values. Providing full information is not necessarily helpful. In addition, even if the information provided is used to update all prior expectations, unless full information is provided, it may not be helpful. The *possibility* of these results should be cause for concern as they illustrate the need to explicitly consider the type of information provided and the mechanisms through which youth may translate information into decisions prior to conducting an intervention. In particular, consideration should be given to the sequential nature of educational investments, heterogeneity in the returns to education and the process of updating expectations.

In 2.6 we test the robustness of our results to changes in key assumptions regarding the ROPE model and the data. For example, we consider the possibility of there being heterogeneity in the returns to education according to gender rather than race, and consider varying our assumptions about career length. In summary, the results of all sensitivity checks are qualitatively the same as the original results. This demonstrates that our conclusions about the dangers of information provision are not simply due to a very specific combination of model and data assumptions. Lastly, section 2.6 of this chapter concludes.

2.2 Sequential investment in education and subjective expectations

2.2.1 Real options and sequential investment in education

Although important advances have been made in the analysis of the returns to education (see Psacharopoulos and Patrinos (2004) for a review of the literature), several aspects of this decision have received much less attention. In particular, investment in education has traditionally been modelled without much consideration for the fact that such decisions are made in sequence and in consideration of a range of possible outcomes. Every young person must make a series of sequential decisions about education and employment based on their expectations regarding costs and benefits of the available options. Keane and Wolpin (1997) use a dynamic human capital model with data from the National Longitudinal Surveys of Labour Market experience and demonstrate that youths' decisions are consistent with this model of maximising behaviour that is based on sequential and periodic consideration of the expected costs and benefits of school and work options (e.g. school fees and lifetime earnings) over a career that lasts to the nominal age of 65. However, as noted by Heckman, Lochner, and Todd (2006), the relevant expectations of youth may be inaccurate if youth do not have full information on costs and benefits, potentially leading them to make sub-optimal decisions, a difficulty that is shared by economists trying to understand these decisions in order to design policies that may affect them (Manski, 1993).

The inclusion of uncertainty associated with the returns to each level of education (for example, through the inclusion of the variance in the wages of graduates that find jobs and/or the possibilities of unemployment or failure to complete a

level of education) makes real options analysis an attractive approach to the modelling of this decision. Weisbrod (1962) is one of the earliest explicit discussions of the option value of different education levels as a way to overcome such limitations. Using data from Schultz (1961), he concludes that existing estimates of returns to education would be adjusted upward if they were to include the value of higher education options, and that such adjustments could be substantial: for example, the returns to primary school would be adjusted up from 35% to 53.9%.

Based on recent empirical work that considers the sequential nature of education choices (Altonji, 1993, Keane and Wolpin, 1997, Eckstein and Wolpin, 1999, Belzil, 2007, Pugatch, 2012, Lindset and Matsen, 2011), a broader concept of the private value of education should consider the following components:

1. The expected net value of entering the labour market with a particular level of education given current and expected labour market conditions, in particular employment probability and the distribution of wages (for example, Levhari and Weiss, 1974, Keane and Wolpin, 1997).

2. The value of learning about the individual's academic ability and capacity to successfully complete further education (for example, Altonji, 1993, Arcidiacono, 2004, Stange, 2012).

3. The expected value of the options to invest in higher levels of education that are opened by completing lower levels of education. For example, completing grade 12 creates the option of enrolment in tertiary education (Weisbrod (1962) and Lindset and Matsen (2011)).

Traditional measures of the returns to education do not deal with the uncertainty regarding the first component and ignore the second and third components.

Heckman, Lochner, and Todd (2006) argue that for this reason, such estimates most likely suffer from systematic downward bias. They argue, like Weisbrod (1962), that the inclusion of an option value to years of education increases the value of each year of education, which may be especially important for lower years of education. Equally important seems to be the need to link the different components listed above: for example, there has been little work linking uncertainty and sequential investment in the analysis of educational choices (but see Eckstein and Wolpin (1999) as an example of an exception). [2] Also, there is no research that we are aware of that directly compares observed returns and subjectively expected returns in a real options framework. In this chapter, the need for a real options framework (and a sequential model) comes from the way information is linked together in such a framework, which may have implications for modelling the impact of exogenous information shocks. Without using a sequential model, the impact of an exogenous information shock (such as an information intervention) can only be assessed in terms of (1), but not in terms of (3). For example, in order to understand how the provision of information about university may impact on an individual's choice about grade 12, both (1) and (3) must be considered as they are related to each other sequentially. If this were not the case and grade 12 was not related to university entrance, information about the returns to university may have no impact on choices about grade 12 completion. It is important to note that learning about ability (2), or learning about the labour market (which could also be included in a sequential model of education) are beyond the scope of this pa-

[2]Heckman, Lochner, and Todd (2006, p.6) conclude that their analysis of the literature on uncertainty about the returns to education "points to a need for more empirical studies that incorporate the sequential nature of individual schooling decisions and uncertainty about education costs and future earnings to help determine their importance."

per though they remain areas that should be further researched. Next, we briefly review the literature on expectations of returns to education, before providing a link between a real options framework and subjective expectations in sections 2.3 and 2.4.

2.2.2 Subjective expectations of returns to education and information interventions

As noted in the introduction, the usefulness of information interventions rests on the hypothesis that education investment choices are made based on inaccurate expectations of the benefits and/or costs of education. The general importance of considering this possible divergence is highlighted by Manski (2004, pg.1360), who notes that "researchers have many reasons to be interested in the correspondence between subjective expectations and objective realities. Economists invoking rational expectations assumptions should want to know how well such assumptions describe real decision makers." The importance of this question in the context of education choices was highlighted early by Manski (1993) and the answer has clear policy implications: if it is evident that youth have inaccurate expectations about returns to education, policies that provide information and assist youth with making choices (such as careers counselling) might be appropriate, while if expectations are accurate, policy makers should focus on other factors that may be leading to inefficient levels of investment in education.

A general scepticism about the usefulness and reliability of subjective expectations data has meant that, until recently, there have been limited attempts at the collection and analysis of this data (Manski, 2004, Delavande, Giné, and McKenzie,

2011, Attanasio, 2009). [3] The limited research on expectations of the returns to schooling demonstrates that the accuracy of these expectations varies according to country and level of education. Betts (1996) uses survey data of the income expectations of students at the University of California San Diego to find that students "made very good estimates of current salaries of young workers" (p.39), where students average estimate of the wage differential between high school graduates and college graduates is about 57.8% while Betts' estimates the real differential to be 50.9%. Webbink and Hartog (2004) provide evidence from the Netherlands on the correlation between students' expected starting salaries (measured when they were still enrolled in higher education) and what they were observed to earn four years later and find that students have very accurate expectations. Another study, in the USA by Rouse (2004) also finds students to have accurate income expectations. Rouse used Dominitz and Manski's (1996) data, from high income backgrounds, and a small sample of low-income minority high school seniors in Baltimore city to test the hypothesis that youth from low-income backgrounds have lower expectations leading to lower levels of investment in education compared to youth from high income backgrounds. The analysis shows that there is no significant difference in expectations between the two groups of students, leading to the conclusion that expectations cannot explain differences in rates of educational attainment.

As noted by Jensen (2010, p.516) youth in developing countries may have more limited information on the returns to education available when compared to youth in developed countries (where such information is more widely available). Indeed, Jensen (2010) and Nguyen (2008), in the Dominican Republic and Madagascar,

[3]Dominitz and Manski (1996), is a notable early study that demonstrated that students are willing and able to respond to earnings expectations questions in probabilistic form (as opposed to responding just with averages/point estimates).

respectively, find that students' expectations are inaccurate. In the Dominican Republic, for example, there were no published estimates on the returns to education available, nor was there the equivalent of a labour force survey that could be used to estimate them. Given the lack of publicly available information, it is perhaps not too surprising that youth in this context underestimate wages conditional on the completion of secondary school by about 14% while also overestimating wages conditional on primary school completion by 11% (both contributing to reduce the expected net return to secondary school). Such deviations provide a justification for interventions intended to influence education choices through the provision of information. The conclusion of these two studies is that when students are informed about average earnings conditional on different education outcomes they change their expectations (reducing the large gap between students' reported expectations and the information provided) and invest more in education: Nguyen (2008) reports that, for those students with perceived returns lower than the average returns (about which they were informed), school test scores improved significantly, by 0.37 standard deviations, while Jensen (2010) reports an average increase in number of years of schooling of 0.20-0.35 years for those students who receive information. Together, these results support the remaining two assumptions that validate the potential utility of information interventions: students respond to the provision of information by changing their expectations and, as a result (and when feasible), invest in education. [4] In both cases, the information provided is very

[4]As noted in the introduction, in the Dominican Republic, changes in expectations were not followed by changes in education choices among the poorest youth. Similarly (although not in the context of an information intervention), Attanasio and Kaufmann (2009) find, using data from Mexico, that expected returns to schooling and perceived employment probabilities are related to school and college attendance decisions in all but the poorer individuals. They hypothesise that this is due to credit constraints faced by poor students that "break" the link between

simple: average returns to different levels of education. This simplicity is justified on the basis that it would be the easiest type of relevant information that the students are likely to understand (Jensen, 2010, pg. 523). [5] However, in more or less indirect ways, both Jensen (2010) and Nguyen (2008) address the two issues that are central to this chapter, the importance of future options and of heterogeneity in returns to education.

Focusing on the first issue, Jensen (2010) provides information on average earnings conditional on *all* education options, rather than the simplistic approach of providing information only on the next choice available to the beneficiaries of the information intervention. Although this characteristic of this intervention is not framed in terms of sequential nature of education choices, it implicitly raises the option value of different grades, as they open the possibility of further studying. In section 2.5 we explore the consequences of a more explicit treatment of the sequential nature of educational investment for the design of information interventions.

The importance of heterogeneity is addressed differently. For example, Nguyen (2008) tested the relative effectiveness of role models who tell a personal success story (either as substitute or complement to information on average returns to education) and hypothesises that one explanation for their relative lack of effectiveness

expectations and behaviour.

[5]However, in cases where there is less concern about the capacity for youth to analyse a wider range of information, interventions may be more complex. For example, Kerr et al. (2013) evaluates a randomized field experiment of information given to graduating high-school students about the returns to higher education in Finland. Unlike other interventions, there was no need to limit information complexity as in the interventions by Jensen (2010) and Nguyen (2008) and students were given very specific and detailed information regarding employment probabilities and mean wages according to individual degrees. The study found that such information had an impact on the distribution of applications across higher education courses, though it was too early to draw any conclusions about the welfare implications of the intervention.

would be that role models highlight the importance of aspects such as background as determinants of returns to education. This effect was particularly pronounced among poor children, who possibly assumed that success stories of students coming from a wealthier background were not relevant to their own expected future (while being positively influenced by role models who were from a poor background). Given this, this analysis demonstrates that information can change expectations and behaviour but that, importantly, heterogeneity in the returns to education may matter for the design of information interventions.

Contrary to this position, the importance of heterogeneous returns to education is clearly minimized in the design and analysis presented in Jensen (2010). The reason for this neglect is presented as following:

"And although there is likely to be heterogeneity in the returns (say by school quality or race) and students may be aware of that heterogeneity, this alone could not explain why students on average have low expectations for the returns they would personally face. (...) black youths may believe the returns are lower for them, but white youths should then also believe that the returns for them are higher than average, so the average across a representative sample of youths should still hit the correct average return." (Jensen, 2010, p531, footnote 22)

Implicit in this argument is the assumption that heterogeneity in the returns to education should neatly correspond to heterogeneity in expectations. That is, for groups of youth for which the returns to education are high, their expectations of the returns to education are also comparatively high (and vice versa). However, it is possible that students who are advantaged in some way actually have lower than average expectations and that students who are disadvantaged in some way actually have higher than average expectations. If this were the case, it is not clear

ex ante that providing information on the returns to education would necessarily increase the accuracy of expectations and lead to optimal investment in education. In fact, it is possible that providing information on average returns could be counterproductive in such a situation.

This point is best illustrated with a hypothetical example similar to that presented at the start of this chapter. Suppose that the returns to grade 12 and university are on average (70, 180). However, there are heterogenous returns with 50% of students who are 'advantaged' facing returns of (90, 260), while the remaining 50% are 'disadvantaged' and face returns of (50, 100). All of the 'advantaged' underestimate the returns to education for themselves by the same degree with expectations of (80, 200). In contrast, all of the 'disadvantaged' overestimate returns to the same degree and have expectations of (55, 120). However, the 'disadvantaged' under-invest in education because of credit constraints. A researcher that analysed this situation without explicitly measuring group differences, would see average expectations to be (67.5, 160). The researcher would rightly conclude that some youth underestimate the returns to education and they would be correct if they assumed that some youth under-invest in education. However, if they provided information on the average returns to all youth (70, 180), this would lead to an increase in the expectations for 'disadvantaged' students but have no effect on investment because they are credit constrained. While simultaneously, this intervention would lead to lower expectations for the 'advantaged' youth possibly lowering their investment in education (depending on education costs). Importantly, unless heterogeneity was explicitly considered when designing the intervention, the researcher would not be able to predict *ex ante* that this intervention would have a 0% chance of increasing the accuracy of expectations. Neither would it increase in-

vestment in education (as it could only *raise* the expectations of the disadvantaged who are credit constrained).

While this is a highly stylised example, the key point not recognised in Jensen (2010) is that it is not sufficient to consider the average impact of an intervention without considering heterogeneity in the observed and expected returns to education across groups. [6]. Also, if heterogeneity is significant, then a randomised control trial would only give a consistent estimate of the effect of an intervention by sub-group if it were randomised at that level and the information was appropriately targeted. [7]

Nguyen (2008, p.7-8) considers the relationship between the population average of the return to education and its use by individuals when there is heterogeneity in the returns. It is assumed that all individuals are aware of the relationship between the population average and their own expected return (which is a linear function of individual characteristics and the population average). This assumption remains untested. It may be equally reasonable to assume that rather than the national population average, individuals expected return should be modelled based on the average return for a particular demographic group (for example, according to gender, race, age, location etc.). Deciding on one particular statistic on which individuals' base their expectations (whether it is a national average, local average, or average for a race group) is arbitrary but critically, it is by no means innocuous. As the numerical example about heterogenous returns across

[6]Though the realistic possibility of disadvantaged/poor youth facing credit constraints and not being helped by information *is* something explicitly recognised by Jensen (2010).

[7]Or if the idea of giving different information to different students seems unfair, all students could be explicitly made aware of differences in the returns to education due to demographic, individual or other factors and be provided with a range of information on the returns to education conditional on these differences.

groups demonstrates, making assumptions about the representativeness and appropriateness of averages may lead to a misunderstanding of the likely impact of an intervention and particularly, the possibility that the intervention might be harmful.

Finally, one other problem with existing research on information interventions is that they generally focus on increasing investment in education rather than increasing the accuracy of expectations and the optimality of investment decisions. This is important because if the market return to education is very low for some students, increasing their level of investment in education may not be efficient. Increasing investment *per se* should not be the goal of policy; rather, the goal should be increasing the efficiency of choices (which would implicitly mean increasing investment in education for those that currently invest less than what they should). [8]

2.3 Real Options Pathways for Education (ROPE): a model of investment in education

This section presents a model of educational investment for those who have just completed grade 11 and can either enrol in grade 12 or enter the labour market.

[8] While we only consider the private efficiency of choices and ignore the public benefits of education, this should not be a problem for our analysis in sections 2.3 and 2.4 and our later assumption that both underinvestment and overinvestment in education may be a problem. If it's the case that the privately efficient level of investment in education is less than the socially optimal level of investment in education, attempting to mislead students into overvaluing the private benefits of education so that they increase their investment is arguably not an ethical or reliable solution to the problem. Rather, it could be argued that it would be better for the government to consider ways of incentivising educational investment so that the privately efficient investment level matches the socially efficient investment level.

We focus on a very stylized version of this decision, in order to present the main differences that originate from the consideration of the sequential decision of this investment and postpone the consideration of the specificities of the South African context, in particular regarding its education system and labour market, until the next section. We will refer to this model as Real Options Pathways for Education (ROPE), as it emphasises the fact that investment in grade 12 can be made in consideration of the options for higher education that open upon its completion.

The purpose of ROPE is to provide a framework in which the impact of information interventions may be considered. Specifically, information on monthly earnings conditional on different levels of education, provided to individuals that have just finished grade 11. Frameworks similar to ROPE have been suggested as tools to assist decision making in other contexts. For example, Hertzler (2007) presents a series of "decision diagrams" that may be used by farmers when making production decisions in the face of new risks stemming from climate change. Also of relevance is the use of diagrams in support of the sequential educational choice model of Breen and Goldthorpe (1997). They investigate differences in educational attainments according to class and gender as a function of the existence of heterogeneity in access to resources for educational investment according to class (that impact on educational choice through the rational evaluation of costs and benefits).

The complete decision problem of the individual and all the possible outcomes are linked together in this diagram (figure 2.1), which includes the probabilities of pass and failure for grade 12 and higher education (here, labeled as grade 13). Decisions are represented as solid lines while dotted lines indicate events (passing or failure of a grade) with the probabilities of these events specified within brackets.

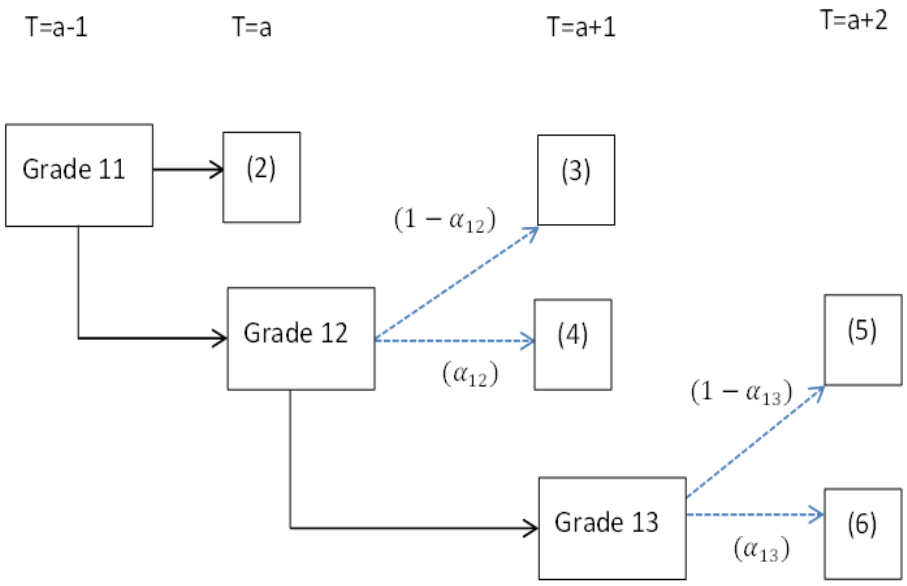

Figure 2.1: Real Options Pathways for Education (ROPE)

The boxes, numbered 2-6, correspond to possible final outcomes.

At the start of each time period, individuals can choose whether to continue in school or to join the labour market. We assume that once they enter the labour market they do not return to education and when in education they can only decide to leave or continue to the next level at the start of the next time period. The outcome of a year of education is binary, with pass/fail as possible outcomes. We further assume that utility is a linear function of income only and that agents are risk neutral. [9] The model can be solved recursively to arrive at the income maximising choice between investing in grade 12 and leaving school after grade 11.

At the start of the first time period (that is, having completed grade 11), individuals maximise expected lifetime utility by deciding to either continue in school (for which they expect utility of $E(U12)$), or to enter the labour market (for which they expect utility of $E(UL)$).

$$E\left(U\right) = Max\ \left\{\ E\left(UL\right),\ E\left(U12\right)\right\} \tag{2.1}$$

Expected utility for one year in the labour market would be a function of the full time employment probability (EP_{11}) and expected wage conditional on the completion of grade 11 (W_{11}). Expected lifetime utility from entry into the labour market is given in equation 2.2 as a function of the discounted sum of expected labour market returns for duration of her/his career, given by the difference between the minimum of retirement age and life expectancy (y) and the age at which they complete grade 11 (a). Subscript t denotes the age of the individual in each

[9] We assume risk neutrality for simplicity but it is important to note that risk neutrality is not synonymous with risk being unimportant. For example, Hogan and Walker (2007) find that risk affects educational investment independently of the degree of risk aversion.

year of their work career from the time they leave school. R is the discount rate and g is an expected constant growth rate in wages. [10]

$$E\left(UL_2\right) = \left(\sum_{t=a}^{y} EP_{11t} \times W_{11t} \ \times \frac{(1+g)^{y-a}}{(1+R)^{y-a}}\right) \tag{2.2}$$

Expected utility from continuing onto grade 12 depends on different possibilities. Firstly it depends on the expected wage and employment probability if grade 12 is failed, (which occurs with probability $(1-\alpha_{12})$). We assume that this expected wage and employment probability are the same as for entering the labour market with grade 11 (W_{11} and EP_{11} , respectively). Expected utility from this outcome is expressed in equation 2.3, which differs from equation 2.2 because it includes the direct cost of grade 12 (C_{12}) and the opportunity cost associated with one extra year of education (as the sum of wages is now between *a+1* and y).

$$E\left(UL_3\right) = \left(\sum_{t=a+1}^{y} EP_{11t} \times W_{11t} \ \times \frac{(1+g)^{y-a-1}}{(1+R)^{y-a-1}}\right) - C_{12} \tag{2.3}$$

Those who successfully complete grade 12 (with probability α_{12}) can opt between entering the labour market with a different expected wage and employment probability (W_{12} and EP_{12}, respectively), as expressed in equation 2.4.

$$E\left(UL_4\right) = \left(\sum_{t=a+1}^{y} EP_{12t} \times W_{12t} \ \times \frac{(1+g)^{y-a-1}}{(1+R)^{y-a-1}}\right) - C_{12} \tag{2.4}$$

[10]For simplicity, we assume that the growth rate in wages is equal and constant for all levels of education and is equal in magnitude to the discount rate. Extensions to the model could easily incorporate both a discount rate and a growth rate in wages that vary. Also, when considering the presentation of data to youth we assume that nominal mean monthly earnings would be provided to youth and that they would use this information to infer total career earnings for each educational pathway. While youth may not think exactly in correspondence with ROPE, it is designed to be an approximation of the sequential choice framework they would use.

or, alternatively, can invest in higher education. For simplicity we assume that there is only one higher education option, that lasts one year. This option has a direct cost of C_{13} and is associated with W_{13} and EP_{13} as the expected wage and expected employment probability, respectively. Two outcomes are then possible, as the student may fail higher education (with probability $(1-\alpha_{13})$) or, alternatively, pass. Equations 2.5 and 2.6 represent the lifetime earnings associated with each of these outcomes, respectively.

$$E\left(UL_5\right) = \left(\sum_{t=a+2}^{y} EP_{12t} \times W_{12t} \ \times \frac{(1+G)^{y-a-2}}{(1+R)^{y-a-2}}\right) - (C_{12} + \frac{C_{13}}{(1+R)^2}) \quad (2.5)$$

$$E\left(UL_6\right) = \left(\sum_{t=a+2}^{y} EP_{13t} \times W_{13t} \ \times \frac{(1+G)^{y-a-2}}{(1+R)^{y-a-2}}\right) - (C_{12} + \frac{C_{13}}{(1+R)^2}) \quad (2.6)$$

Youth will choose an educational investment pathway with the greatest expected return. Clearly, no matter how low the cost of grade 12 is, or high the return to grade 12 is (and the higher education option), as long as there is the possibility of failing grade 12, there is the possibility of a lower utility outcome than if the individual entered the labour market immediately following grade 11. [11] That is, the following condition will always hold: $E(UL_3) < E(UL_2)$.

Also of significance is the inclusion of the option value of tertiary education if the individual finishes grade 12 successfully which can be separated from its direct value. Equation 2.7 expresses the direct net value of grade 12, ED (a net value because it is the added value of grade 12 compared to grade 11), while equation 2.10 expresses the option value, EO, (the pathway value). Equation 2.16 expresses the total value of grade 12, ET which is equal to $ED + EO$.

[11] This result reflects the existence of opportunity costs of education and it holds even if the direct costs of education C_{12}, are zero.

$$ED = (1 - \alpha_{12}) \times E(UL_3) + \alpha_{12} \times E(UL_4) - E(UL_2) \tag{2.7}$$

$$EO = [\alpha_{12} \times Max\ \{E(UL_4),\ (1 - \alpha_{13}) \times E(UL_5) + \alpha_{13} \times E(UL_6)\}] - [\alpha_{12} \times E(UL_4)] \tag{2.8}$$

$$ET = (1 - \alpha_{12}) \times E(UL_3) + \alpha_{12} \times$$

$$Max\ \{E(UL_4),\ (1 - \alpha_{13}) \times E(UL_5) + \alpha_{13} \times E(UL_6)\} - E(UL_2) \tag{2.9}$$

As noted by Weisbrod (1962) and Heckman, Lochner, and Todd (2006), estimates of returns to grade 12 obtained through a model without option values would be downward biased. In addition, as we will address in more detail in section 5, an information intervention that only provides information on outcomes 2-4, would not be enough to decide on the optimal investment decision, unless one is willing to assume that individuals were perfectly informed about outcomes 5-6. If this assumption were incorrect, then no matter how effective the intervention was in correcting expectations in 2-4, individuals may still not make the utility maximising investment decision following the intervention. In short, the consideration of the sequential nature of educational choices requires the provision of information on *all* possible educational outcomes.

In addition it is clear that providing information only on expected earnings, will impact individuals' decisions differentially, depending on their probabilities of passing different levels of education and also on employment probabilities (even if information is provided on earnings conditional on all possible levels of education). For example, someone who is 100% certain that they would fail grade 12 would not be influenced by information on earnings conditional on grade 12 or the higher

education option. Finally, nothing has been said so far about the nature of these different quantities used to model this decision other than that they are conditional on different educational levels. In particular, we have not yet taken into account the possibility of other wage determinants, and said nothing about the nature of the probabilities used to guide decisions through the different levels of education. In the next section we take these steps, using data from South Africa.

2.4 Using ROPE in the South African context

The decision model presented in figure 2.1 was deliberately simplified to highlight the main elements of a model that incorporates the sequential nature of educational investments. However, its use in an analysis of the possible effects of information campaigns on such decisions requires a more realistic description of the decision tree facing individuals and information on expected labour market outcomes. In this section we specify the model for the education system and labour market of South Africa.

There are three important features of the South African context that make it especially interesting. Firstly, given the high level of income inequality (Leibbrandt and Woolard, 2001) and youth unemployment (Kingdon and Knight, 2007, 2004) in South Africa, it is not surprising that the economic returns to education are a central motivation in education choices (Cosser and du Toit, 2002). In addition to education being valued for its association with higher earnings, higher levels of education are also valued because they are associated with increased probability of employment (Branson et al., 2012). Secondly, there is a pronounced convexity in the return structure to education (Keswell and Poswell, 2004). Using data from a national household living standards survey in 1993, Mwabu and Schultz (2000) find

that for all races, the earnings premiums for higher education are greater than the they are for secondary education and primary education. This pattern of returns suggests that the option values of higher education may be especially important to the decision to invest in grade 12.

Lastly, there are pronounced differences across racial groups in both educational achievement and average earnings, with Whites earning more (and being more educated) than Coloureds who in turn earn more (and are more educated) than Africans (Keswell and Poswell, 2004). In addition, there are differences in the returns to education between races. Interestingly however, while Whites earn more than other races for any given level of education, there is evidence that from the last years of apartheid until now, *ex post* relative returns to secondary and higher education have been consistently greatest for Africans. For example, Mwabu and Schultz (2000, p22) estimate that tertiary education earnings premiums for African men and women were 29% and 40% per year, respectively, while they were 15% and 14% per year for Whites and about 19% and 30% for Coloureds. Similarly, Branson et al. (2012, p8) using annual data from the Labour Force Survey, estimate that in 2007, African men with two years of post-secondary schooling were earning 70% more than African men with only grade 12 completion. But despite the relatively high tertiary education earnings premiums, the proportion of individuals completing tertiary education has not dramatically increased across generations. For example, Branson et al. (2012, p21) estimate that 10.2% of males born in 1953-1955 have a tertiary qualification while 9.8% of males born in 1983-1985 have a tertiary qualification. There are several possibilities for why this might be the case. Two are of particular interest in this chapter. Firstly, youth may underestimate the very high rates of return to tertiary education. Secondly, even if youth are

aware of the earnings premiums, tertiary education may be difficult to enter and to complete. The *ex ante* expected return to tertiary education for a grade 11 student must account for the possibility of failure to pass in grade 12 and failure to pass in tertiary study.

One characteristic of South Africa's higher education system that is particularly relevant to this discussion is the wide range of options accessible to youth once they complete high school. Students can enrol in one year diplomas or technical certificates, two year diplomas or technical certificates and undergraduate degrees (three or four years in length).[12] The choice between these different degrees is dependent on the individual's results at the end of grade 12. Those enrolling in grade 12 may "pass with exemption", "pass without exemption" or fail. Passing with exemption is required for some higher education options, namely three and four year university degrees, but it is not required for the lower levels of higher education.

For each year of study in higher education, there is a chance of failure. For simplicity we assume that if the student fails a year s/he enters the labour market in the following year. We will also assume that incomplete education (such as passing the first year of a bachelor degree but failing its second year) has no effect on expected wages and employment probability, which are then assumed to be identical to the highest concluded degree: for example, a youth with incomplete higher education will expect to receive the wage that he would have received had

[12]There are also postgraduate degrees, such as masters degrees. However, we ignore this possibility as we have no information on such degrees in the data we use. No individuals in our sample are recorded as having completed more than 4 years of post-high school education. Also, we do not distinguish between different courses and subject areas within higher educational categories (for example, we do not distinguish between a 3 year bachelor of arts degree and a 3 year bachelor of science degree).

s/he entered the labour market after completing year 12.

To specify the ROPE model for South Africa, we need information about pass probabilities, employment probabilities and average wage levels conditional on different levels of education. We use data from the Cape Area Panel Study (CAPS) which is a longitudinal study of a random sample of 4758 youth aged 14-22 in 2002 living in the Cape Town Metropolitan Area. [13] The study focuses on a wide range of important outcomes for youth during the transition to adulthood, including schooling and labour market participation. All youth were initially interviewed in 2002 (wave 1) and then in 2003-2004 (wave 2), 2005 (wave 3) and 2006 (wave 4). Importantly for our discussion of the potential impact of information campaigns, respondents were asked about their subjective expectations of wages, through the question "What is the typical take-home monthly wage for other people like you (same age, education, and skills) who have full-time jobs?". We restrict our analysis to wave 3, as this is the only wave for which this question was asked of the whole sample. One limitation of the data is that each individual was only asked for expectations regarding their own level of education. I assume however, that conditional on race and education, individuals are the same in all other regards. Later, in sections 2.6.1 and 2.6.1 we consider the possible importance of gender and self-selection into employment. Another limitation of the data is that the sample is not representative of youth around South Africa. This is particularly in regards

[13]The Cape Area Panel Study Waves 1-2-3 were collected between 2002 and 2005 by the University of Cape Town and the University of Michigan, with funding provided by the US National Institute for Child Health and Human Development and the Andrew W. Mellon Foundation. Wave 4 was collected in 2006 by the University of Cape Town, University of Michigan and Princeton University. Major funding for Wave 4 was provided by the National Institute on Ageing through a grant to Princeton University, in addition to funding provided by NICHD through the University of Michigan. See Lam, Seekings, and Sparks (2008).

to proportions of each race in the sample. While Coloureds are the largest group in the sample this is only because they are over-represented in Western Cape province compared to the rest of the country and CAPS was designed to be representative of that province.

Table 2.1 presents some summary statistics for observed and expected earnings, by education level and race. We consider only youth aged 17-25 (which represents 98% of the respondents in wave 3 of CAPS) because focusing on this age range ensures that all respondents are old enough to have finished grade 11 and that some are old enough to have completed a 4 year degree. For the subjective expectations data, we include both employed and unemployed youth, and both youth that are still in school and those that have left school. To make it comparable with the subjective expected wage, we restrict the analysis of observed earnings to full-time workers, who represent 78% of those who worked any number of hours. [14] Earnings data in CAPS was reported as monthly net wages measured in Rands, which we simply multiply by 12 to estimate expected yearly earnings. [15]

Table 2.2 presents the probability of the two events we consider, namely finding

[14]There is no official definition of full-time work in South Africa, nor is full-time a term that is defined explicitly in CAPS. We define a full-time worker as someone who works 140 hours a month or more, corresponding to working 7hrs a day, 5 days a week. We also compute the rate of "not employed full-time", rather than a more conventional unemployment rate.

[15]It should be noted that the number of observations for Africans with 3 years of higher education that are working is zero. However, unless the pathway for 3 years of higher education were in reality more valuable for these youths than the pathway for 4 years of higher education (and this seems unlikely as for Whites and Coloureds it is not), this is inconsequential to our estimates of the net value of grade 12 in the ROPE framework. A more critical limitation of the version of ROPE we specify for South Africa is that we ignore all direct costs of education. As the direct costs of university education are substantial in South Africa, many youth may not be able to afford to go to university and so may not consider the university pathways when valuing grade 12. This issue is addressed later in the paper, in section 2.6.4.

Table 2.1: Summary statistics: observed and expected monthly earnings in Rands

	Observed			Expected			Over-
Grade	Obs	Mean	SD	Obs	Mean	SD	estimation
All races							
11	82	1756	780	315	2217	1553	**26.3%**
12	395	2793	1966	670	3328	2062	**19.2%**
13	21	3160	2242	36	3383	1363	**7.1%**
14	11	3092	1786	33	5061	3440	**63.7%**
15	10	4408	3153	68	5175	3000	**17.4%**
16	9	11611	9955	16	10564	8330	**-9.0%**
African							
11	40	1469	709	186	2133	1459	**45.2%**
12	105	2076	2050	271	3082	2134	**48.5%**
13	2	5250	6010	11	3836	1614	**-26.9%**
14	2	3300	424	15	5647	3999	**71.1%**
15	0	0	0	15	5520	3309	**0.0%**
16	2	10000	7071	4	9875	7554	**-1.3%**
Coloured							
11	38	1977	748	103	2369	1750	**19.8%**
12	242	2713	1331	332	3140	1829	**15.7%**
13	15	3057	1897	17	2947	993	**-3.6%**
14	5	2500	1127	12	3858	2915	**54.3%**
15	6	4097	4083	23	4839	3138	**18.1%**
16	2	6000	1414	4	6506	5793	**8.4%**
White							
11	4	2525	709	26	2219	1378	**-12.1%**
12	52	4613	2954	67	4261	2549	**-7.6%**
13	4	2500	1061	8	3688	1557	**47.5%**
14	5	3600	2584	6	6000	2588	**66.7%**
15	4	4875	1250	30	5260	2806	**7.9%**
16	5	14500	12560	8	12938	9682	**-10.8%**

Table 2.2: Event probabilities

	All		African		Coloured		White	
	Obs	%	Obs	%	Obs	%	Obs	%
Grade 12, fail	259	23.2%	119	37.0%	98	15.3%	42	2.4%
Grade 12, pass, no exempt.	325	54.8%	134	52.7%	126	65.9%	65	33.0%
Grade 12, pass, with exempt.	325	22.0%	134	10.3%	126	18.8%	65	64.6%
Tertiary, pass	379	88.7%	111	77.5%	155	93.6%	113	92.9%
Employ. prob. grade 11	227	40.5%	129	32.6%	91	50.6%	7	57.1%
Employ. prob. grade 12	889	53.0%	299	39.1%	477	61.2%	113	54.5%
Employ. prob. grade 13	46	60.9%	7	42.9%	28	71.4%	11	45.5%
Employ. prob. grade 14	26	57.7%	9	33.3%	11	63.6%	6	83.3%
Employ. prob. grade 15	29	41.4%	3	0.0%	10	70.0%	16	31.3%
Employ. prob. grade 16	12	75.0%	3	66.7%	4	50.0%	5	100.0%

a job (employment probabilities) and successfully completing a grade (passing probabilities). The full-time employment rate was calculated as those employed full-time divided by the total number of youth who are not studying. We do not use a traditional definition of unemployment as there is evidence in South Africa of a large number of discouraged workers (Kingdon and Knight, 2000). Pass probabilities for each educational level were calculated using data on the proportion of students in grade 12 and in tertiary education that successfully completed their level. For example for grade 12, the pass probability is the proportion of those students who had been enrolled in grade 12 in the previous year and that had reported grade 12 as having being completed in the current year. Strictly speaking, the pass probability is a 'completion probability' as it counts those students that drop out as students that 'fail'. The probability of students who pass grade 12 with

exemption was taken from a CAPS question that specifically asks students in wave 3 if they have such an exemption or not. However, we cannot distinguish between pass rates for different years of tertiary education and are forced to assume that the average pass rate is the same for all years of tertiary education and for all types of tertiary education. Finally, we assume that youth are aged 17 in year 1 and that life expectancy at birth is 56.3 (the life expectancy in the year 2000, the year of the national census). [16] This information will allow us to estimate expected benefits for all the different possibilities in an expanded version of the ROPE model. [17]

The summary statistics presented in table 2.1 reflect what is known regarding the structure of returns to education in South Africa. Wages for those who finished grade 12 are significantly greater than for those who only completed grade 11, less than the wages earned by those with 1-3 years of tertiary education, and significantly smaller than for a 4 year degree. Also, while wages for African youth

[16] As this assumption about the length of career corresponding neatly to life expectancy is arbitrary, we consider changes to this assumption in a sensitivity analysis presented in section 2.6.3. While life expectancy may differ sharply by race in South Africa, we use the non-conditional life expectancy as there are no official statistics on life expectancy by race. Also, as reported in section 2.6.3, changing the life expectancy makes no difference to the results.

[17] One potential problem with our sample regards self-selection of those who are not working, as they may have different ability, preferences and expectations from those who are working. We address this issue in section 2.6.2 and find that restricting the sample to only those that are employed has no substantial consequence for the results and conclusions. Another possible issue regards self-selection of those with different levels of education. Expectations of earnings for each grade are measured only for individuals with that level of education. This could be important if individuals with different ability levels invest differently in education (for example, they may have different expected and actual pass rates). For simplicity we assume that the only cause of heterogeneity in expectations and observed returns other than grade of education completed is race. However, in section 2.6.1 we consider the alternative assumption that gender drives heterogeneity and find that our conclusions are unaffected by this changed assumption about the source of heterogeneity.

that finish a 4 year degree are lower than for White youth (R10000 versus R14500), the percentage wage premium for a 4 year degree relative to grade 12 is higher for African youth than it is for White youth (481% versus 314%). In addition, and as shown in table 2.2, higher levels of education are also associated with higher probabilities of working full time.

Comparing expected and actual earnings for each race group in table 2.1, it is possible to conclude that, conditional on race, we find both cases of average overestimation and average underestimation of returns to education: for example, Coloured youth with grade 12 overestimate the returns to this grade by 16% while Coloured youth with 13 years of education underestimate the returns to this level by 4%. Across all race groups and grades, average overestimation ranges from -27% to 71% and this combination of of overestimation and underestimation for each race group makes it *a priori* unclear whether the overall net value of grade 12 in a ROPE model would be an overestimate or underestimate. Also, the large range in magnitude of overestimation and underestimation may be due in part to small sample sizes for observed wages and expected wages for some levels of education. For example, while overestimation of 14 years of education for Africans is measured to be 71%, this may be because the two African individuals that are working with this level of education just happen to be earning less than average. While the accuracy of the estimates may be a problem because of small samples, we do not have reason to believe that the estimates are biased. It is also important to note that an information intervention that addresses inaccuracy in expectations for only some levels of education might leave inaccuracy in the expectations of other levels of education unchanged.

It might be argued that the relationship between earnings and education would

be made clearer by OLS or IV estimation. However, this chapter does not use econometric estimates of the returns to education in the ROPE framework for a simple reason- we want to consider information interventions that could be easily communicated to youth with little or no training in econometrics. While average earnings conditional on education could be communicated to youth in high-school, there would be no easy way of communicating OLS or IV estimates without gross simplification (which may make the information misleading). Hence, we follow the precedent set by Jensen (2010) and consider average observed earnings as the information that is communicated to youth.

Figure 2.2 summarises the pathways available to students in South Africa, under the assumptions discussed and using the probabilities and wage estimates presented in tables 2.2 and 2.2, with estimated expected lifetime earnings reported in Rands. Each year after the completion of grade 11 is characterized by a choice to join the labour market (work) or to invest in another year of education. The first column for each year constitutes the set of possible choices for that year, conditional on the outcome of the previous year, with the links being indicated by arrows. If the last year was a period of study that was either failed or was the final year of a degree being passed, the individual would choose to enter the labour market in the proceeding year. If the the previous year was grade 12 or in a non-terminal year of a degree, the individual can continue studying in the following year or can enter the labour market. The outcomes of years where the individual enters the labour market are specified by an expected lifetime earnings equation, with the general form specified in equation 2.10 and where s is the number of years spent in education since grade 11, x is the highest grade completed and $g = R$.

$$E\left(UL_x\right) = \left(\sum_{t=a+s}^{y} EP_{xt} \times W_{xt} \ \times \frac{(1+g)^{y-a+s}}{(1+R)^{y-a+s}}\right) \tag{2.10}$$

While there are many pathways the model can easily be solved recursively to arrive at values for the overall expected returns to grade 11 and grade 12.The numbers presented in figure 2.2 are based on actual earnings unconditional on race and while not presented, the same model is also specified for each race group individually and for expected earnings rather than observed earnings (for a total of 8 specifications). As an example of how the values of the outcomes were calculated, consider the number representing the lifetime earnings that should objectively be expected if an individual passes a 4 year degree and enters the labour market (R3 584 350, presented at the bottom right corner of figure 2.2). Using equation 2.10, and the data from tables 2.2 and 2.2 for mean annual wage (116111×12), employment probability (0.75) and with a career length of life expectancy (56.3) minus age at degree completion (22), yields 2.11:

$$E(UL_{16}) = R11611 \times 12 \times 0.75 \times (56.3 - 22) = 3584350 \tag{2.11}$$

To find the optimal pathway from grade 11 completion onward, an individual would solve the general ROPE equation 2.11, where the return to grade 12 is expanded to equation 2.12:

$$E(U12) = 0.23 \times E(UL_{11})$$

$$+0.55 \times Max\{E(UL_{11}), T1, T2\} + 0.22 \times Max\{E(UL_{11}), T1, T2, T3, T4\} \tag{2.12}$$

In this equation, T1-T4 represent the returns to each level of tertiary education. T1-T3 are presented in equations 2.13, 2.14, and 2.15, respectively. While T4 is not explicitly presented in the interests of space, it follows the exact same pattern

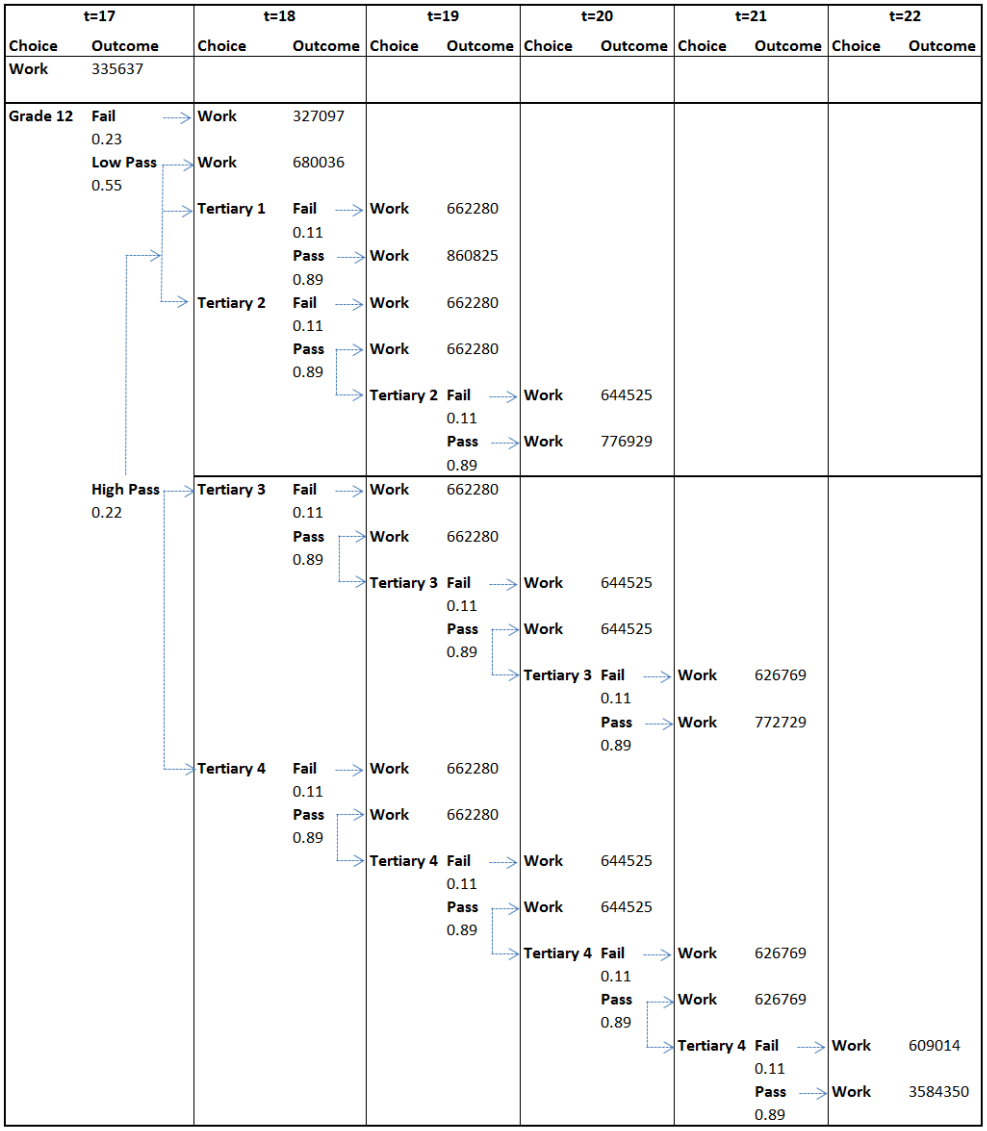

Figure 2.2: ROPE Specified for the South African Context

as T3, except with four points where the individual considers the maximum of either entering the labour market or continuing to study.

$$T1 = 0.11 \times E(UL_{12}) + 0.89 \times E(UL_{13}) \tag{2.13}$$

$$T2 = 0.11 \times E(UL_{12}) + 0.89 \times Max\{E(UL_{12}), 0.11 \times E(UL_{12}) + 0.89 \times E(UL_{14})\} \tag{2.14}$$

$$\begin{aligned} T3 = {} & 0.11 \times E(UL_{12}) + 0.89 \times Max[E(UL_{12}), 0.11 \times E(UL_{12}) + \\ & 0.89 \times Max\langle E(UL_{12}), 0.11 \times E(UL_{12}) + \\ & 0.89 \times Max\{\{E(UL_{12}), 0.11 \times E(UL_{12}) + 0.89 \times E(UL_{14})\}\rangle] \end{aligned} \tag{2.15}$$

Table 2.3 presents, for the whole sample and then for each race group, the values for the return to grades 11 and 12 estimated in the ROPE framework using observed and expected wages. The values for observed wages come from the full-time working sample for each race, while the sample used for expected wages includes both those who are working and those who are not working (the samples in table 2.1). The grade 11 total value is the value of entering the labour market with grade 11. The grade 12 total value is composed of the value of entering the labour market with grade 12 (grade 12 direct value) plus the value of creating the option of enrolling in higher education (grade 12 pathway value). The equations used to calculate these values are the same as equations 2.7, 2.10, and 2.10, except they are expanded to include the greater number of options in the South African context. This pathway value is also presented as a percentage of the total value

of grade 12 (which is equal to pathway value divided by the total value). The last row of table 2.3 presents the difference between the total values of grade 11 and grade 12 (grade 12 total net value), which we assume is the criterion used in the investment decision rule. Finally, for each race and grade, we compare observed and expected values through the calculation of an overestimation percentage.

Table 2.3: Returns to grade 12 in the ROPE framework

	Observed	Expected	Overestimation
All races			
Grade 11 total value	335637	423844	**26.3%**
Grade 12 direct value	598260	718377	**20.1%**
Grade 12 pathway value	477701	523780	**9.6%**
Grade 12 total value	1075960	1242156	**15.4%**
Pathway value (%)	44.4	42.2	**-5.0%**
Grade 12 net value	740323	818312	**10.5%**
African			
Grade 11 total value	225573	327580	**45.2%**
Grade 12 direct value	316559	467373	**47.6%**
Grade 12 pathway value	344587	157908	**-54.2%**
Grade 12 total value	661145	625280	**-5.4%**
Pathway value (%)	52.1	25.3	**-51.5%**
Grade 12 net value	435572	297701	**-31.7%**
Coloured			
Grade 11 total value	471330	564739	**19.8%**
Grade 12 direct value	716745	832383	**16.1%**
Grade 12 pathway value	199284	187977	**-5.7%**
Grade 12 total value	916029	1020360	**11.4%**
Pathway value (%)	21.8	18.4	**-15.3%**
Grade 12 net value	444699	455622	**2.5%**
White			
Grade 11 total value	680417	598021	**-12.1%**
Grade 12 direct value	1142814	1054864	**-7.7%**
Grade 12 pathway value	2346794	2354386	**0.3%**
Grade 12 total value	3489609	3409250	**-2.3%**
Pathway value (%)	67.3	69.1	**2.7%**
Grade 12 net value	2809191	2811229	**0.1%**

The results in table 2.3 show that there is considerable variation in returns to education along race lines, as expected. The pattern from highest to lowest

returns is from White, to Coloured to African and this is true both for expected and observed net returns. The investment decision for grade 12 is made based on the total net returns of grade 12. Based on this criterion, the average individual should want to invest in grade 12 regardless of their race or gender, as long as the costs of education are lower than this return. [18]

This conclusion holds even after decomposing the total value of grade 12 to its direct and pathway values, which are R598 260 and R477 701, respectively. The pathway value corresponds to 44% of the total value and increases the difference between the returns to grade 11 and grade 12. With no tertiary options considered, grade 12 would be worth R262 623 more than grade 11. However, including the pathway value means that grade 12 is worth R740 323 more than grade 11. Similarly, the expected net return to grade 12 is always positive (table 2.3), though for African youth it is underestimated by about 32%. This is because they underestimate the pathway value of grade 12. If we were to only consider expectations regarding the direct value, we would not have reason to fear that African youth might under-invest in education. This is because African youth overestimate the direct return to grade 12 by 47.6%. These results suggest two conclusions. The first is that the pathway value of grade 12 is always important and information campaigns that neglect the sequential dimension of educational

[18]If race is ignored and all the data is pooled together to estimate the net return to grade 12 in the ROPE framework, overestimation percentage is estimated to be 10.5%. This may seem peculiar given that overestimation percentage is lower for each race individually but it is important to understand that ROPE estimates for all are not just a weighted average of the ROPE estimates by race. The observed and expected mean earnings for each level of education by race do indeed form weighted averages that are *inputs* into the ROPE model for all. However, the model is not a simple linear combination of these inputs and so there is no reason why the net value of grade 12 should correspond to what it is for the ROPE model by race.

decisions risks undervaluing education. The second is that overall, while White and Coloured youth in our sample seem to have fairly accurate expectations of the net return to grade 12, there may still be motivation for an information intervention to reduce underestimation (and presumably under investment in education) among African youth. In the next section we will consider different approaches to information interventions and the possible consequences of these approaches.

2.5 Using ROPE to evaluate the potential impacts of information interventions

In this section we consider the impact of returns to education of four information interventions that might possibly be considered, *ex ante*, as ways of improving the accuracy of expectations and, potentially, improving education investment choices. All four interventions involve the provision of mean monthly earnings conditional on years of education completed. Also, we only consider information interventions that would occur at the end of grade 11, when youth need to make a choice about whether to stay in school or enter the labour market. [19] The first intervention we consider consists of the provision of information on mean monthly wage in a full time job, conditional on completion of grade 11 and on completion of grade 12. We label it the "simple intervention" as it does not take into account heterogeneity in the returns to education or the pathway value of grade 12. The second intervention we consider consists in the provision of information on mean monthly wages in a full time job to all higher education options in addition to the information included

[19]It must be emphasised that the four interventions proposed are meant to be fairly simple interventions that correspond to the designs of previous real-world interventions. That does not imply that any of these interventions are ideal, only that they are all practical."

in the "simple intervention". However, we still do not consider the possibility of heterogeneous returns to education. Because its structure parallels the intervention used by Jensen (2010), we label it the "standard intervention". Thirdly, we consider a "targeted intervention" which, like the "simple intervention", provides information only on wages conditional on completion of grade 11 and grade 12 but takes into account possible differences across races. Lastly, we consider the "full intervention", that provides youth with information on average wage in a full time job for all grades (including higher education) conditional on race. This intervention combines consideration of the full range of education options (as in the standard intervention) with consideration of heterogeneity (as in the targeted intervention). The information provided in each intervention by race is reported in table 2.4. As can be seen here however, the information given in the simple and standard interventions is the same for all races as these interventions do not provide mean earnings conditional on race. [20]

We are interested in evaluating these interventions in terms of whether they improve the accuracy of subjective expectations. The criterion we use does not

[20]It should be noted that is not totally clear from the question eliciting expectations in CAPS if participants specifically reported a mean expected wage, though we assume that they did (as they were asked to report a 'typical take-home monthly wage'). It is possible though, that instead of a mean wage participants reported a median wage. For this reason, we also tested the provision of median earnings rather than mean earnings for all interventions. There was no qualitative difference to the results except in one case of the full intervention (when there is no perfect updating of expectations). However, for the proceeding analysis we use only the mean rather than the median for two key reasons. Firstly, expected earnings in the ROPE model are based on mean earnings, not median earnings, with the assumption that youth are risk neutral so they would have no reason to care about the median. Secondly, we assume that all youth by grade 11 understand the concept of an average monthly wage but may not be familiar with the concept of a median wage. Any provision of such information would need to be accompanied by an explanation of what a median wage is.

Table 2.4: Contents of information interventions

	Grade 11	Grade 12	1yr dip.	2yr dip.	3yr degree	4yr degree
Simple intervention						
African	R1756	R2793				
Coloured	R1756	R2793				
White	R1756	R2793				
Standard intervention						
African	R1756	R2793	R3160	R3092	R4408	R11611
Coloured	R1756	R2793	R3160	R3092	R4408	R11611
White	R1756	R2793	R3160	R3092	R4408	R11611
Targeted intervention						
African	R2133	R3082				
Coloured	R2369	R3140				
White	R2219	R4261				
Full intervention						
African	R2133	R3082	R3836	R5647	R5520	R9875
Coloured	R2369	R3140	R2947	R3858	R4839	R6506
White	R2219	R4261	R3688	R6000	R5260	R12938

specify anything about the magnitude of the change in expectations, only the direction. We define an improvement in the accuracy of expectations (and hence intervention success) as when the following criteria are met[21]:

1. That the average overestimation of the net value of grade 12 for groups that overestimate this value is reduced following the intervention and is equal to or greater than zero following the intervention. That is, average overestimation does not become underestimation. For example, if a group had a 2% overestimation before the intervention but a 2% underestimation following the intervention, this criterion would not be met.

2. That the average underestimation of the net value of grade 12 for groups that underestimate this value is reduced following the intervention and is equal to or less than zero following the intervention. That is, average underestimation does not become average overestimation.

3. For groups that do not underestimate or overestimate the net value of grade 12, following an intervention they should not underestimate or overestimate the net value of grade 12.

Because our goal is to consider information that could be explained to youth at the end of high school, we need to make some assumptions regarding the types of

[21]As we cannot directly measure the efficiency of decisions without information on costs, we cannot know how many agents will change their decisions following an intervention, regardless of whether or not it meets the criteria we specify. However, if an intervention were to meet the criteria we specify and there is no variance in expectations within groups, then that intervention would be a Pareto improvement: It would potentially make some groups better off without making any groups worse off. If our criteria are not met by an intervention, there would be no easy way of establishing Pareto efficiency *ex ante*.

information that can be easily provided and understood. In particular, we assume that:

1. Youth can understand mean earnings conditional on demographic characteristics, that is, they should understand what is meant by a statement such as "the average monthly earnings for a white person with a university degree";

2. They can understand conditional probabilistic outcomes, such as the meaning of the probability of finding a job conditional on having finished grade 12, versus the probability of finding a job if they were to leave school after grade 11;

3. They use the information provided in a rational and predictable way that mirrors the way information is used in the ROPE model;

4. The different event probabilities are known to youth. (As there is no quantitative data on subjective expectations regarding event probabilities, it would not be possible to test otherwise, using the data we have.)

With the possible exception of assumption 4, these assumptions seem relatively weak. They are supported by a growing literature, discussed in section 2.2 and reviewed by Delavande, Giné, and McKenzie (2011), that finds that people in developing countries can generally understand and answer probabilistic questions. In addition, the seemingly sensible responses to expectations questions in the data we use (for example, the fact that expectations are higher for higher levels of education) seem to support the assumption that survey respondents can understand the concepts of expected average earnings and employment probability. While the ROPE model may seem complicated, it is meant to represent an approximation of

how youth would compare information about different education pathways when deciding whether or not to invest in grade 12. Given the the theoretical and empirical work by Keane and Wolpin (1997), which demonstrates that youths' behaviour is consistent with a model that includes many different education and career options, and where youth implicitly calculate what is analogous to the pathway values of different options, assumption 3 seems reasonable.

We also need to consider how the information provided is used. We model youth's updated expectation after receiving the new information ($E2$) as a linear function of their original expectation ($E1$) and the wage reported in the information they receive (I)

$$E2 = \alpha E1 + (1 - \alpha) I \tag{2.16}$$

where $0 \leq \alpha \leq 1$. In the extreme case of $\alpha = 0$, then $E2 = I$. In the other extreme, the information provided is ignored and there is no updating of expectations. In between theses two extremes lies a continuum of different combinations of $E1$ and $E2$ with the midpoint point being when $\alpha = 0.5$. However, we must remain agnostic about what the most realistic value of α is as we have no data on how sensitive the individuals in our sample are to new information. In the analysis that follows, we will consider the impact of interventions when $\alpha = 0$ and when $\alpha = 0.5$, concentrating our discussion on the results for $\alpha = 0$ as these results represent the *maximum potential* impact of information provision. For all interventions, theses two values of α are a good indication of the range of possible impacts for different values of α. While values for $\alpha = 0.25$ and for $\alpha = 0.75$ were also tested for each intervention the results are not qualitatively different from when $\alpha = 0$ or $\alpha = 0.5$ so are not presented.

Depending on prior expectations, the information given and the value of α, there is a range of possibilities for how the accuracy of expectations (as defined above) may change following an intervention. We are particularly interested in understanding how important the underestimation of returns to education is after each of the interventions defined in table 2.4, given that such divergence between subjectively expected and objectively estimated returns is proposed as the justification for these types of interventions (given that they may be associated with underinvestment in education). In the analysis that follows we show that, in three of the four interventions, it is possible that underestimation of the returns to education may *increase* following the intervention and that this divergence may be especially important for African youth, the group that we consider to be most disadvantaged in education and the labour market.

Although the possibility of underestimating the returns to education receives much attention, the possibility of overestimation of the net returns to grade 12 may also be of concern as it may lead to overinvestment in education. This could be a problem if it is associated with debt incurred to invest in education (which cannot be repaid later on, because the individual finds a job with lower than expected wages). In the analysis that follows, we show that in three of the four interventions we consider, overestimation of the return to grade 12 may result for at least one race group.

Lastly, it must be emphasised that for some youth a change in expectations may be inconsequential even if an intervention leads them to have less accurate expectations. For example, it may be optimal for a particular youth to invest in grade 12 because of very high returns, which they accurately predict. If, after an intervention, such expectations are raised to an unrealistic level, it will still be

optimal to invest and therefore no harm will be caused by the intervention, even if it reduces the accuracy of expectations.

As we do not have information on the direct costs of education, we do not estimate the proportion of youth for which any of the interventions analysed here could actually lead to a change in behaviour. Our focus on whether expectations would be more or less accurate after an intervention (measured through the importance of overestimation or underestimation) is justified by our interest in determining whether an intervention is *potentially harmful.* This is no trivial issue as if we have reason to believe that an intervention might be harmful to most or even some youth, then that would be an argument in favour of changing the design of the intervention. Our results demonstrate that issues of sequential decision making, heterogeneity in the returns to education, and the way youth update their expectations in response to an intervention should be considered *before* an intervention. We now turn to the specific results of each intervention.

2.5.1 The simple intervention

In table 2.5 we present the observed and expected net value of grade 12, estimated using the ROPE model. In addition, this table contains the expected net value of grade 12 following an intervention where all youth are provided with just two values, the average monthly earnings conditional on completion of grade 11 (R1756) and the average monthly earnings conditional on completing grade 12 (R2793). As mentioned above, when $\alpha = 0$, following this intervention, expected earnings will match the values provided by the intervention (or, in other words, that youth replace their subjective expectations of returns to education with the values provided by the intervention). The last three columns in this table compare

the overestimation percentage of returns to grade 12 before and after the intervention, for $\alpha = 0.5$ and for $\alpha = 0$, respectively. We will firstly consider the results in the last column, for $\alpha = 0$. The immediate conclusion is that this intervention results in an average increase in overestimation of total net value of grade 12, from 11% to 16% if we consider the intervention using ROPE specified for the whole sample, and not conditional on race. This change results directly from considering the effect of an information intervention when decisions are made sequentially and option values are important: because the total value of grade 12 depends on higher education options, overestimation or underestimation of the total value of grade 12 may not be reduced if only information on wages conditional on completion of grades 11 and 12 is provided. In this case because youth initially overestimate the value of grade 11, following the intervention they have a lower expectation of the value of grade 11 and consequently *a higher expectation* of the net value of grade 12. It is important to emphasise that the net value of grade 12 depends on the *relative values of different pathways* not on their absolute values.

Table 2.5: Simple intervention and the total net value of grade 12 (1000R's)

	Observed	Expected			Overestimation		
		Original	$\alpha = 0.5$	$\alpha = 0$	Original	$\alpha = 0.5$	$\alpha = 0$
All	740.3	818.3	839.9	861.6	**10.5%**	**13.5%**	**16.4%**
African	435.6	297.7	311.1	321.0	**-31.7%**	**-28.6%**	**-26.3%**
Coloured	444.7	455.6	530.7	569.1	**2.5%**	**19.3%**	**28.0%**
White	2809.2	2811.2	2813.1	2860.6	**0.1%**	**0.1%**	**1.8%**

However, the conclusion of increased overall overestimation of the total net value of grade 12 masks important differences across races. These differences can

only be seen when ROPE is specified by race for all event probabilities and wage rates. In particular, if we consider the impact of the simple intervention for each race, we see that significant underestimation of the returns to grade 12 among African youth has not been significantly reduced (from -32% to -26%). This is because African youth initially underestimate the pathway value of grade 12 and this is not effected by the simple intervention. In contrast, Coloureds now overestimate the returns to education by a large percentage (28%, against an initial overestimation of only 2%). This is because they initially overestimate the value of grade 11, which partially offsets their overestimation of other levels of education but following the intervention the gap between grade 11 and higher education becomes unrealistically large, leading to an overall increased overestimation.[22]

As previously noted, these results represent the *maximum potential* impact of this intervention. However, they are qualitatively similar if we consider the impact when α=0.5. In this case overall overestimation would increase to 13%, and the disaggregation of these values along race lines would remains similar, with Africans, Coloureds and Whites overestimating the total net value of grade 12 by -29%, 19% and 0%, respectively. So in conclusion, it seems that regardless of the value of α, the simple intervention would not meet our three criteria for intervention success in this context.

2.5.2 The standard intervention

A natural improvement on the simple intervention would be to provide information on the returns to higher education. The information intervention would then

[22]As previously noted, because ROPE is not a simple linear model, the results using the sample in aggregate rather than splitting the sample by race are not a weighted average of the results by race.

involve providing information on additional earnings conditional on the conclusion of 13-16 years of education (R3160 for 13 years, R3092 for 14 years, R4408 for 15 years, and R11,611 for 16 years). Its potential impact is presented in table 2.6.

Table 2.6: Standard intervention and the total net value of grade 12 (1000R's)

	Observed	Expected			Overestimation		
		Original	$\alpha = 0.5$	$\alpha = 0$	Original	$\alpha = 0.5$	$\alpha = 0$
All	740.3	818.3	755.3	740.3	**10.5%**	**2.0%**	**0.0%**
African	435.6	297.7	286.8	283.1	**-31.7%**	**-34.1%**	**-35.0%**
Coloured	444.7	455.6	520.9	675.7	**2.5%**	**17.1%**	**51.9%**
White	2809.2	2811.2	2553.9	2296.6	**0.1%**	**-9.1%**	**-18.2%**

If we focus on the overall importance of overestimation of returns to education, it would seem that the intervention is useful, given that such overestimation disappears or reduces, depending on α. However, the results look very different once we consider its impact disaggregated by race. When we do, it is clear that the standard intervention does not meet our criteria for intervention success.

For African youth, the standard intervention would actually *increase average underestimation* of the net returns to grade 12. This is because they would now underestimate the value of having 13 years of education as for this level of education they have higher than average observed returns than Coloured and White youth. Conversely, Coloured youth would greatly overestimate the returns to grade 12 following the intervention (by up to 52%). This is because observed returns to higher education (particularly two and four year degrees) are lower for Coloured youth than for White youth and African youth. Lastly, White youth, who had accurate expectations before the intervention, would underestimate returns by up to

18% following the intervention. This is because White youth generally have higher observed returns to higher education than African and Coloured youth. Overall, it is clear that for all race groups, the standard intervention is *inferior* to the simple intervention. Again this intervention demonstrates that *more information* is not always *better* or *more helpful* to individuals.

Given these results, it seems important to include the possibility of heterogeneous returns to education in the design of these interventions. In what follows, we will consider this change to the two interventions analysed so far. Essentially, we want to address the possibility that the potential for harm identified, particularly in the case of the standard intervention, is due to the provision of *irrelevant* or *misleading* information. However, as we show in the analysis of the results of the targeted and full interventions, the provision of information that accounts for heterogeneity according to race, may not guarantee an improvement in the accuracy of expectations.

2.5.3 The targeted intervention

Surprisingly, conducting the simple intervention with returns to grades 11 and 12 specified by race (the "targeted intervention") would lead to very similar results to the simple intervention. In table 2.7 it can be seen that African youth are no better off with the targeted intervention than with the simple intervention, while in the case of Coloured youth the increase in overestimation is significantly smaller (17% versus 28% when $\alpha = 0$, with the intuition for this being the same as for the simple intervention). For White youth however, the intervention results in a small underestimation of up to 2% which may be worse than the 2% overestimation that

would result from the simple intervention when $\alpha = 0$. [23]

Table 2.7: Targeted intervention and the total net value of grade 12 (1000R's)

	Observed	Expected			Overestimation		
		Original	$\alpha = 0.5$	$\alpha = 0$	Original	$\alpha = 0.5$	$\alpha = 0$
All	740.3	818.3	839.9	861.6	**10.5%**	**13.5%**	**16.4%**
African	435.6	297.7	307.1	323.6	**-31.7%**	**-29.5%**	**-25.7%**
Coloured	444.7	455.6	488.7	521.7	**2.5%**	**9.9%**	**17.3%**
White	2809.2	2811.2	2779.7	2748.1	**0.1%**	**-1.1%**	**-2.2%**

Overall, this intervention is *unambiguously better* than the standard intervention in this context, given that the latter reduces the accuracy of expectations of each race group to a greater extent. However, it is unclear whether or not this intervention is better than the simple intervention. We can only conclude that providing information that takes into account heterogeneity but neglects the sequential nature of investment in education may reduce the accuracy of expectations for some youth.

2.5.4 The full intervention

The previous analysis suggests that the cause of a reduction in expectations accuracy for at least some youth in each intervention is due to the type of information provided. If true, a natural improvement on these interventions would be to provide values for the returns to all levels of education conditional on race. We would consider this to be the provision of full information and following this intervention

[23] As for the simple and standard interventions, these conclusions are independent of the value of α. While not reported, the results are similar for values of $\alpha = 0.25$ and for $\alpha = 0.75$

youth would have *all the information needed* to estimate the average return to grade 12. Under the assumption that $\alpha = 0$, this intervention would reduce overestimation to 0% for all race groups as presented in table 2.8. However, it is not the case that the full intervention will always result in an increase in expectations accuracy: the efficacy of the full intervention rests on our assumptions regarding the nature of heterogeneity (that only race matters), and the way that youth use information (that youth completely replace their prior expectation with the new information).

Even if the heterogeneity assumption is correct, if youth do not completely replace their prior expectations with the information provided, the full intervention may lead to a reduction in expectations accuracy for some. While the full intervention would increase expectations accuracy for African and White youth regardless of the value of α, surprisingly when $\alpha = 0.5$, Coloureds are induced to underestimate the value of grade 12 by 6.6%. This is a net change in expectations accuracy of -11%, which is greater than the net change in expectations accuracy of 9% for Africans and 0% for Whites at this value of α. The relationship between α and overestimation of total net value of grade 12 by Coloureds is graphed in figure 2.3, with the intervention being least effective when $\alpha = 0.71$.

The results would not be possible in a simple linear model of the value of grade 12. In a real options model (such as ROPE), expectations may change in such a way that even though a particular option increases in value, another option is preferred and so the increase in value of the first option is inconsequential. At the same time, the preferred option may reduce in value, reducing the overall value of education. For Coloureds, while 14 years of education is expected to have

Table 2.8: Full intervention and the total net value of grade 12 (1000R's)

	Observed	Expected			Overestimation	
		Original	$\alpha = 0.5$	$\alpha = 0$	Original	$\alpha = 0.5$	$\alpha = 0$
All	740.3	818.3	755.3	740.3	**10.5%**	**2.0%**	**0.0%**
African	435.6	297.7	363.1	435.6	**-31.7%**	**-16.6%**	**0.0%**
Coloured	444.7	455.6	415.5	444.7	**2.5%**	**-6.6%**	**0.0%**
White	2809.2	2811.2	2810.2	2809.2	**0.1%**	**0.0%**	**0.0%**

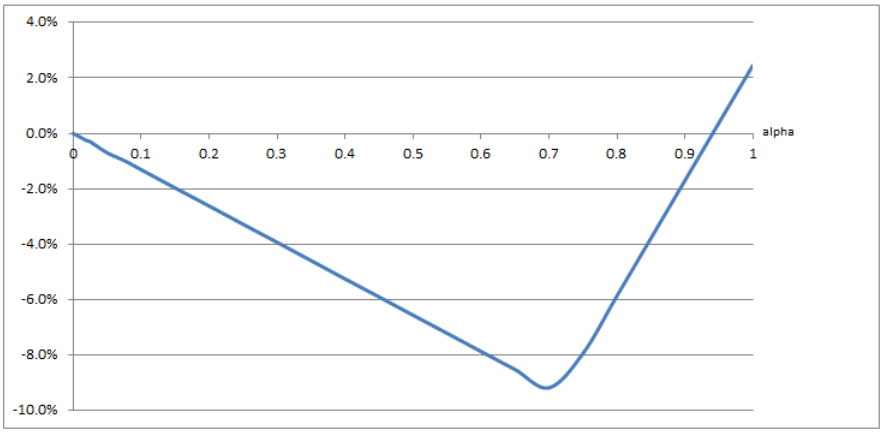

Figure 2.3: Relationship between α and overestimation % for Coloureds

a higher return than 13 years of education, the observed return to 13 years of education is actually higher than the observed return for 14 years of education. This means that when combining their expectations of wages with the new (and correct) information, at some level of α, ROPE will switch from specifying grade 14 as being preferred to grade 13 being preferred. [24] However, unless that switching point is reached, changes in expectations regarding these levels of education do not lead to an overall increase in the net value of grade 12.

This provides definitive evidence that even if information is completely accurate and relevant (given our assumptions, the full intervention is the provision of perfect information), there is no guarantee that it will lead to an increase in expectations accuracy. The impact will depend on how individuals respond to the information, particularly as this relates to their prior expectation. It must be emphasised that assuming that decision makers are rational (as we have done) is not enough to guarantee that correct information will be beneficial when using a real options model (though in a simple linear model it would be). More research on how agents respond to information is needed as the literature in this area is sparse. [25]. In addition, it must be emphasised that assuming the information in this intervention constitutes full information is likely to be an unrealistic assumption, which may invalidate the results when $\alpha = 0$. For example, if gender is an importance source of heterogeneity in the returns to education, our intervention would be misleading as it implies that only race matters, and not gender.

[24]This value of α is about 0.95 and for all values of between 0 and 0.95, the impact of the intervention is to lead to an underestimation of the net value of grade 12 among Coloureds, violating our criteria for intervention success.

[25]Wiswall and Zafar (2011) investigate with experiments, how college students form and update their beliefs about future earnings. They find that while updating behaviour seems logical, students display larger revision when the information provided is more specific and is good news.

In summary, the results for the four interventions provide the major contribution of this chapter: unless heterogeneity, the sequential nature of investment and the way youth respond to information are all considered together, there is no way of knowing whether or not an intervention may improve expectations accuracy. The conclusions on the potential harm caused by information are robust to sensitivity analysis of the key assumptions regarding our data and the ROPE model. The results of this sensitivity analysis are presented in section 2.6. Firstly in section 2.6.1, we consider whether using gender as the variable associated with heterogeneity in the returns to education, leads to different results for the impact of the four interventions. Secondly, in section 2.6.2, we consider whether including the entire sample (instead of restricting the analysis only to full time employed individuals) changes our conclusions regarding the effects of these interventions. Thirdly, in section 2.6.1, we consider the impact of changing our assumption regarding the average anticipated length of an individual's career. In section 2.6.1, we consider potential changes if we limit the set of higher education possibilities to 1 and 2 year diplomas and certificates, with no university level education being possible (possibly because of their higher cost). Lastly, in section 2.6.1, we discuss one possible limitation of our criteria for intervention success. For readers that wish to skip the section on robustness, it is sufficient to note that all of the sensitivity checks produce results that are qualitatively similar to the results reported in this section.

2.5.5 Policy implications: what to consider when informing youth about the returns to education?

While this chapter cannot provide a definitive answer to the question of how youth should be informed about the returns to education, it does make a clear case for the ways in which youth *should not be* informed about the returns to education. In doing so, this chapter hopes to influence the planning of interventions by economists or others wishing to help youth make better education investment choices. Interventions 1-3 demonstrate that unless both heterogeneity in the returns to education and the sequential nature of investment in education are considered, information interventions may reduce the accuracy of expectations. In addition, our consideration of different possibilities for how youth may respond to information provided (that is, different values of α) demonstrates that a full intervention may also reduce the accuracy of expectations.

Also, even if youth were to be provided with more targeted and structured information that accounts for the issues of sequential choice and heterogeneity, it is not clear what the most practical way of doing this would be. Youth could be explicitly presented with the information in the ROPE framework. This information could be provided in the form of decision diagrams (as presented in the previous sections) with outcomes conditional on individual characteristics, and inclusive of all event probabilities. Without guidance to explain such diagrams to youth, they could not be expected to understand such information. Thus even if such an intervention were effective, it may not be low cost, like the Jensen intervention, or even practically possible in some contexts. In addition, the possibility of this type of intervention would depend on three things: 1) the potential of researchers to accurately estimate the returns to education for different types of individuals 2) the

potential of researchers to accurately measure the expectations of these youth and 3) An understanding of how youth will use the information provided. Our analysis of different types of interventions demonstrates that any one of these alone could be a critical problem. Our model and results suggest that research on each of these issues should be conducted for every particular context where an intervention is being considered prior to the design and implementation of the intervention. Not doing so may risk causing harm to some youth.

Lastly, we have only considered giving youth data on conditional mean earnings. Without giving individuals information about the variance in earnings, the interventions we considered would not be giving youth much information about the risk associated with investment in education (though the probability of unemployment conditional on different education levels would be some information about the riskiness of investment). If our assumptions that youth are risk neutral is incorrect, then this may be an important issue. Even so, it is not clear how the variance in earnings could easily be communicated. While Kerr et al. (2013) is one study that provides more complex information in the context of Finland, this is an area that should be researched further, especially for developing countries.

2.6 Robustness Checks

2.6.1 Heterogeneity in the returns to education by gender

For simplicity, we assumed that only heterogeneity in the returns to education according to race was important, as only one dimension of heterogeneity is needed to illustrate it's potential importance to the impact of information interventions.

However, ignoring other potential sources of heterogeneity may have an impact on the results of the four interventions. There is evidence in South Africa that women earn less than men, even after controlling for human capital and type of employment (Heintz and Posel, 2008, p.37). To test if the possibility of heterogeneity in the returns to education by gender may matter, we specify the ROPE model for gender instead of race. The results for the four interventions are presented in table 2.9, which also include for comparison, the original results for the ROPE model specified unconditionally (All).

The first thing to note is that males have a higher observed return to grade 12 and also, a higher expectation of the return to grade 12 than females, which is not surprising. Overall, expectations seem reasonably accurate for both genders with males overestimating the net value of grade 12 by 2% and females overestimating it by only 1%. While it may seem strange that overestimation in ROPE specified for the whole sample (All), is much higher than overestimation when ROPE is specified for males and females, it must be remembered that ROPE is not a linear model and that the results for (All) *are not a weighted average* of the results for males and females.

For an intervention to be deemed worthwhile according to the criteria in section 5, it would have to reduce the overestimation for both genders without leading to underestimation for either gender. None of the four interventions achieve this except the full intervention when $\alpha = 0$.[26] The simple intervention would result in an increase in average overestimation for both genders when $\alpha = 0$. The standard intervention could result in greater overestimation for females (up to 20%) while

[26]However, this finding as only reliable if there are no other sources of heterogeneity to the returns to education apart from gender (which as evidenced in the main results, is not true as there is substantial heterogeneity according to race).

Table 2.9: Gender, information interventions and the total net value of grade 12 (1000R's)

	Observed	Expected			Overestimation		
		Original	$\alpha = 0.5$	$\alpha = 0$	Original	$\alpha = 0.5$	$\alpha = 0$
Simple intervention							
All	740.3	818.3	839.9	861.6	**10.5%**	**13.5%**	**16.4%**
Male	980.7	1002.7	1120.7	1155.8	**2.2%**	**14.3%**	**17.9%**
Female	703.5	710.7	705.4	719.5	**1%**	**0.3%**	**2.3%**
Standard intervention							
All	740.3	818.3	755.3	740.3	**10.5%**	**2.0%**	**0.0%**
Male	980.7	1002.7	814.9	675.7	**2.2%**	**-16.9%**	**-36.1%**
Female	703.5	710.7	749.9	845.3	**1%**	**6.6%**	**20.1%**
Targeted intervention							
All	740.3	818.3	839.9	861.6	**10.5%**	**13.5%**	**16.4%**
Male	980.7	1002.7	1059.3	1116	**2.2%**	**8.0%**	**13.8%**
Female	703.5	710.7	725.1	739.4	**1%**	**3.1%**	**5.1%**
Full intervention							
All	740.3	818.3	755.3	740.3	**10.5%**	**2.0%**	**0.0%**
Male	980.7	1002.7	991.7	980.7	**2.2%**	**1.1%**	**0.0%**
Female	703.5	710.7	679.0	703.5	**1%**	**-3.5%**	**0.0%**

resulting in large underestimation for males (as much as -36.1%). The targeted intervention is similar to the simple intervention and results in greater overestimation for both genders. The full intervention interestingly, is not successful for all values of $\alpha \leq 0.92$ (though not reported) and leads to underestimation for females by 3.5% when $\alpha = 0.5$.

Ideally, if gender in addition to other characteristics (such as race) drive heterogeneity in the returns to education, then ROPE should be specified for each possible combination of relevant characteristics (e.g. black female). Due to the limited number of observations in our data, we do not explore heterogeneity in greater detail. However, the results for gender demonstrate that the ineffectiveness of the interventions is not purely due to our selection of race as the key dimension of heterogeneity. Even if there were only heterogeneity due to a binary variable such as gender, the efficacy of interventions may be substantially affected.

2.6.2 Self-selection into employment

The data on wage expectations used include both those who were already in the labour market and those who weren't. It is possible that there are systematic differences between the expectations of the unemployed and the employed. For example, those that are not working may overestimate the earnings of others with the same level of education. Also, some of those that are not working may not want to work and may not have invested in education for economic returns. They may consequently have lower expectations of the return to education than would seem to justify rational investment. As we do not estimate the determinants of expectations we do not know exactly why expectations may differ between those who work and those who are jobless. For our purposes however, all that matters is

whether or not self-selection in this regard impacts on our results. We can analyse this by restricting the sample used to only those that are employed. The results for the four interventions using only data for the employed are presented in table 2.10.

The results for the four interventions are qualitatively and quantitatively similar to our original results. The simple intervention would result in a reduction in average underestimation for Africans but would result in overestimation for Coloureds and Whites. The standard intervention would also result in a reduction in average underestimation of Africans and increased overestimation for Coloureds. It would cause greater overestimation for Whites when $\alpha = 0$ but a slight reduction in overestimation for Whites if $\alpha = 0.5$. The targeted intervention is unambiguously better than the simple and standard interventions as it results in a reduction in underestimation for Africans, a reduction in overestimation for Whites and a smaller increase in overestimation for Coloureds than the first two interventions. However, as the standard intervention still results in an increase in overestimation for one group, it does not meet our criteria for a successful intervention. Lastly, the impact of the full intervention could result in underestimation for Coloureds of about 19%, in a similar way to our original results. Overall, it is clear that none of the interventions are without the potential to cause harm.

2.6.3 Assumptions in ROPE regarding career length

Whether or not an individual appears to overestimates the net value of grade 12 in a sequential decision making framework such as ROPE, may depend on the structure and assumptions of the model. One important assumption we make is that individuals consider the returns to education over their whole career, which

Table 2.10: Self selection of the employed, information interventions and the total net value of grade 12 (1000R's)

	Observed	Expected			Overestimation	
		Original	$\alpha = 0.5$	$\alpha = 0$	Original	$\alpha = 0.5$	$\alpha = 0$
Simple intervention							
All	740.3	862.3	907.2	952.1	**16.5%**	**22.5%**	**28.6%**
African	435.6	216.7	257.3	283.1	**-50.2%**	**-40.9%**	**-35%**
Coloured	444.7	453.7	517.4	587.6	**2.0%**	**16.4%**	**32.1%**
White	2809.2	3835.3	3826.0	3915.5	**36.5%**	**36.2%**	**39.4%**
Standard intervention							
All	740.3	862.3	777.3	740.3	**16.5%**	**5.0%**	**0.0%**
African	435.6	216.7	247.5	283.1	**-50.2%**	**-43.2%**	**-35%**
Coloured	444.7	453.7	515.3	675.7	**2.0%**	**15.9%**	**51.9%**
White	2809.2	3835.3	3065.9	2296.6	**36.5%**	**9.1%**	**-18.2%**
Targeted intervention							
All	740.3	862.3	907.2	952.1	**16.5%**	**22.5%**	**28.6%**
African	435.6	216.7	250.0	288.3	**-50.2%**	**-42.6%**	**-33.8%**
Coloured	444.7	453.7	496.8	540	**2.0%**	**11.7%**	**21.4%**
White	2809.2	3835.3	3819.1	3803	**36.5%**	**36.0%**	**35.4%**
Full intervention							
All	740.3	862.3	777.3	740.3	**16.5%**	**5.0%**	**0.0%**
African	435.6	216.7	323.7	435.6	**-50.2%**	**-25.7%**	**0.0%**
Coloured	444.7	453.7	361.0	444.7	**2.0%**	**-18.8%**	**0.0%**
White	2809.2	3835.3	3322.2	2809.2	**36.5%**	**18.3%**	**0.0%**

lasts until the age of life expectancy at birth.[27] We also assumed that the discount rate for future returns exactly equalled the growth rate in wages. If wages growth occurs mostly early on in the career (e.g. before the age of 35) and individuals are short sighted, we may have overweighed the long run returns to education in ROPE. In case our assumption about long run returns is too strong, we specify the ROPE model in table 2.11 until the age of 35, effectively setting the discount rate beyond the age of 35 to $1 + g$. [28]

Firstly, we note that the change in our assumption does lead to different measures of the initial overestimation and underestimation of the total net value of grade 12. Underestimation for Africans increases from about 32% to about 37%, while Coloureds, who previously overestimated by about 3%, now *underestimate* the net value of grade 12 by about 5%. Conversely, overestimation for Whites increases from about 0% to about 2%. These results demonstrate that changing the assumption about career length/ discounting, does not have straightforward impacts on our measure of the accuracy of expectations. For some groups it may yield higher measures of underestimation, while for other groups it may lead to higher levels of overestimation. In any case, results for the four interventions are qualitatively similar to the original results and none of our conclusions regarding

[27]Another assumption regarding our model that may be important is in regards to grade repetition and re-enrollment, which as found by Pugatch (2012) using CAPS data, are very important in the context of South Africa. While extending the model to account for re-enrollment possibilities would be possible, we do not do so as this would significantly complicate the model without adding to the argument here that assumptions about ROPE may impact on our measures of expectations accuracy.

[28]Also, we use the age of 35 as information interventions such Oreopoulos and Dunn (2013) and Jensen (2010) involve the provision of information on the average earnings of those aged 35 who are working. This implies that youth targeted with information in these studies are concerned with the future returns to education at least up until the age of 35.

Table 2.11: Short career length, information interventions and the total net value of grade 12 (1000R's)

	Observed	Expected			Overestimation		
		Original	$\alpha = 0.5$	$\alpha = 0$	Original	$\alpha = 0.5$	$\alpha = 0$
Simple intervention							
All	283.5	307.3	317.9	328.5	**8.4%**	**12.2%**	**15.9%**
African	175.9	111.2	117.6	124.5	**-36.8%**	**-33.2%**	**-29.2%**
Coloured	171.7	162.5	197.1	215.2	**-5.4%**	**14.8%**	**25.3%**
White	1031.8	1046.9	1050.6	1073	**1.5%**	**1.8%**	**4%**
Standard intervention							
All	283.5	307.3	282.2	283.5	**8.4%**	**-0.5%**	**0.0%**
African	175.9	111.2	109.9	108.5	**-36.8%**	**-37.6%**	**-38.3%**
Coloured	171.7	162.5	196.1	260.7	**-5.4%**	**14.2%**	**51.8%**
White	1031.8	1046.9	947.9	849	**1.5%**	**-8.1%**	**-17.7%**
Targeted intervention							
All	283.5	307.3	317.9	328.5	**8.4%**	%	**15.9%**
African	175.9	111.2	119.6	128	**-36.8%**	%	**-27.3%**
Coloured	171.7	162.5	178.0	193.5	**-5.4%**	%	**12.7%**
White	1031.8	1046.9	1032.0	1017.1	**1.5%**	%	**-1.4%**
Full intervention							
All	283.5	307.3	282.2	283.5	**8.4%**	**-0.5%**	**0.0%**
African	175.9	111.2	143.6	175.9	**-36.8%**	**-18.4%**	**0.0%**
Coloured	171.7	162.5	157.3	171.7	**-5.4%**	**-8.4%**	**0.0%**
White	1031.8	1046.9	1039.3	1031.8	**1.5%**	**0.7%**	**0.0%**

their efficacy change, as each intervention fails to meet our criteria for success. For example, the standard intervention may lead to Whites underestimating the value of grade 12 by about 18% while the full intervention may cause underestimation of Coloureds to increase to about 8% when $\alpha = 0.5$.

We also used the ROPE model under the alternative assumption that the age of retirement (age 65) is used instead of life expectancy. Our original assumption that life expectancy is the upper bound for career length effectively implied a discount rate beyond the age of 56 to be $1 + g$. Setting the age to 65 effectively means setting a discount rate equal to g until the age of 65 and setting the discount rate to $1 + g$ thereafter. This would be appropriate if youth do not anticipate dying before retirement age but rather, anticipate working until retirement age. We report the results in table 2.12. Clearly, these results are very similar to the original results and none of our conclusions regarding intervention efficacy change. Though not reported, all integer life expectancies from 30-65 were tested with no qualitative difference to the results. Also not reported, in the case where expected career length were to differ sharply by race, for example 45 for Africans, 55 for Coloureds and 65 for Whites, all results remain qualitatively the same.

Lastly, it should be noted that a traditional discount rate was not used in the analysis as we assume that youth think in nominal terms (following Jensen (2010)).This assumption is consistent with short-sightedness being modelled as short career length. However, even if youth do not think precisely in terms of a constant discount rate, it may be a better approximation of how they view the future than simply assuming a short career. For this reason, I tested integer discount rates from 1%-15% with $g = 0$ as a further robustness check of results for the ages of 35, 56, and 65 though the results are not reported in the interest of

Table 2.12: Career till retirement age, information interventions and the total net value of grade 12 (1000R's)

	Observed	Expected			Overestimation	
		Original	$\alpha = 0.5$	$\alpha = 0$	Original	$\alpha = 0.5$	$\alpha = 0$
Simple intervention							
All	926.9	1027.0	1053.2	1079.3	**10.8%**	**13.6%**	**16.4%**
African	541.6	374.6	390.9	402.7	**-30.8%**	**-27.8%**	**-25.6%**
Coloured	556.2	575.4	666.9	713.6	**3.4%**	**19.9%**	**28.3%**
White	3535.2	3531.9	3533.0	3590.8	**-0.1%**	**-0.1%**	**1.6%**
Standard intervention							
All	926.9	1027.0	948.5	926.9	**10.8%**	**2.3%**	**0.0%**
African	541.6	374.6	359.1	354.4	**-30.8%**	**-33.7%**	**-34.6%**
Coloured	556.2	575.4	653.6	845.2	**3.4%**	**17.5%**	**52.0%**
White	3535.2	3531.9	3209.9	2887.9	**-0.1%**	**-9.2%**	**-18.3%**
Targeted intervention							
All	926.9	1027.0	1053.2	1079.3	**10.8%**	**13.6%**	**16.4%**
African	541.6	374.6	384.7	403.5	**-30.8%**	**-29.0%**	**-25.5%**
Coloured	556.2	575.4	615.6	655.8	**3.4%**	**10.7%**	**17.9%**
White	3535.2	3531.9	3493.5	3455.2	**-0.1%**	**-1.2%**	**-2.3%**
Full intervention							
All	926.9	1027.0	948.5	926.9	**10.8%**	**2.3%**	**0.0%**
African	541.6	374.6	452.7	541.6	**-30.8%**	**-16.4%**	**0.0%**
Coloured	556.2	575.4	521.0	556.2	**3.4%**	**-6.3%**	**0.0%**
White	3535.2	3531.9	3533.5	3535.2	**-0.1%**	**0.0%**	**0.0%**

space. I tested a total of 45 combinations of discount rate and age, with $\alpha = 0$ for all interventions and also $\alpha = 0.5$ for the full intervention. For all interventions and for all discount rates tested, the interventions failed to meet the criteria for intervention success (except for the full intervention when $\alpha = 0$), though some discount rates did substantially change the impact of the interventions, as is the case with the robustness checks for different ages reported above.

2.6.4 Credit constraints

If youth anticipate not being able to afford some levels of higher education because they face credit constraints, then these options may not be considered in the decision to invest in grade 12. In South Africa, like most countries, higher education can be an expensive investment and some youth many not enrol in university when it is optimal to do so because of credit constraints. Attanasio and Kaufmann (2009) is one study that focuses on the importance of credit constraints and the possibility that they 'break' the link between expectations and investment in education. One limiting simplification of ROPE is that all direct costs to education are assumed to be zero, implicitly ignoring the importance of credit constraints. However, if credit constraints do limit the number of possible education pathways, it may be that the dangers of information provision are not as significant. To test this hypothesis, we specify a variation of ROPE with only two higher education options: 1 year of higher education or 2 years of higher education. Effectively, we are assuming that all youth are credit constrained to the same degree. In table 2.13, we report results for the four interventions with this variation of ROPE.

The results are quantitatively very different but qualitatively similar to the original results. Firstly, it now seems as though Whites greatly overestimate the value

Table 2.13: Credit constraints, information interventions and the total net value of grade 12 (1000R's)

	Observed	Expected			Overestimation		
		Original	$\alpha = 0.5$	$\alpha = 0$	Original	$\alpha = 0.5$	$\alpha = 0$
Simple intervention							
All	384.2	576.9	591.8	615.6	**47.8%**	**54.0%**	**60.2%**
African	399.2	235.0	248.5	259.5	**-41.1%**	**-37.7%**	**-35.0%**
Coloured	413.9	402.1	477.3	516.5	**-2.9%**	**15.3%**	**24.8%**
White	584.8	1388.2	1410.9	1463.1	**137.4%**	**141.3%**	**150.2%**
Standard intervention							
All	384.2	576.9	442.4	384.2	**47.8%**	**15.1%**	**0.0%**
African	399.2	235	209.7	192.9	**-41.1%**	**-47.5%**	**-51.7%**
Coloured	413.9	402.1	399.5	486	**-2.9%**	**-3.5%**	**17.4%**
White	584.8	1388.2	980.7	573.3	**137.4%**	**67.7%**	**-2.0%**
Targeted intervention							
All	384.2	576.9	591.8	615.6	**47.8%**	**54.0%**	**60.2%**
African	399.2	235.0	246.6	266.7	**-41.1%**	**-38.2%**	**-33.2%**
Coloured	413.9	402.1	435.7	469.3	**-2.9%**	**5.3%**	**13.4%**
White	584.8	1388.2	1353.6	1319.0	**137.4%**	**131.5%**	**125.6%**
Full intervention							
All	384.2	576.9	442.4	384.2	**47.8%**	**15.1%**	**0.0%**
African	399.2	235	312.8	399.2	**-41.1%**	**-21.6%**	**0.0%**
Coloured	413.9	402.1	363.5	413.9	**-2.9%**	**-12.2%**	**0.0%**
White	584.8	1388.2	986.5	584.8	**137.4%**	**68.7%**	**0.0%**

of grade 12 (about 138%), while Africans still greatly underestimate the value of grade 12 (48%) and Coloureds slightly underestimate the value of grade 12. like the original results, while underestimation is reduced for African youth by the simple, full and targetted interventions, the standard intervention increases underestimation. Also similar to our original results, the first and third interventions lead to greater overestimation for Coloureds while the full intervention leads to underestimation. Also similar to our original results, overestimation increases for White youth following the simple intervention and becomes an underestimation following the standard intervention. In summary, none of the four interventions meet our criteria for intervention success, leaving our original conclusions unchanged.

2.6.5 Criteria for intervention success

Lastly, we note one weakness of our criteria for intervention success: there is no consideration of the potential impact of interventions on the variance of expectations. While in three out of four interventions, average underestimation among Africans would be reduced this does not imply that *the number of those underestimating* would likewise be reduced or be unchanged. In fact, the number of those underestimating the returns following interventions 1 and interventions 4 could actually increase. This is because before the intervention, as there is variance in expectations, some African youth overestimate the earnings conditional on grades 11 and 12 and some African youth underestimate earnings conditional on grades 11 and 12. There would be no variance if α=0. Following the simple intervention for example, all African youth may underestimate the returns to grade 12 by 26%. How this would affect educational investment is unclear but it cannot be assumed that a reduction in average underestimation would lead to educational investment

that is equal to or greater for African youth than if there had been no intervention, unless inaccuracy is reduced to 0%. In any case, this issue could not affect our conclusion regarding the efficacy of interventions in this context as none of the four interventions meet our (arguably) weak criteria. Making the criteria much more stringent (for example, adding a criterion that following the intervention, no one should underestimate the returns to grade 12) would still imply that all four interventions are not successful.

2.6.6 Implications for each type of intervention

This section has tested the robustness of our results to changes in several key assumptions about our model and the data. For each intervention in each sensitivity check, at least one groups' expectations are made less accurate, violating our criteria for intervention success. When $\alpha = 0$, the simple intervention leads to overestimation for Coloureds and Whites if the sample is restricted to those that are employed, if we change assumptions about the career length, or if we consider credit constraints. Also, it leads to overestimation of males and females if ROPE is specified according to gender (about 18% and 2%, respectively). When $\alpha = 0$, the standard intervention leads to overestimation for females (about 20%) and underestimation for males (about 36%), while it leads to overestimation for Coloureds in all other sensitivity checks. The targeted intervention leads to overestimation for males and females (about 14% and 5%, respectively), while it also leads to overestimation for Coloureds in all other sensitivity checks. Lastly, when $\alpha = 0.5$, the full intervention leads to underestimation for females (about -3%) while in all other sensitivity checks it leads to underestimation for Coloureds just like the results in 2.5.4.

Overall, the results of these sensitivity checks are important because they demonstrate that the dangers of information provision do not occur simply because of a very specific combination of model and data assumptions. If the ROPE model is changed substantially, or if we change our assumption about the source of heterogeneity, it is clear that the dangers of information remain *realistic* possibilities (as opposed to simply being *theoretical* possibilities).

While beyond the scope of this chapter, there are at least three significant extensions to ROPE that could make it considerably more realistic (though more complicated) but that would require either substantial changes to the model or data that is not available in CAPS. These possible extensions are 1) the inclusion of positive education costs, 2) the option of repetition, and 3) the option of reentry into education once an individual has entered the labour market. The likely impact of each of these extensions on the conclusions of this chapter is discussed below:

1. The inclusion of positive education costs for each grade beyond grade 11 should reduce the net value of grade 12 and the pathway value of grade 12. It may also change the level of overestimation for each race. However, the results of the interventions should remain qualitatively similar as they are driven by changes to the expected relative net benefits of different education pathways, and such changes would occur regardless of the costs of education.

2. If youth are able to repeat a grade that they fail in the following year, this option of repetition will increase the gross return to that grade of education. However, whether or not this increases the pathway value of grade 12 will depend on the pass probabilities for each race. Though not reported, we tested one example extension to ROPE, where grade 12 may be repeated one time

with the same probability of failure and where a pass can only occur without exemption on this second attempt. In this situation, compared to the original model, the pathway value percentage for African youth increases from 52.1% to 56.6% but the pathway value for White youth decreases slightly, from 67.2% to 67.0%. Regardless of the pathway value percentage, the overall gross value of the grade being repeated will increase (as repetition is always an *option* that is only chosen when it is relatively better than the alternative of entering the labour market). So it should be noted that both relative values of pathways and the absolute values of different levels of education will be different in a model with repetition. However, this should not change the potential dangers of information interventions. Regardless of the options available for repetition, individuals still need to make choices based on the *expected relative benefits* of different levels of education (particularly in ROPE, the net value of grade 12). In the example extension we tested, the results of the four interventions remain qualitatively the same with no intervention meeting our criteria for success for both values of α.

3. Unlike repetition and positive education costs, the option to re-enter education after a period in the labour market may have implications for the dangers of information provision. In particular, if an individual can change pathways after experiencing a negative outcome (such as unemployment or obtaining a low wage job), then the decision to continue onto grade 12 directly after grade 11 becomes much less critical. By extension, information provision that can impact on this decision also becomes less critical. For example, following the Standard intervention, African youth have lower expectations of the net value of grade 12. Suppose that this caused them to leave school

after grade 11 but that they later realised it was a mistake and so re-enrolled in grade 12 one year later. In this case, they would not have lost a lifetimes' worth of higher income, as would have been the case if re-entry to school was not allowed. Further research would need to be done to explore how significant the possibility of re-entry into education is likely to be.

2.7 Conclusion

This paper developed a real options model of investment in education and applied it to an *ex ante* evaluation of the potential impacts of information interventions aimed at improving the accuracy of expectation.

Using data from the Cape Area Panel Study in South Africa, it showed the potential policy relevance of information interventions for at least a subset of individuals, by identifying substantial differences between the expected and observed total value of grade 12 among African youth (though not for Coloured and White youth). This difference is due mainly from differences in the expected and observed pathway value of grade 12 (that is, the value of the option to pursue higher education) for African youth.

The impacts of information provision regarding the returns to education were then considered using the Real Options Pathways for Education (ROPE) model, considering the issues of education investment options and heterogeneity in the returns to education that received less attention in the existing interventions. Firstly, it was found that information that does not account for possible heterogeneity in the returns to education or higher education options (the simple intervention), would do very little to reduce African youths' underestimation of the net value of grade 12. Providing information on returns to grade 12 (but not on higher edu-

cation) according to race (the targeted intervention) would not seem to be more helpful than the simple intervention, at least for Africans (and Whites) who, as noted, undervalue the options open with the completion of higher education. The standard intervention, which provides information on all wage outcomes (and, as such, addresses this limitation) but does not consider heterogeneity of returns, would potentially be greatly harmful leading to significant underestimation by African and White youth (35% and 18%, respectively). Lastly, results for the full intervention demonstrate that even when both heterogeneity and the sequential nature of investment are considered, the intervention may lead to a substantial increase in underestimation, the extent of which depends on how new information is used by youth to update their prior expectations.

These results should be seen as a cautionary note regarding the potential and shortcomings of information interventions. To our knowledge, this paper is the first to demonstrate this possibility using real data on observed outcomes and expected outcomes. Similar analyses to this one could be conducted prior to any information intervention where option values and/or heterogeneity in returns are important to evaluate, *ex ante*, the potential limitations of different types of information provision. Lastly, it must be emphasised that the main contribution of this paper was not to investigate the accuracy of earnings expectations in South Africa and suggest policies for that particular context, but rather to demonstrate a specific analytical approach to planning and evaluating information interventions *ex ante*. In particular, this approach allows interventions to be checked for their potential to reduce the accuracy of expectations, and in doing so, potentially reduce the efficiency of economic choices.

Chapter 3

Social Networks and Wages in South Africa

3.1 Introduction

An extensive literature in economics and other social sciences, recently reviewed in Marsden and Gorman (2001), Ioannides and Loury (2004) and Topa (2011), analyses the importance of social networks for the functioning of labour markets. [1] One of the stylised facts that emerges from this work is that their use is pervasive and generally productive. As a result, a large fraction of jobs are found through social networks. [2]

Given this empirical regularity, an important economic question that follows is whether the use of social networks affects labour market outcomes, in particular wages and tenure, and if they do, for whom. As Mouw (2003, p. 868) notes, "This

[1] The use of social networks refers to job seekers communicating with friends, relatives and acquaintances about employment opportunities and firms using informal referrals and recommendations to find workers.

[2] This is true in both developed and developing countries. See, for example, Bewley (1999) who finds that, in the USA, 30-60% of the jobs are found through social contacts. See also Granovetter (1974) and Holzer (1987) for the USA and Bentolila and Suarez (2010) for a recent analysis of the role of social networks in several countries of the European Union. In developing countries see Magruder (2010) and Burns, Godlonton, and Keswell (2010) for evidence in South Africa, Calvo-Armegnol and Zenou (2005) for an analysis of the role of social networks in the functioning of the labour market in Egypt, and Lawler et al. (1995) for a comparison of human resource management practices of firms in India and in Thailand and the extent of use of internal referrals to fill different jobs.

is no trivial issue. If using contacts seems to have little overall impact on labour market outcomes, then perhaps economic models of the labour market can safely ignore 'embeddedness' ... without sacrificing explanatory power."

The growing number of studies, both theoretical and empirical, that address this question can be organized around two competing hypotheses. The first is the "good matches" hypothesis, which dates back to a seminal study by Rees (1966). More formally, Montgomery (1991), Saloner (1985) and Simon and Warner (1992), interpret social networks as screening or monitoring mechanisms that increase the quality of the match between worker and firm by reducing informational asymmetries, leading to higher productivity which would be reflected by a higher wage. Alternatively, the "limited choices" hypothesis, initially suggested by Loury (2006), interprets social networks as mechanisms that allocate workers to jobs at a relatively low cost for both firms and prospective workers. Because search through networks is limited by network characteristics, including the job information that its members have access to, this type of search may lead to sub-optimal matches and associated wage penalties.

We briefly review both of these different interpretations in section 3.2. We emphasise that they share one important similarity: that the importance of the referee, both in terms of motivation and identity, is usually unexamined. [3] Rather, a simple assumption is usually made that "present workers tend to refer people like themselves ..." (Rees, 1966, p. 562). Although the importance of homophily, that "birds of a feather flock together", has long been recognized in the social network literature (see McPherson, Smith-Lovin, and Cook, 2001, for a review), these two competing hypotheses emphasize different aspects of similarity between members

[3] A referee is someone that refers an individual to a job and/or refers a job to an individual.

of a network: productive ability in the first case, identity in the second.

In section 3.2 we also discuss some of the recent work that addresses the possible conflict between these two dimensions of the referee's decision. We also raise the possibility of referee opportunism, that is, that referees may indicate someone who is not of high ability but who is socially close. An intuitive consequence of this possibility that we explore in this paper, is that firms may react by applying a wage penalty that is a positive function of how close the referee is to the worker. This prediction is close in spirit, to the important study by Granovetter (1974) which emphasises the positive impact of increased social distance (weak ties) between referee and job seeker, on the probability of finding a job.

Focusing on the referee and their incentives is important because previous research has found it to impact on the type of person connected to the job and the quality of their work (e.g. Magruder and Beaman (2012)). We do not directly observe the incentives of the referee but assume that they are linked strongly to the identity of the referee (which is observed in our data). Particularly, if the difference in social proximity between friends and relatives is systematic and in the direction we assume (that relatives are closer than friends), we can infer that relatives have a greater incentive to risk referring unsuitable workers.

We empirically estimate the effects on wages of finding a job through the use of different social contacts using data from the Cape Area Panel Study (CAPS), which we present, succinctly, in section 3.3. The CAPS data is particularly valuable for this purpose, for two reasons. The first is the wealth of information in CAPS on respondents' employment outcomes, job search methods, and personal characteristics such as education and working experience, for a large sample of youth across several years (2002-2006). The second is the importance of the func-

tioning of the labour market in addressing issues of poverty and inequality in the South African context, a point that we also discuss, briefly, in section 3.3.

In section 3.4 we present our results. We find that, while controlling for individual and job characteristics, including unobserved heterogeneity, finding a job through social networks carries a wage penalty for all race groups. This suggests that the "good matches" hypothesis finds no support in the context of South Africa. This negative effect is compounded by social proximity, especially in the case of African youth, with higher penalties associated with the use of contacts who are socially closer. [4]

We conclude in section 3.5. An understanding of the significance of social networks to individual job search outcomes may have important policy implications in the high unemployment context of South Africa. If youth are using social networks to acquire jobs even though they are associated with large wage penalties, this may be a sign that labour market institutions do not facilitate formal job search for these individuals to find more suitable employment.

3.2 Wages and social networks: a brief review of the literature

Most of the economic studies of the role of social networks in the functioning of labour markets build on the job matching approach and its assumption that not all

[4] A note on terminology. We adopt the labels used in CAPS and by Seekings and Nattrass (2005, p.ix) and use the word 'African' to refer to peoples classified under apartheid as "native", "bantu" or "black", the word 'White' to refer to people classified as European or white, and the word 'Coloured' to refer to the people that don't fit the first two categories, including descendants of the indigenous Khoi and San, the descendants of "Malay" slaves from Indonesia and those with mixed race ancestry.

matches between workers and firms are optimal. Prospective workers are heterogeneous in preferences for different types of jobs and in their productivity in those jobs, while jobs are heterogeneous in the skills required and in their non pecuniary characteristics (Johnson, 1978, Jovanovic, 1979, Viscusi, 1980). Uncertainty and imperfect information about any of these aspects may lead some hired workers to be poor matches for the jobs they hold, either because the job seeker misjudges the job or the firm misjudges the prospective worker. The use of social networks in job search is then seen as an effective method for firms and potential workers to improve the matching process (Montgomery, 1991, Saloner, 1985, Simon and Warner, 1992), at relatively low costs (Holzer, 1987). The label "better matches" derives from this interpretation.

If the use of job contacts makes it more likely that job seekers and firms will obtain accurate information about each other, then this should impact positively on wages and tenure. For example, in the model proposed by Jovanovic (1979) wages are positively related to tenure. If a job match turns out to be poor, it is more likely that job separation will occur earlier rather than later (that is, as soon as the low quality of the match becomes apparent to the worker and the firm) while longer tenure is a sign that the worker is suitable for the job. Similarly, if wages are set equal to worker's marginal product, they should be higher on average for those workers that are proven to be productive and consequently have longer tenure. Simon and Warner (1992) extend this reasoning to suggest that wages may be higher from the start of employment for a worker that is referred to the firm as the firm anticipates higher than average productivity.

Implicit in this analysis is the assumption that workers refer others like themselves and that because of this, "worker referrals . . . usually provide good screening

for firms who are satisfied with their present workforce"(Rees, 1966, p.562). This assumption is also explicitly made in Montgomery (1991), where the motivations of the referee are replaced with the assumption that abilities are correlated across individual members of a network. However, the benefits of a network are conditional on its characteristics and there is abundant evidence that homophily in networks is not limited to one specific characteristic (in this case, ability). Rather, homophily spans several characteristics such as gender, race and age, that are not inherently productive.

The "limited choices" hypothesis emphasises that there may be costs associated with the use of social networks precisely because productivity irrelevant characteristics play a role in network formation and, as such, in access to the information it may transmit. While networks may be used because of convenience and low monetary costs relative to formal methods, they provide the job seeker with a range of job opportunities limited by the characteristics of their social network. These job opportunities may not be suited to the characteristics of the job seeker, unlike the wider range of jobs that are potentially available through formal methods (Bentolila and Suarez, 2010). As a result, wages in jobs found through social networks will be equal to or lower than in jobs found through formal methods. The clearly contrasting wage implications of the limited choices and good matches hypotheses can be used to empirically test between these two hypotheses. [5]

[5]It is important to note, as Loury (2006) does, that whilst longer job tenure should be the result of good matches it may also be consistent with the limited choices hypothesis and so can not be a definitive test between the theories. Simon and Warner (1992) argue that jobs obtained through contacts with more information (i.e., recruiters or acquaintances inside the firm) lasted significantly longer than jobs obtained through want ads or private agencies and that this is evidence of the benefits of using networks. In contrast however, Topel and Ward (1992) argue that the average young worker holds seven jobs in the first 10 years of his working life and that job mobility accounts for one-third of wage growth during this

Several recent studies have provided support to the "limited choices" hypothesis. Combining personnel files and job-histories of workers at a manufacturing firm, Antoninis (2006) found that new recruits receive a higher starting wage when recommended to the job by an individual with direct experience of their productivity but that the use of referrals from friends and relatives has no effect on the starting wage and may even be negatively related to wages in unskilled jobs. Similarly, Bentolila and Suarez (2010) found that contacts are associated with wage discounts of at least 2.5%, although they also reduce unemployment duration by 1-3 months on average, while Pellizzari (2010), using data from the European Community Household Panel, finds that penalties and premiums are equally common across the countries, a result that he attributes to the mediating effect of the labour market institutions of each country. [6]

Of particular interest in this paper is the idea, suggested by Loury (2006) and Calvo-Armegnol and Jackson (2007), that heterogeneity in the characteristics of social networks may result in the possibility of the limited choices hypothesis finding support for some job seekers but not for others, that is, that different population groups may experience different network effects according to the differences in the characteristics of their networks. An extensive literature provides empirical support to this possibility.

Korenman and Turner (1996) reported that, among young workers in inner-city

period. If workers chose to change jobs voluntarily this must be because of having better alternative jobs available and so such job mobility should increase wages. In contrast, young workers that stick with jobs for longer than average periods of time may miss out on opportunities to obtain higher wages. Keith and McWilliams (1995) and Abbott and Beach (1994) show that the effects of job mobility on wages depend on whether the job change was voluntary or involuntary. Both studies find that voluntary separation increases wages while involuntary separation does not.

[6]Finally, several other studies have found no wage effect associated the use of informal search methods (Bridges and Villemez, 1986, Holzer, 1987, Mouw, 2003).

Boston, Whites who found jobs through contacts received much larger wage gains (19% higher) than Blacks with similar observable characteristics. Smith (2000) showed that gender wage differences were small for those using formal job-search methods. In contrast, she found larger wage differences between Hispanics and Whites who used personal contacts to find jobs compared to those who used more formal means.

Using data on contacts amongst a group of Dartmouth college seniors, Marmaros and Sacerdote (2002) found that the students networking with fraternity/sorority alumni were more likely to get better paying and more prestigious jobs. However, there were significant gender and race differences and the type of contacts used had an impact on the quality of employment found. For example, women were less likely to get fraternity/sorority help, more likely to use help from professors and equally likely to get help from relatives. Loury (2006) also finds heterogeneity in network effects according to the gender of the contact, with prior generation male relatives having a positive and larger effect than female friends and relatives.

It is also important to consider that networks are not formed with the sole purpose of disseminating information important for the functioning of the labour market, rather they may provide a variety of functions (including, and importantly in developing countries, insurance). It seems likely that there will exist spill-overs from one context/purpose to another. This possibility was also raised by Rees (1966), who notes that "the few firms who deliberately avoid informal sources are either those who are seeking to upgrade their work force or *those who have had bad experience with nepotism or cliques*" (p. 562, emphasis added). Similarly, Antoninis (2006, p.135) remarks, in criticising Montgomery's (1991) assumption of correlated abilities between referees and job seekers, that "...one observes that

individuals take active steps to ensure employment for members of their family network for reasons that are often unrelated to whether their relative has a similar level of ability. Firms often respond to such recommendations favorably despite their knowledge that they may have little to do with the new recruits' productivity potential".

Some recent work (Magruder and Beaman, 2012, Fafchamps and Moradi, 2009, Bandiera, Barankay, and Rasul, 2010) analyzes the possible implications of this overlap for productivity in a variety of contexts, concluding that referee opportunism may lead to lower productivity. The results of Magruder and Beaman (2012) are particularly interesting in this context. The authors asked workers to refer people for a task. Referees were randomly allocated to two groups: they would either receive a fixed amount per referral or an amount that depended on the performance of the referred person. When referral is made contingent on worker's productivity, participants are less likely to refer a relative and more likely to refer someone with high productivity. Additionally, the effect of the incentive depends on the quality of the referee: participants of low ability do not seem capable of recruiting high ability workers. This work then provides mixed support to the "good matches" hypothesis: although referee and worker's ability may be correlated (as assumed by Montgomery (1991)), they do not need to be in the absence of the right incentive, in which case referee's are motivated by the exchange of favors with friends and relatives.

The consequences of such opportunistic behavior are intuitively simple: firms may withdraw from reliance on social networks as a recruitment mechanism (as mentioned in Rees (1966)) or may take into account such productivity-irrelevant motivations when making a (lower) wage offer. Because the motivation to refer

someone who is not necessarily a good match should increase the closer the referee is to the worker, so should the wage penalty. If relatives of the worker are generally socially closer to the worker than friends, as we assume, then wage penalties associated with matches through relatives should be higher than wage penalties associated with matches made through friends. In section 3.4 we present estimates of the effect of social networks which support this interpretation, particularly among the group who experience higher unemployment rates (African youth).

3.3 Context and Data: Social Networks and Wages in South Africa

In South Africa, unemployment has been persistently high, having increased in the decade following the end of apartheid and remaining high ever since (Kingdon and Knight, 2004, 2007). As in other countries, youth unemployment is higher than the overall rate. [7] Race still plays an important role in describing unemployment. Morrow, Panday, and Richter (2005), using data from a 2002 Labour Force Survey, found a very high rate of youth unemployment for Africans (above 60%), compared to White and Coloured youths (less than 10%).

Burns, Godlonton, and Keswell (2010) argue that, in such a context of high unemployment, it may be rational to decide not to actively search for work through formal methods if the cost of such search is high and there are significant opportunities to find work through networks. They also study the productivity of social networks in the context of the South African labour market, defining social

[7] At the end of 2010, Statistics South Africa reported unemployment for the last quarter of that year at 24% (defined as individuals aged 15-65 searching for work and not currently working).

networks as language groups and finding that its use increases employment probability between 3-12%. The productivity of social networks in South Africa is also studied by Magruder (2010), who focuses on the role of the employment status of fathers on the employment probabilities of sons, and finds that a 10% growth rate in the father's industry led to an increase of 3-4% in the employment probability of sons. His conclusion that fathers are critical network members for their sons has important implications for the persistence of inequality in South Africa which, as shown by Leibbrandt and Woolard (2001), is determined primarily by the functioning of the labour market. The higher productivity of social networks in terms of employment probability would then limit intergenerational mobility (Magruder, 2010), creating a cycle that is akin to the mechanism of inequality traps presented in Bourguignon, Ferreira, and Walton (2007).[8]

However, the variation in employment probability is not the only way in which social networks may relate to inequality. Wage inequality among the employed may also matter if jobs found through networks vary systematically from jobs found through formal search or, more simply, if the use of contacts does not improve the quality of the match between firms and job seekers. Previous studies have not related wages and job search method to explore this possibility.

This paper uses data from the Cape Area Panel Study (CAPS), described in Lam, Seekings, and Sparks (2008), to address the question of whether the use of social networks has an impact on wages. [9] CAPS is a longitudinal study of a random

[8]Recent theoretical work suggests that the implications of the use of social networks in the labour market are in fact not trivial, and can range from differences in the aggregate unemployment rate (Kuzubas, 2010) to long run inequality between groups (Calvo-Armegnol and Jackson, 2004, Arrow and Borzekowski, 2003).

[9]The Cape Area Panel Study Waves 1-2-3 were collected between 2002 and 2005 by the University of Cape Town and the University of Michigan, with funding provided by the US National Institute for Child Health and Human Development

sample of 4758 youth aged 14-22 in 2002 living in the Cape Town Metropolitan Area. The study focuses on a wide range of important and related outcomes for youth during the transition to adulthood, including schooling and labour market participation. All youth were initially interviewed in 2002 (wave 1) and then in 2003-2004 (wave 2), 2005 (wave 3) and 2006 (wave 4). [10]

In this paper, we limit the analysis to the effect of social networks on the wages of respondents from 15 to 24 years of age who were employed full time. This age interval corresponds to the United Nations definition of "youth", a period best understood as one of transition from the dependence of childhood to the independence of adulthood. [11] The emphasis on independence as a defining characteristic of adulthood motivates our emphasis on full time employment.

As discussed at greater length in the World Development Report 2007 (World Bank, 2006), the transition to adulthood is an economically important time for at least two reasons. Firstly, youth begin to reap the rewards of investment in human capital as they make critical decisions, in an uncertain environment, about when to leave education and enter the labour market. Part of that uncertainty derives from youth inexperience with employment and the search for employment:

and the Andrew W. Mellon Foundation. Wave 4 was collected in 2006 by the University of Cape Town, University of Michigan and Princeton University. Major funding for Wave 4 was provided by the National Institute on Aging through a grant to Princeton University, in addition to funding provided by NICHD through the University of Michigan.

[10]We do not use data from wave 1 in the analysis as the questions regarding the key variables of interest, job finding method, were not coded in the same way as in subsequent waves. After the first versions of this paper were written, one more wave (wave5, 2009) became available. However, job history of respondents was not collected in a way similar to previous waves. For these reasons, the analysis is limited to the period 2003-2006 (waves 2 to 4).

[11]The lower bound of this interval also matches the age at which schooling stops being compulsory while the upper bound matches the one in the definition of youth by the United Nations.

the low quality of a match with a job may only be discovered by experience. Secondly, early bad experiences in the transition to the labour market, such as long periods of unemployment or jobs with low wages and low skills, are scarring experiences, disadvantaging those who suffer them in terms of their future labour market outcomes. For that reason, whether social networks act as mechanisms that facilitate good matches or rather simply allocate workers to jobs at low cost (and potentially limit the range of matches available to prospective workers) is potentially more important to youth than to older workers.

Table 3.1 presents information on both job and individual characteristics in our sample. The first observation is that Whites' wages are about twice that of Africans and Coloureds. [12] Such differences in wages between race groups may be due, at least in part, to systematic differences in education. The second observation is that while White youth have, on average, slightly more years of education than Africans and Coloureds, this difference is not very large, an unsurprising fact given what others have written about generally high enrolment rates in primary and high school across all race groups (Anderson, Case, and Lam., 1999). However, years of education tells us nothing about differences in educational quality. We can capture some of these differences in the results of the literacy and numeracy ability test conducted by CAPS. [13] As shown, there are wide differences across races, with the mean literacy and numeracy score for Whites (close to 85%) being much higher than for Africans (barely above 50%).

Apart from differences in education, differences in family background may lead

[12]Hourly pay in Rands was calculated from data on monthly pay and hours worked each month. The sample includes only those participants that provided data on both monthly pay and monthly hours worked.

[13]The literacy and numeracy score measures the number of correct answers in a test consisting of 45 questions, conducted in the the first wave of CAPS.

to differences in a range of personal characteristics (or in the quality of a network) that may be valued in the labour market. Again, we find large differences across races, with a higher proportion of Whites coming from a financially well-off background and living mostly in formal urban areas throughout their lives than Africans. [14] Parents of White youth are also more educated (again measured in years of education) than parents of youth of other race groups, and the differences are much larger than in the current generation. The data on educational background of mothers and fathers allow us to address one other background difference: although the importance of missing values for father's education is always higher than for mother's education, this difference is relatively small in the case of Whites (354 observations on mother's education versus 326 observations on father's education) and relatively large in the other two race groups (2128 versus 1577 and 778 versus 511 in the case of Coloureds and Africans, respectively). If we are willing to speculate that such differences reflect that fewer youths are connected with their fathers than with their mothers while growing up, then such differences are relatively minor in the case of White youth, but relatively important in the case of Coloured and Black youth. This result seems particularly relevant in light of Magruder's (2010) conclusion regarding the importance of fathers as job contacts for their sons.

A key advantage of CAPS for the purposes of this paper is the existence of a complete, month by month, labour market history constructed for the years 2003-

[14] "Family well-off during childhood" corresponds to an individual who responded either "very comfortable" or "comfortable" when asked to characterise their family's financial situation as a child, while "Family poor during childhood" corresponds to those who responded either "very poor" or "poor" to that same question. "Lived mostly in formal urban areas" corresponds to those who reported having spent most of their life in formal urban areas, as opposed to rural or informal urban areas.

2006 (waves 2 to 4) which also includes detailed information about job finding methods. [15] This information is used in three ways. Firstly, it allows us to construct different measures of work experience, including accumulated experience in the same occupation and in the same industry, rather than simpler measures of overall experience in the labour market. Secondly, it allows us to construct a set of industry and occupation fixed effects which help to control for the possibility that firms in different industries may have different search strategies for different jobs. Finally, it allows us to quantify the importance of job search method on labour market outcomes.

Table 3.1: Summary Statistics: Full time Employed Respondents by Race

Variables	**African**	**Obs**	**Coloured**	**Obs**	**White**	**Obs**
Wages (Rands per hour)	8.68	900	12.15	2351	21.26	361
	(7.25)		(7.64)		(21.98)	
Male	0.53	900	0.52	2351	0.48	361
	(0.50)		(0.50)		(0.50)	
Years of education	10.60	875	10.35	2335	12.12	355
	(1.90)		(2.03)		(1.50)	
Literacy and numeracy score	23.31	889	28.40	2338	38.00	356
	(7.63)		(7.47)		(5.13)	
Urban	0.43	895	0.96	2323	0.96	361
	(0.49)		(0.19)		(0.20)	
Family well off during childhood	0.24	835	0.54	2126	0.90	232
	(0.42)		(0.50)		(0.31)	
Family ok during childhood	0.51	835	0.42	2126	0.10	232
	(0.50)		(0.49)		(0.30)	
Family poor during childhood	0.25	835	0.04	2126	0.00	232
	(0.43)		(0.20)		(0.07)	
Mother's Education	7.95	778	8.36	2128	12.08	354
	(3.14)		(2.64)		(1.83)	
Father's Education	7.00	511	8.50	1577	12.57	326
	(3.76)		(3.02)		(2.01)	
Experience in current job	9.67	845	13.14	2319	12.31	349
	(5.10)		(11.29)		(11.30)	
Previous experience	2.01	872	6.64	2317	9.92	352

Continued on next page

[15] Observations of self employment, working for the family business, missing observations and the category "other", have been dropped. We also dropped those jobs where the respondent was in contact with the firm because of past employment with them.

Table3.1 – continued from previous page

Variables	African	Obs	Coloured	Obs	White	Obs
	(11.51)		(8.91)		(10.39)	
Experience in same occupation	12.02	853	16.75	1886	17.47	337
	(11.57)		(12.95)		(13.95)	
Experience in same industry	11.62	555	15.00	1750	14.31	299
	(11.64)		(12.60)		(12.44)	
Relative	0.32	795	0.36	2121	0.25	291
	(0.47)		(0.48)		(0.43)	
Friend	0.36	795	0.38	2121	0.33	291
	(0.48)		(0.49)		(0.47)	
Formal	0.32	795	0.26	2121	0.41	291
	(0.47)		(0.44)		(0.49)	
Occ: Legislators, officials, managers	0.03	853	0.03	1886	0.06	337
	(0.18)		(0.16)		(0.23)	
Occ: Professionals	0.02	853	0.05	1886	0.13	337
	(0.16)		(0.21)		(0.34)	
Occ: Technicians and assoc. professionals	0.03	853	0.07	1886	0.17	337
	(0.16)		(0.26)		(0.37)	
Occ: Clerks	0.13	853	0.17	1886	0.15	337
	(0.33)		(0.38)		(0.36)	
Occ: Service workers and shop sales workers	0.24	853	0.14	1886	0.29	337
	(0.43)		(0.35)		(0.45)	
Occ: Skilled agriculture and fishery	0.02	853	0.01	1886	0.00	337
	(0.13)		(0.08)		(0.06)	
Occ: Craft and related trades workers	0.13	853	0.20	1886	0.09	337
	(0.34)		(0.40)		(0.29)	
Occ: Plant and machine operators	0.05	853	0.07	1886	0.01	337
	(0.21)		(0.26)		(0.12)	
Occ: Elementary	0.31	853	0.22	1886	0.08	337
	(0.46)		(0.41)		(0.27)	
Occ: Other	0.04	853	0.04	1886	0.02	337
	(0.04)		(0.19)		(0.14)	
Ind: Private households	0.04	555	0.01	1750	0.02	299
	(0.20)		(0.11)		(0.13)	
Ind: Agriculture, Hunting, Forestry & fishing	0.01	555	0.01	1750	0.00	299
	(0.10)		(0.12)		(0.06)	
Ind: Manufacturing	0.10	555	0.31	1750	0.11	299
	(0.31)		(0.46)		(0.31)	
Ind: Electricity, gas and water supply	0.01	555	0.00	1750	0.00	299
	(0.07)		(0.06)		(0.06)	
Ind: Construction	0.11	555	0.11	1750	0.04	299
	(0.32)		(0.31)		(0.20)	
Ind: Wholesale and retail trade	0.45	555	0.31	1750	0.451	299
	(0.50)		(0.47)		(0.49)	
Ind: Transport, storage and communication	0.03	555	0.02	1750	0.01	299
	(0.17)		(0.16)		(0.08)	
Ind: Financial intermediation, insurance, real estate	0.13	555	0.12	1750	0.23	299
	(0.34)		(0.33)		(0.42)	
Ind: Community, social and personal services	0.11	555	0.08	1750	0.18	299

Continued on next page

Table3.1 – continued from previous page

Variables	African	Obs	Coloured	Obs	White	Obs
	(0.31)		(0.28)		(0.38)	
Standard deviation in brackets						

Regarding this last use of the data, we consider three mutually exclusive possibilities: formal methods, assistance from a relative or assistance from a friend. Formal methods comprise responding to a newspaper advert, getting work through an employment agency, submitting CV's directly to, or visiting firms. Assistance from 'relatives' comprises either a household member or a relative outside the household 'getting' or having told the youth about the job. This includes both the possibility that the job was at the workplace of the relative or at a different workplace (though we exclude family businesses where youth work directly for their family as these jobs are infrequent and sometimes associated with no formal wage). For jobs at the workplace of the relative, it is unknown how involved the relative might have been in the hiring process. They may or may not have directly recommended the job seeker to the employer. Assistance from friends is defined similarly to assistance from relatives, with the category excluding household members and relatives. We will refer to friends and relatives that link a job seeker with a job as 'referees'. The two variables, 'friends' and 'relatives', express the importance of social networks and differ in the assumed social proximity between the job seeker and the referee, as we assume that relatives are socially closer than friends. What Table 3.1 shows is that there are important differences in terms of how youth find a job: although it is clear that all race groups use social networks to find a large portion of their jobs, Whites successfully used formal methods in about 40% of the jobs while the corresponding number for Africans and Coloureds is slightly less

than 1/3 and slightly above 1/4 of the jobs, respectively. There are also differences regarding the relative importance of relatives versus friends: conditional on finding a job through social networks, Whites are slightly more reliant on friends (57% of their network found jobs are through friends) than Coloureds and Africans (52% and 53%, respectively).

Before we present the analysis of the relation between search method and wage, it is important to qualify the importance of social networks. Table 3.2 shows that, of the 1132 individuals who had more than one job (representing more than 80% of our sample of individuals), 193 (or 17% of the sub-sample) found all jobs using formal methods. If we add to this number the 279 individuals (25% of the sub-sample) who found jobs using a mix of formal methods and social contacts (either friends, relatives or both), we see that over 40% of the individuals in the sample have some experience with formal search methods. As expected, given what is stated above, this percentage is slightly higher for Whites (64%) than Africans and Coloureds (43% and 39% respectively).

Table 3.3 further qualifies the importance of social networks. It presents the choice of job search method by respondents who are full time employed and actively searching for a different job, by how they found their last job. The most noticeable aspect is that in only 3% of these searches do respondents rely exclusively on social networks, and this is true irrespective of whether they had been successful in using this method in securing their current job. In contrast, the rule is to use both formal methods and social networks, a strategy adopted in almost 80% of the cases for which we have information. Similarly, table 3.3 presents the job search methods of the full time employed by education level. Regardless of education level, there is a strong tendency to use both formal search and network search in combination.

Table 3.2: Number of individuals by job finding method

	All	African	Coloureds	Whites
Found jobs using relatives, friends and formal methods	27	4	19	4
Found jobs using relatives and formal search	106	25	122	10
Found jobs using friends and formal search	146	34	95	17
Found jobs using friends and relatives	159	27	122	10
Found all jobs using relatives	247	72	171	4
Found all jobs using friends	254	69	170	15
Found all jobs using formal methods	193	66	106	21
Total	1132	297	755	80
Used formal methods at least once	472	129	292	51
	(41.7%)	(43.4%)	(38.7%)	(63.8%)

Also, for all races and for all education levels, exclusive network search is never more popular than exclusive search with formal methods. Clearly, there does not seem to exist any specialization in job search method (either the use of formal methods or social networks) in this sample. [16]

Tables 3.5 and 3.6 seem to suggest that this conclusion is not specific to any occupation or industry. Table 3.5 also allows us to conclude about the relative importance of social networks in finding jobs with different skill levels. [17] There is some evidence that differences in required skills lead to privileging hiring through

[16]These results, taken together with the importance of social networks as evidenced in table 3.1, seem to lend support to the conclusion that social networks are productive, in the sense that, when used, they tend to lead to higher offer rates. See Mouw (2003) for a review and Fernandez, Castilla, and Moore (2000), Fernandez and Weinberg (1997), Petersen and Seidel (2000) for examples.

[17]Occupations are classified according to the South African Standard Classification of Occupations. These are mapped into skill levels using an adaptation of the International Standard Classification of Occupation due to Ziervogel and Crankshaw (2009) and also used by Keswell, Girdwood, and Leibbrandt (2012).

Table 3.3: Current job search method by how last job was found

Current search:	Network search only				Formal search only				Both search methods			
Last job was found using:	All	Africans	Coloureds	Whites	All	Africans	Coloureds	Whites	All	Africans	Coloureds	Whites
Relative found	8	2	6	0	27	6	17	4	190	38	146	6
Friend found	10	0	8	2	35	10	21	4	212	54	144	14
Formal found	3	2	1	0	38	13	21	4	115	36	73	6
total	21	4	15	2	100	29	59	12	517	128	363	26

Table 3.4: Current job search method by years of education

Current search:	Network search only				Formal search only				Both search methods			
Years of education:	All	Africans	Coloureds	Whites	All	Africans	Coloureds	Whites	All	Africans	Coloureds	Whites
9 yrs or less	13	3	9	1	38	9	27	2	181	35	145	4
10 yrs	2	0	2	0	13	3	8	2	59	17	41	1
11 yrs	2	1	1	0	3	3	0	0	45	22	23	0
12 yrs	4	0	3	1	34	10	20	4	208	50	138	20
13-14 yrs	0	0	0	0	3	2	1	0	14	3	10	1
15-16 yrs	0	0	0	0	9	2	3	4	10	2	6	3
total	21	4	15	2	100	29	59	12	517	128	363	26

Table 3.5: Job finding method by occupation

Variable	Obs	Networks		Skill
		Proportion	SD	Level
Occ: Legislators, officials, managers	86	0.63	0.48	4
Occ: Professionals	123	0.61	0.49	4
Occ: Technicians and associate professionals	174	0.56	0.50	3
Occ: Clerks	436	0.60	0.49	2
Occ: Service workers and shop sales workers	487	0.58	0.49	2
Occ: Skilled agriculture and fishery	22	0.86	0.34	2
Occ: Craft and related trades workers	476	0.85	0.36	2
Occ: Plant and machine operators	173	0.83	0.38	2
Occ: Elementary	648	0.80	0.40	1
Occ: Other	96	0.74	0.44	1
Observations	2806			

Table 3.6: Job finding method by industry

Variable	Obs	Networks	
		Proportion	SD
Ind: Private households	47	0.79	0.41
Ind: Agriculture, Hunting, Forestry & fishing	28	0.91	0.29
Ind: Manufacturing	570	0.81	0.39
Ind: Electricity, gas and water supply	7	0.52	0.50
Ind: Construction	239	0.88	0.33
Ind: Wholesale and retail trade	819	0.67	0.47
Ind: Transport, storage and communication	62	0.81	0.39
Ind: Financial intermediation, insurance, real estate	311	0.63	0.48
Ind: Community, social and personal services	205	0.50	0.50
Observations	2288		

social networks, but it points in an opposite relation from what we would expect from the "good matches" hypothesis: the importance of social networks increases as the skills level of the job decreases (79% of the jobs in the lowest skill category are filled through social networks, substantially more than in the highest skill category, 62%).

Summarizing this discussion so far, while networks are widely used and are productive, there does not seem to exist a strong preference for specialization, either at the level of individuals (or firms, defined by industry or occupation) in terms of searching (or hiring) through networks. In the next section we will show that, nevertheless, finding a job through social networks does have an important and independent impact on wages.

3.4 Empirical Analysis

We are interested in estimating a wage equation of the type

$$W_{ij} = \alpha X_i + \gamma Z_j + \beta(S * R)_{ij} + \delta T_{ij} + \theta_i + \varepsilon_{ij} \tag{3.1}$$

where W_{ij} is the log of wage of individual i in job j (in Rands per hour), X_i is a vector of individual characteristics that are valued in the labour market (in this case gender and years of education), Z_j is a vector of job characteristics (in practice, a vector of industry and occupation fixed effects), T_{ij} is a vector of time variant characteristics (previous experience and tenure in current job, as well as year fixed effects), and, finally, $(S * R)_{ij}$ are our main variables of interest, the interaction between race (African, Coloured and Whites) and the method used to find the specific job under analysis (Relatives, Friends or Formal).

We estimate equation 3.1 using observations from waves 2, 3 and 4 of CAPS for which we have information on wages, hours worked and the job search method that led to the job.[18] There are multiple observations per individual, forming an unbalanced panel. This structure of the data is reflected in the decomposition of the error term in two components, one of which is individual-specific and time invariant (θ_i).

We start by estimating equation 3.1 using a random effects estimator, under the assumption that all the right hand side variables are uncorrelated with the error term, $\theta_i + \varepsilon_{ij}$. Although this is perhaps a strong assumption, it seems defensible given the nature of our data and the variety of individual and job controls that we

[18]The CAPS data include a "mega job" table that assigns a unique job number to each job per respondent and records the months in which each job was worked in across waves 2, 3 and 4. We excluded observations from wave 1 of CAPS because the job search variable was coded differently.

can include in the estimation. Additionally, random effects estimators are more efficient than the fixed effects estimators and this may matter when, as in our sample, there is limited variation across jobs for some of the independent variables of interest (in this case, job search method) which could lead to large standard errors.

The results, for different sets of control variables, corresponding to different assumptions regarding what firms value and can observe, are presented in table 3.7. In column (1), we present the estimates for a simple model where wage depends on job finding method, human capital variables, gender and race. The human capital variables consist of years of education and months of work experience in current and previous jobs, months of occupation and industry-specific experience, and education quality measured through the literacy and numeracy score. Focusing on the coefficients of the interaction between race and job search method, it is immediately obvious that there are two types of penalties (the comparison category is White $\times$ formal): firstly, and as expected, Africans receive lower wages than Coloured, who in turn receive lower wages than Whites and, secondly, finding a job through social networks (either relatives or friends) carries an additional wage penalty.

It is possible that job characteristics vary systematically according to search methods, that is, that the wage discounts identified in column (1) simply reflect the nature of the work that social networks connect job seekers with. For that reason, we include a set of industry and occupation fixed effects in the model presented in column (2) (although the estimates associated with these variables are omitted). Finally, in column (3), we also control for individual characteristics such as family background, as proxy both to characteristics that are potentially valued in the

Table 3.7: Wages and search method: random effects estimates

Variables	(1)	(2)	(3)
African x relative	-0.587***	-0.584***	-0.383***
	(0.091)	(0.093)	(0.131)
African x friend	-0.632***	-0.626***	-0.448***
	(0.088)	(0.089)	(0.129)
African x formal	-0.427***	-0.409***	-0.320**
	(0.095)	(0.094)	(0.132)
Coloured x relative	-0.318***	-0.349***	-0.194**
	(0.079)	(0.080)	(0.091)
Coloured x friend	-0.342***	-0.369***	-0.177*
	(0.079)	(0.080)	(0.092)
Coloured x formal	-0.219***	-0.239***	-0.107
	(0.081)	(0.081)	(0.093)
White x relative	-0.258**	-0.222*	-0.154
	(0.108)	(0.113)	(0.129)
White x friend	-0.171*	-0.161*	-0.116
	(0.092)	(0.091)	(0.103)
Male	0.175***	0.143***	0.136***
	(0.026)	(0.026)	(0.037)
Years of education	0.071***	0.064***	0.077***
	(0.008)	(0.008)	(0.011)
Literacy and numeracy score	0.008***	0.007***	0.003
	(0.002)	(0.002)	(0.003)
First recorded job	-0.114***	-0.090***	-0.075**
	(0.031)	(0.029)	(0.037)
Not working or studying prior to current job			0.048
			(0.037)
Experience in current job	0.028***	0.024***	0.021***
	(0.003)	(0.003)	(0.004)
Previous experience	0.010***	0.009***	0.009***
	(0.003)	(0.003)	(0.003)
Experience in same occupation	0.004**	0.003*	0.005*
	(0.002)	(0.002)	(0.003)
Experience in same industry	-0.015***	-0.011***	-0.010***
	(0.003)	(0.003)	(0.004)
Urban			0.055
			(0.064)
Family background: well-off			-0.074
			(0.067)
Family background: ok			-0.103*
			(0.061)
Mother's education			0.012
			(0.007)
Father's education			0.011*
			(0.006)
Observations	2,154	2,154	1,145
Number of respondents	1,406	1,406	736

Robust standard errors in parentheses. An interaction variable for "already working at start of CAPS" x "previous experience", time dummies, and a constant were included but not reported Occupation and industry dummies were included in (2) & (3) but are not reported. *** p<0.01, ** p<0.05

Table 3.8: Wages and search method: random effects estimates

Variables	(1)	(2)	(3)
African x relative	-0.444***	-0.442***	-0.318***
African x friend	-0.468***	-0.465***	-0.361***
African x formal	-0.348***	-0.336***	-0.274**
Coloured x relative	-0.272***	-0.295***	-0.176**
Coloured x friend	-0.290***	-0.309***	-0.162*
Coloured x formal	-0.197***	-0.213***	-0.101
White x relative	-0.227**	-0.199*	-0.143
White x friend	-0.157*	-0.149*	-0.110
Observations	2,154	2,154	1,145
Number of respondents	1,406	1,406	736

*** p<0.01, ** p<0.05, * p<0.10

labour market or to the quality of the social contacts to which youth have access to.

It is noticeable that the inclusion of industry and occupation fixed effects does not change our earlier conclusions in any meaningful way (as perhaps would be expected, given the data on job search presented in tables 3.5 and 3.6, discussed above). This suggests that the cause of wage penalties is not simply because social networks allocate prospective workers to low paying jobs. The same is not true when considering the estimates presented in column (3): not only are the estimates of the effect of social networks generally much lower, they are no longer statistically significant in the case of White youth. Apart from father's education, the remaining four family background variables are not individually significant though they do pass a t-test for joint significance at the 5% level of significance. In principle, we

Table 3.9: Respondents with all data vs. respondents with missing data

	All data	Missing data	
Variables	mean	mean	p-value
African	0.165	0.230	0.000
	(0.011)	(0.013)	
Coloured	0.73	0.67	0.004
	(0.013)	(0.014)	
White	0.106	0.097	0.513
	(0.009)	(0.009)	
Relative	0.166	0.174	0.600
	(0.011)	(0.011)	
Friend	0.359	0.400	0.048
	(0.014)	(0.015)	
Formal	0.300	0.258	0.031
	(0.014)	(0.014)	
Male	0.544	0.505	0.073
	(0.015)	(0.016)	
Years of education	10.740	10.374	0.000
	(0.057)	(0.064)	
Literacy and numeracy score	28.882	27.704	0.001
	(0.236)	(0.259)	
Observations	1,145	1009	

Robust standard errors in parentheses.

would prefer this last estimation which has a larger number of control variables, particularly given that we are using a random effects estimator and we may worry about problems of correlation between some of the explanatory variables and the error term. However, it is important to note that due to missing values in some of the variables, the inclusion of these variables substantially reduces the sample size. To check the possibility that some of the differences between coefficients in column (3) and column (2) are due to differences in sample composition, we present in table 3.9, t-tests of the means of the sample in column 3, and the sample of additional individuals in the first two estimations that did not provide data for all the family background variables. It can be seen that for most key demographic variables and two of the job finding variables, these two samples of youth are the same. However, we cannot rule out possible differences between the samples due to unobservables.

Although they are not the focus of this analysis, it seems important to notice that the estimates of the effect of education and experience are positive, as expected, and relatively robust to the inclusion of additional covariates. The main surprise is the general lack of importance of the variables measuring individual background, although we must note again the importance of missing values, which reduces the sample size to roughly half the original sample.

As it is known the coefficients of dummy variables in semi-logarithmic models such as these cannot be directly interpreted. In table 3.8 we present the wage penalties associated with the race $\times$ job search variables using the correction proposed by Halvorsen and Palmquist (1980). We find no evidence that using networks leads to wage premiums, and thus no evidence of the good matches hypothesis. Consistent with the limited choices hypothesis, while more jobs are found through

social networks for Africans than for Whites (see table 3.2 above), this is clearly not because Africans can secure higher wages through social networks. Finally, we notice that, for all races, the coefficients for the 'relative' and the 'friend' interaction variables are not significantly different to each other. This means that there is no evidence from these results that social distance between the worker and their referee has an effect on wage.

A fixed effects estimator may still be preferred if the random effects estimates are inconsistent. For that reason, we estimate a fixed effects specification of equation 3.1

$$W_{ij} = \theta_i + \alpha X_i' + \gamma Z_j + \beta S_{ij} + \delta T_{ij} + \varepsilon_{ij} \tag{3.2}$$

where X_i' includes those individual characteristics that are valued in the labour market and are time variant, S_{ij} is a vector of dummy variables indicating the search method, θ_i is an individual fixed effect and all other variables have the same meaning as above. The estimates, separate for each race group, are presented in table 3.10. [19] It should be noted that while fixed effects estimation controls for all time invariant heterogeneity, we cannot discount the possibility that there is also some unobservable time variant heterogeneity across individuals that may affect our estimates and that is not controlled for. However, that different types of individuals may specialise in different job search methods is likely to be the key source of heterogeneity that may lead to inconsistency of the random effects estimates. This is not a problem with the fixed effects estimates as they exclude individuals that only find jobs using one type of search method. These estimates

[19] Because race is time invariant, it is not possible to include race and search interaction terms in a fixed effects estimation, as we did in table 3.7.

are identified by the existence of changes in successful job search strategy, that is, by individuals who had more than one job between 2002-2006 and who found those jobs using different search methods. As shown in table 3.2, the identifying sample is approximately 1/3 of the total sample.

The coefficients for the main variables of interest (finding a job through friends or through relatives) are presented, as elasticities, in Table 3.11, again using the Halvorsen and Palmquist (1980) correction. Two conclusions emerge from these results. The first is that the use of either friends or relatives carries a penalty in terms of wages. This is true for all race groups, reinforcing our earlier conclusion that there is no evidence to support the good matches hypothesis. Considering that there is much evidence to support positive wage premiums for social network use (Topa, 2011) and that few studies find wage discounts for several major population groups (and not in excess of 10%), this is a significant finding that is broadly consistent with the limited choices hypothesis.

The second conclusion is that, unlike the random effects estimates, the wage penalties associated with the use of relatives as referees are always larger than the penalties associated with the use of friends. The differences are relatively small in the case of Coloured and White youth (as in the random effects estimates) but quite large in the case of African youth. These results suggest that employers may suspect some degree of opportunistic behaviour from referees and react by discounting the value of the information provided by those referees who are closer to job seekers.

Table 3.10: Wages and search method: fixed effects estimates

	(1)	(2)	(3)
Variables	African	Coloured	White
Relative	-0.645***	-0.161**	-0.361**
	(0.185)	(0.077)	(0.154)
Friend	-0.120	-0.135*	-0.347***
	(0.099)	(0.079)	(0.108)
Years of Education	-0.018	0.013	0.001
	(0.063)	(0.044)	(0.071)
First recorded job	-0.326*	-0.141***	0.074
	(0.195)	(0.046)	(0.135)
Experience in current job	0.043***	0.016**	-0.011
	(0.016)	(0.008)	(0.022)
Previous experience	-0.026	0.013*	-0.006
	(0.018)	(0.008)	(0.022)
Experience in same occupation	0.045***	0.007**	0.005
	(0.017)	(0.004)	(0.010)
Experience in same industry	-0.066***	-0.008**	0.005
	(0.019)	(0.004)	(0.009)
Observations	426	1,521	221
Number of respondents	328	941	147

Robust standard errors in parentheses. An interaction variable for "already working at start of CAPS" x "previous experience", time dummies, industry and occupation dummies and a constant were also included but are not reported.

*** p<0.01, ** p<0.05

Table 3.11: Wages and search method: fixed effects estimates

	(1)	(2)	(3)
Variables	African	Coloured	White
Relative	-0.475***	-0.148**	-0.303**
Friend	-0.113	-0.126*	-0.293***
Observations	426	1,521	221
Number of respondents	328	941	147

*** p<0.01, ** p<0.05, * p<0.10

3.5 Conclusion

It is well known that social networks play a significant role in labour markets in many countries, including developing countries. However, the significance of this fact for labour market outcomes, namely wages, is still not clear. This paper is motivated by two competing hypotheses regarding these effects, the "good matches" hypothesis (which predicts that the use of social networks would lead to higher wages) and the "limited choices" hypothesis (which predicts that wages will be lower or unaffected by social networks when people limit their choices to the jobs available through one's social networks). We contribute to this discussion by focusing on the motivation of the referee and the possibility that she/he may not select individuals with the highest possible ability - instead privileging individuals who are socially closer. Firms react by offering lower wages.

We use data from the Cape Area Panel Study (CAPS) of South Africa to estimate the wage effect of contact use and of social proximity. Controlling for individual heterogeneity in characteristics that are valued in the labour market

(through a range of human capital and family background variables) and also for unobserved heterogeneity (through the use of fixed effects), we show that there are penalties associated with the use of social networks for all race groups, ranging from about 11% - 48%. In addition, for Africans these penalties are compounded by the proximity between job seeker and referee: referrals from relatives carry a very large penalty in contrast to no penalty associated with friends. Interestingly, we do not find any convincing evidence of this for Coloureds or Whites. This may reflect either historical discrimination as an inheritance of apartheid, or different labour market conditions faced by African youth. This result also suggests that if the referred individual is from a disadvantaged group, faced with limited employment choices (as is more likely the case for African youth), then the probability that referees will suggest less capable workers should be higher - and so would be the associated wage penalties imposed by employers.

Lastly, it must be noted that even if using contacts leads to lower wages, their use may be associated with a higher probability of finding a job, which may be especially important in the case of young people who are starting their careers. Hence, our results cannot be interpreted as suggesting that youth should not search for jobs using their social connections. Instead, it would be better if youth had more opportunities to productively use formal search methods. Previous literature (Magruder, 2010) has already argued that social networks may perpetuate inequality across generations in South Africa. In addition to this issue of equity, we have identified an inefficiency in the way workers are matched to employment that results from social network use. Together, these two findings suggest that policy makers should consider ways of further supporting formal job search methods that would facilitate relatively better matches and lessen the prevalence of youth being

disadvantaged by limited choices.

Chapter 4

Conclusion

4.1 Conclusion

The two key economic transitions to adulthood regarding education and the labour market both involve problems of imperfect information. Firstly, youth have to decide on the right amount of investment in education and may not have accurate information about its returns, leading to sub-optimal investment. Secondly, youth may use friends and family to search for work as they are not always able to find employment through formal search methods. These informal search methods could be a problem if they lead to poor matches between jobs and workers, indicated by wage discounts.

Both decisions underlying the economic transition to adulthood may possibly be improved through information interventions targeting either of these issues. However, there is limited previous research regarding the best way to design information interventions about the returns to education. Particularly, the potential for such interventions to cause harm has been ignored. In addition, there is limited empirical research on the accuracy of expectations regarding education returns in developing countries. Also, in the context of South Africa there is no previous research on the wage impacts of using social networks to find jobs. More generally, there is no consensus in the empirical literature across other countries regarding whether the impact of using social networks is positive or negative. This thesis investigates these issues in chapters 2 and 3, respectively.

In chapter 2, the potential impact of information provision on the returns to education is analysed using a real options model of investment in grade 12 with data from CAPS on observed and expected earnings conditional on different levels of education. Simulations for the potential impact of four designs of information interventions demonstrate that each intervention may reduce the accuracy of expectations of at least one race group and thus, none can be recommended for the context we consider. The main contribution of this chapter is to show the importance of considering three critical factors when designing an information intervention and anticipating its potential impacts. Unless the sequential nature of education investment choices, heterogeneity in the returns to education and the process by which expectations are updated are all considered *ex ante*, the potential for harm may go undetected. This is true regardless of the accuracy of the information that is to be provided. This, to my knowledge, is the first research to demonstrate that the provision of accurate information (even full information), may result in reduced accuracy of expectations. Particularly concerning in the context we consider, is that underestimation of the returns to education (and hence underinvestment) may be caused by accurate information provision to African youth (who may be considered to be relatively disadvantaged in both education and the labour market).

Further research is needed regarding the effectiveness of different types of information interventions designed to assist youth with education choices. Chapter 2 only considered 4 specific types of intervention designs and their potential impact on expectations accuracy. Firstly, one limitation of most previous research is that the effectiveness of interventions is measured only in terms of changes in expectations and investment decisions. Measuring long term welfare measures such as life

time earnings or job satisfaction may be a more useful indicator when job characteristics other than wages are important, though only one paper, by Borghans, Golsteyn, and Stenberg (2011), has done this. They use variation in study counseling practices in Dutch Schools as an instrumental variable (IV) to evaluate the impact of counseling on the quality of educational choice, measured by future job satisfaction. They find that one standard deviation more of study counselling is associated with a 2 percent reduction in the probability of having made a low quality education investment choice though the exact variation in counselling that leads to this is unknown. Further research could explore the long term impacts of different designs of information interventions.

In addition, further research could explore a broader range of short term impacts from information interventions. For example, Dinkelman, Martínez et al. (2012) find information provision to be largely ineffective when targeting grade 8 students in Chile. They examined the impact of showing DVD's to these students, of young disadvantaged adults who tell their stories of transitions toward college or vocational school. While the presentation did result in increased knowledge of financial aid among the students and also, decreased absenteeism, overall enrolment decisions and grade scores were unaffected. This suggests that some types of information interventions may have a complex range of effects that existing theoretical and empirical research cannot explain. Lastly, beyond having implications for information interventions targeted at education choices, the analysis and results of chapter 2 have much broader implications.

Any decision that could be modelled with a real options framework and be the target of information provision, such as investment in agriculture (Hertzler, 2007), could be analysed *ex ante* using an approach that is similar to ours. Also, when-

ever there is quantifiable heterogeneity in the outcome of an economic decision and that decision is the target of information provision, a similar approach to ours could be used to evaluate the impact of the information provided. For example, recommendations for alcohol consumption before driving are sometimes provided in the measure of standard drinks per hour for men and women, but without quantitative information about the exact nature of this heterogeneity. The incidence of low range drink driving and the defence of some such drivers (that they only had a few drinks and didn't *expect* to be over the limit), suggests that more accurate and detailed information regarding the relation between blood alcohol level and alcohol consumption might be better, though this would depend on the sequential nature of alcohol consumption, and the way in which alcohol consumers update their expectations. In general, the cost of analysing different designs of interventions and their potential impacts *ex ante* should generally be very small compared to testing interventions using field experiments (though the latter has clearly been the preference of recent research). Also, *ex ante* analysis provides the benefit of anticipating negative impacts before they occur and avoiding them through careful intervention design. More experimental research regarding the way people respond to different types of information would also be useful, but it may better to conduct such research in lab experiments where no important life decisions are at stake.

Chapter 3 investigates two competing hypotheses regarding the effect of social networks and wages and consequently, whether or not the widespread use of social networks in the South African labour market is cause for concern. In addition, we contribute to the broader debate on the "good matches" hypothesis (which predicts that the use of social networks would lead to higher wages) and the "limited choices" hypothesis (which predicts that wages will be lower or unaffected by social

networks when people limit their choices to the jobs available through one's social networks). Unlike most previous research, we also address the motivation of the referee and the possibility that s/he may not select individuals that are the most suitable for a particular job. Firms may be willing to hire referred individuals who are not suitable because it is less costly than finding workers through formal mechanisms. As a result however, network found jobs on average should be lower paid than formally found jobs.

We use data from the Cape Area Panel Study (CAPS) of South Africa to estimate the wage effect of contact use and of social proximity. Controlling for individual heterogeneity in characteristics that are valued in the labour market, job characteristics and also for time invariant unobserved heterogeneity, we show that there are penalties associated with the use of social networks for all race groups, ranging from about 11% - 48%. In addition, for Africans these penalties are compounded by the proximity between job seeker and referee: referrals from relatives carry a very large penalty in contrast to no penalty associated with friends. Interestingly we do not find any convincing evidence of this for Coloureds or Whites. This may reflect either historical discrimination as an inheritance of apartheid, or different labour market conditions faced by African youth. This result also suggests that if the referred individual is from a disadvantaged group, faced with limited employment choices (as is more likely the case for African youth), then the probability that referees will suggest less capable workers should be higher - and so should be the associated wage penalties imposed by employers. However, further research is needed on the role played by social networks in labour markets, specifically on the exact mechanisms that may lead to wage penalties.

While chapter 3 did not consider information interventions directly, the identi-

fication of these wage effects, reflecting a potential inefficiency in the way workers are matched to with jobs through social network use, together with the finding of Magruder (2010), that social networks may perpetuate inequality across generations, implies that policy makers should consider ways of further supporting formal job search methods, for example, through policies that provide youth with information about suitable job opportunities. Job search assistance interventions may produce efficiency gains in two ways. Firstly, they may reduce the average duration of unemployment. Secondly, they may lead to better quality matches. While the possibility of the first efficiency gain arises simply from the incidence of unemployment, the possibility of the second efficiency gain depends on the relative matching performance of formal search compared to informal search (as investigated in chapter 3). Little is known regarding the effectiveness of different designs of information interventions targeting job search behaviour, though there are some recent studies in this area. Beam (2013) conducted an information intervention experiment in the Philippines, that tested the provision of a voucher to attend a job fair on randomly treated individuals' job search behaviour. In the two months following the fair, individuals' chance of searching for work was more than doubled. Also, Dammert, Galdo, and Galdo (2013) test the impact of sending job-market information to job-seekers via sms messages in Peru and find that the information increases expectations of finding work. If it is true that these types of information interventions can help youth to find work through formal matching mechanisms, these policies may be particularly applicable in contexts where informal matching mechanisms do not work well (indicated by wage penalties) and are responsible for a large proportion of job matches, such as South Africa. More research in this area would be worthwhile.

Bibliography

BIBLIOGRAPHY

Abbott, M., and C. Beach. 1994. "Wage changes and job changes of Canadian women: Evidence from the 1986-1987 labour market activity survey." *Journal of Human Resources* 29:429–460.

Altonji, J.G. 1993. "The Demand for and Return to Education When Education Outcomes Are Uncertain." *Journal of Labor Economics* 11:48–83.

Anderson, K., A. Case, and D. Lam. 1999. "Causes and Consequences of Schooling Outcomes in South Africa: Evidence from Survey Data." *Social Dynamics* 27:120–121.

Antoninis, M. 2006. "The Wage Effects from the Use of Personal Contacts as Hiring Channels." *Journal of Economic Behavior and Organization* 59:133–146.

Arcidiacono, P. 2004. "Ability sorting and the returns to college major." *Journal of Econometrics* 121:343–375.

Arrow, K.J., and R. Borzekowski. 2003. "Limited Network Connections and the Distribution of Wages." Unpublished, Department of Economics, Stanford University working paper.

Arulampalam, W., P. Gregg, and M. Gregory. 2001. "Unemployment scarring." *The Economic Journal* 111:577–584.

Attanasio, O., and K. Kaufmann. 2009. "Educational choices, subjective expectations, and credit constraints." Working paper, National Bureau of Economic Research.

Attanasio, O.P. 2009. "Expectations and perceptions in developing countries: their measurement and their use." *The American Economic Review* 99:87–92.

Bandiera, O., I. Barankay, and I. Rasul. 2010. "Social Connections and Incentives in the Workplace: Evidence from Personnel Data." *Econometrica* 77:1047–1094.

Beam, E.A. 2013. "Incomplete Information in Job Search: Evidence from a Field Experiment in the Philippines." Unpublished.

Belzil, C. 2007. "The return to schooling in structural dynamic models: a survey." *European Economic Review* 51:1059–1105.

Bentolila, C.M., Samuel, and J. Suarez. 2010. "Social Networks and Occupational Choice." *Economica* 77:20–45.

Betts, J.R. 1996. "What do students know about wages? Evidence from a survey of undergraduates." *Journal of Human Resources*, pp. 27–56.

Beutel, A.M., and K.G. Anderson. 2007. "The educational expectations of South African youth." *Sociological Focus* 40:348–369.

Bewley, T. 1999. *Why wages don't fall during a recession*. Cambridge, MA: Harvard University Press.

Borghans, L., B.H. Golsteyn, and A. Stenberg. 2011. "Does Expert Advice Improve Educational Choice?" In *EALE meeting, Cyprus, September*. pp. 22–24.

Bourguignon, F., F.H. Ferreira, and M. Walton. 2007. "Equity, efficiency and inequality traps: A research agenda." *The Journal of Economic Inequality* 5:235–256.

Branson, N., J. Garlick, D. Lam, and M. Leibbrandt. 2012. "Education and Inequality: The South African Case.", pp. .

Breen, R., and J.H. Goldthorpe. 1997. "Explaining educational differentials towards a formal rational action theory." *Rationality and society* 9:275–305.

Bridges, W., and W. Villemez. 1986. "Informal hiring and income in the labor market." *American Sociological Review* 51:574–582.

Burns, J., S. Godlonton, and M. Keswell. 2010. "Social networks, employment and worker discouragement: Evidence from South Africa." *Labour Economics* 17:336–344.

Calvo-Armegnol, A., and M.O. Jackson. 2004. "The Effect of Social Networks on Employment and Inequality." *American Economic Review* 94:426–454.

—. 2007. "Networks in Labor Markets: Wage and Employment Dynamics and Inequality." *Journal of Economic Theory* 132:27–46.

Calvo-Armegnol, A., and Y. Zenou. 2005. "Job Matching, Social Networks and Word-of-Mouth Communication." *Journal of Urban Economics* 57:500–522.

Cosser, M., and J.L. du Toit. 2002. *From School to Higher Education?: Factors Affecting the Choices of Grade 12 Learners*, vol. 1. HSRC Press.

Daley, B., and B. Green. 2012. "Waiting for News in the Market for Lemons." *Econometrica* 80:1433–1504.

Dammert, A.C., J.C. Galdo, and V. Galdo. 2013. "Digital labor-market intermediation and job expectations: Evidence from a field experiment." *Economics Letters*, pp. .

Delavande, A., X. Giné, and D. McKenzie. 2011. "Measuring subjective expectations in developing countries: A critical review and new evidence." *Journal of Development Economics* 94:151–163.

Dinkelman, T., A. Martínez, et al. 2012. "Investing in schooling in Chile: The role of information about financial aid for higher education." *Review of Economics and Statistics (forthcoming)*, pp. .

Dominitz, J., and C.F. Manski. 1996. "Eliciting Student Expectations of the Return to Schooling." *Journal of Human Resources* 31:1–26.

Eckstein, Z., and K.I. Wolpin. 1999. "Why youths drop out of high school: The impact of preferences, opportunities, and abilities." *Econometrica* 67:1295–1339.

Fafchamps, M., and A. Moradi. 2009. *Referral and job performance: Evidence from the ghana colonial army*. Centre for Economic Policy Research Discussion Paper No. DP7408.

Fernandez, R., E. Castilla, and P. Moore. 2000. "Social Capital at Work: Networks and Employment at a Phone Center." *American Journal of Sociology* 105:1288–1356.

Fernandez, R.M., and N. Weinberg. 1997. "Personal Contacts and Hiring in a Retail Bank." *American Sociological Review* 62:1288–1356.

Granovetter, M.S. 1974. *Getting a job: A study of contacts and careers*. Chicago: University of Chicago Press.

Halvorsen, R., and R. Palmquist. 1980. "The interpretation of dummy variables in semilogarithmic equations." *American Economic Review* 70:474–475.

Heckman, J.J., and G.J. Borjas. 1980. "Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence." *Economica* 47:247–283.

Heckman, J.J., L.J. Lochner, and P.E. Todd. 2006. "Earnings functions, rates of return and treatment effects: The Mincer equation and beyond." *Handbook of the Economics of Education* 1:307–458.

Heintz, J., and D. Posel. 2008. "Revisiting informal employment and segmentation in the South African labour market." *South African Journal of Economics* 76:26–44.

Hertzler, G. 2007. "Adapting to climate change and managing climate risks by using real options." *Crop and Pasture Science* 58:985–992.

Holzer, H. 1987. "Job search by employed and unemployed youth." *Industrial and Labor Relations Review* 40:601–611.

Ioannides, Y.M., and L.D. Loury. 2004. "Job Information Networks, Neighborhood Effects, and Inequality." *Journal of Economic Literature* 42:1056–1093.

Jensen, R. 2010. "The (perceived) returns to education and the demand for schooling." *The Quarterly Journal of Economics* 125:515–548.

Johnson, W. 1978. "A theory of job shopping." *Quarterly Journal of Economics* 92:261–78.

Jovanovic, B. 1979. "Job matching and the theory of turnover." *Journal of Political Economy* 87:972–990.

Keane, M.P., and K.I. Wolpin. 1997. "The career decisions of young men." *Journal of Political Economy* 105:473–522.

Keith, K., and A. McWilliams. 1995. "The wage effects of cumulative job mobility." *Industrial and Labor Relations Review* 49:121–137.

Kerr, S.P., T. Pekkarinen, M. Sarvimäki, and R. Uusitalo. 2013. "Educational Choices and Information on Labor Market Prospects: Evidence from a Randomized Field Experiment." Unpublished.

Keswell, M., S. Girdwood, and M. Leibbrandt. 2012. "Educational Inheritance and the Distribution of Occupations: Evidence from South Africa." Unpublished, Southern Africa Labour and Development Research Unit WP73, The University of Cape Town.

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

Kingdon, G., and J. Knight. 2000. *Are Searching and Non-searching Unemployment Distinct States when Unemployment is High?: The Case of South Africa*. Centre for the Study of African Economies, Institute of Economics and Statistics, University of Oxford.

—. 2006. "The Measure of Unemployment When Unemployment is High." *Labour Economics* 13:291–315.

—. 2007. "Unemployment in South Africa, 1995–2003: causes, problems and policies." *Journal of African Economies* 16:813–848.

—. 2004. "Unemployment in South Africa: The nature of the beast." *World Development* 32:391–408.

Korenman, S., and S. Turner. 1996. "Employment contacts and minority-white wage differences." *Industrial Relations* 35:106–102.

Kuzubas, T.U. 2010. "Endogenous Social Networks in the Labor Market." Working paper, Bogazici University, Department of Economics.

Lam, D., C. Ardington, and M. Leibbrandt. 2006. "Progress through School in Urban South Africa: Evidence from Panel Data'." In *DPRU/TIPS Conference on" Accelerated and Shared Growth in South Africa: Determinants, Constraints and Opportunities", Johannesburg*. pp. 18–20.

Lam, D., J. Seekings, and M. Sparks. 2008. *The Cape Area Panel Study: Overview and Technical Documentation for Waves 1-2-3-4*. The University of Cape Town.

Lawler, J.J., H.C. Jain, C.S.V. Ratnam, and V. Atmiyanandana. 1995. "Human resource management in developing economies: a comparison of India and Thailand." *The International Journal of Human Resource Management* 6:319–346.

Leibbrandt, M., and I. Woolard. 2001. "The Labour Market and Household Income Inequality in South Africa: Existing Evidence and New Panel Data." *Journal of International Development* 13:671–689.

Levhari, D., and Y. Weiss. 1974. "The effect of risk on the investment in human capital." *The American Economic Review* 64:950–963.

Lindset, S., and E. Matsen. 2011. "Human capital investment and optimal portfolio choice." *The European Journal of Finance* 17:539–552.

Loury, L.D. 2006. "Some Contacts Are More Equal than Others: Informal Networks, Job Tenure and Wages." *Journal of Labor Economics* 24:299–318.

Magruder, J., and L. Beaman. 2012. "Who gets the job referral? Evidence from a social networks experiment." *American Economic Review* 102:3574–3593.

Magruder, J.R. 2010. "Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa." *American Economic Journal: Applied Economics* 2:62–85.

Manski, C.F. 1993. "Adolescent econometricians: How do youth infer the returns to schooling?" In C. T. Clotfelter and M. Rothschild, eds. *Studies of supply and demand in higher education*. University of Chicago Press, pp. 43–60.

—. 2004. "Measuring expectations." *Econometrica* 72:1329–1376.

Marmaros, D., and B. Sacerdote. 2002. "Peer and Social Networks in Job Search." *European Economic Review* 46:870–879.

Marsden, P., and E. Gorman. 2001. "Social networks, job changes and recruitment." In I. Berg and A. L. Kallerberg, eds. *Sourcebook of Labor Markets: Evolving structure and processes*. New York: Kluwer Acad//Plenum.

McPherson, J.M., L. Smith-Lovin, and J.M. Cook. 2001. "Birds of a feather: homophily in social networks." *Annual Review of Sociology* 27:415–444.

Montgomery, J.D. 1991. "Social Networks and Labor Market Outcomes." *American Economic Reveiw* 81:1408–1418.

Morrow, S., S. Panday, and L. Richter. 2005. *Umsobomvu Youth Survey- Where we're at and where we're going*. Human Sciences Research Council.

Mouw, T. 2003. "Social Capital and Finding a Job: Do Contacts Matter?" *American Sociological Review* 68:868–898.

Mwabu, G., and T.P. Schultz. 2000. "Wage premiums for education and location of South African workers, by gender and race." *Economic Development and Cultural Change* 48:307–334.

Nguyen, T. 2008. "Information, role models and perceived returns to education: Experimental evidence from Madagascar." *MIT Working Paper*, pp. .

Oreopoulos, P., and R. Dunn. 2013. "Information and College Access: Evidence from a Randomized Field Experiment*." *The Scandinavian Journal of Economics* 115:3–26.

Pellizzari, M. 2010. "Do friends and relatives really help in getting a good job?" *Industrial and Labor Relations Review* 63:494–510.

Petersen, I.S., Trond, and M.D.L. Seidel. 2000. "Offering a Job: Meritocracy and Social Networks." *American Journal of Sociology* 106:763–816.

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

Pugatch, T. 2012. *Bumpy rides: School to work transitions in south africa*. IZA Discussion Paper.

Rees, A. 1966. "Information Networks in Labor Markets." *American Economic Review* 56:559–566.

Rouse, C.E. 2004. "Low-Income Students and College Attendance: An Exploration of Income Expectations*." *Social Science Quarterly* 85:1299–1317.

Saloner, G. 1985. "Old boy networks as screening mechanisms." *Journal of Labor Economics* 3:255–267.

Schultz, T.W. 1961. "Education and economic growth." In N. B. Henry, ed. *Social forces influencing American education*. University of Chicago press Chicago, IL, chap. 3.

Seekings, J., and N. Nattrass. 2005. *Class, Race, and Inequality in South Africa*. New Haven and London: Yale University Press.

Simon, C., and J. Warner. 1992. "Matchmaker, matchmaker: The effect of old boy networks on job match quality, earnings, and tenure." *Journal of Labor Economics* 10:306–330.

Smith, S. 2000. "Mobilizing Social Resources: Race, Ethnic, and Gender Differences in Social Capital and Persisting Wage Inequalities." *The Sociological Quarterly* 41:509–537.

Stange, K.M. 2012. "An empirical investigation of the option value of college enrollment." *American Economic Journal: Applied Economics* 4:49–84.

Statistics South Africa. 2008. *Statistical Release P0211.1. Quarterly Labour Force Survey. Additional Aspects of the Labour Market in South Africa: Informal Employment; Underemployment and underutilised Labour; Unemployment*. Pretoria: Statistics South Africa.

Topa, G. 2011. "Labour Markets and Referrals." *Handbook of Social Economics* 2:1193–1221.

Topel, R., and M. Ward. 1992. "Job mobility and the careers of young men." *Quarterly Journal of Economics* 107:439–479.

Viscusi, K. 1980. "A theory of job shopping: A Bayesian perspective." *Quarterly Journal of Economics* 94:609–614.

Webbink, D., and J. Hartog. 2004. "Can students predict starting salaries? Yes!" *Economics of Education Review* 23:103–113.

Weisbrod, B.A. 1962. "Education and investment in human capital." *The Journal of Political Economy* 70:106–123.

Wiswall, M., and B. Zafar. 2011. "Belief updating among college students: evidence from experimental variation in information." Working paper, Staff Report, Federal Reserve Bank of New York.

World Bank. 2006. *World development report 2007: development and the next generation*. Washington, DC: World Bank.

Ziervogel, C., and O. Crankshaw. 2009. "Inter-generational Occupational Mobility Amongst Blacks in the Mitchells Plain Magisterial District, Cape Town: Evidence from the Khayelitsha/Mitchells Plain Survey." *Urban Forum* 20:235–251.