



A Structural Approach to Modelling South African Labour Market Decisions

By Cobus Burger

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

Supervisor: Prof Servaas van der Berg

March 2016

Declaration

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

March 2016

Acknowledgements

I am grateful to my supervisor, Prof Servaas van der Berg, who provided valuable guidance in refining the ideas contained in the dissertation. I would also like to thank Rulof Burger (who co-supervised my honours and master's theses), for not only introducing me to structural modelling, but also for continually assisting me with the pesky technical issues that are bound to arise when one strays from the beaten track and decides to develop one's own models.

Earlier versions of this research have been presented at the annual Centre for the Studies of African Economies Conference in Oxford, the ESSA conference in Bloemfontein, the Microeconomic Analysis of South African Data conference in Durban as well as the SALDRU workshop at the University of Cape Town.

Abstract

Youth unemployment is high in South Africa and especially high among black males. The slow absorption of young black males into the employment is somewhat surprising given that the descriptive statistics suggest that employment mobility in South Africa is high. In the first chapter, I investigate whether the labour market is truly as mobile as reported or whether the transition estimate is rather a reflection of misclassification error or unobserved individual heterogeneity. Thereafter, in chapter two, I proceed to examine the role of reservation wages on unemployment. Unlike previous studies I do not make use of self-reported reservation wages. Instead I use a job search model to recover the reservation wages that are consistent with the behaviour we observe in the labour market. In the final chapter, I look at the role of education on labour market outcomes. More specifically, at whether ability bias is present and whether current estimates are inflated. I do so through a dynamic programming model that mimics the schooling decision for forward-looking optimizing agents.

Opsomming

Werkloosheid onder die jeug is hoog in Suid-Afrika en veral hoog onder swart mans. Die stadige indiensnamekoers van swart mans is ietwat verrassend gegewe dat konvensionele modelle daarop dui dat die mobiliteit van indiensname in Suid-Afrika hoog is. In die eerste hoofstuk, word daar ondersoek of die arbeidsmark werklik so mobiel is soos gerapporteer word en of die huidige beraamings van mobiliteit nie liever 'n aanduiding van metingsfout of heterogeniteit is nie. Daarna, in hoofstuk twee, word daar gekyk na die rol van reserwelope op werkloosheid. Liever as om gebruik te maak van selfgerapporteerde reserwelope ontwikkel ek 'n indiensname-model om die reserwelope te vind wat ooreenstem met die gedrag in die arbeidsmark. In die laaste hoofstuk, kyk ek na die rol van onderwys op arbeidsmark uitkomst. Ek ondersoek of die huidige oprengskoers vir opvoeding sydig is wanneer mens versuim om te kontroleer vir die effek van vermoens. Ek gebruik 'n dinamiese programmering model wat my in staat stel om die optimale hoeveel opvoeding vir toekomsgerigte individu te beraam.

Table of Contents

Introduction	1
i. The Contribution of this Thesis to our Knowledge.....	2
ii. Why Structural Modelling?	3
iii. Why the Focus on Black Males?	4
iv. States, Decisions and Exogenous Transitions	4
Chapter 1: Employment Mobility in South Africa: A Reconsideration	6
1. Introduction.....	6
2. Previous Work.....	7
2.1. Transitioning out of Employment.....	7
2.2. Transitioning into Employment	8
2.3. Validity of Markov Assumption	9
2.4. Misclassification	9
2.4. Heterogeneity	10
3. The Data	11
4. Descriptive analysis.....	12
4.1. Patterns of Employment.....	12
4.2. Entry and Exit Rates	12
4.2.1. Variation in Entry and Exit Rates by Age	13
4.2.2. Variation in Entry and Exit Rates by Recent Employment History.....	14
4.2.3. Variation in Entry and Exit Rates by Education Level	16
5. Classic Transition Estimates.....	17
5.1. Justification for Only Having Two Groups	17
5.2. Classic Estimates.....	18
5.3. Concerns with Classic Estimates	19
6. Refined Model.....	21
6.1. Constant Entry and Exit Rates	24
6.2. Misclassification	24
6.3. Observed Heterogeneity	28
6.4. Unobserved Heterogeneity.....	30
7. Conclusion.....	32
Appendix 1.....	34
Appendix 1.A: Additional Tables.....	34
Appendix 1.B: Cleaning ‘Tenure’ and ‘Time Gap’.....	38
Appendix 1.C: Linear Probability Model	38
Appendix 1.D: The Stationarity Assumption	39
Appendix 1.E: The Markov Assumption.....	40
Appendix 1.F: The Initial State Assumption	41
Appendix 1.G: Identification.....	41

Chapter 2: Using Dynamic Programming to Model Job Search and Reservation Wages	43
1. Introduction.....	43
2. Previous Work.....	44
2.1. South African Literature	44
2.2. Job Search Literature.....	45
3. A Dynamic Job Search Model.....	47
3.1. Home Production and Search Cost.....	48
3.2. Reward for Working (Wage)	49
3.3. Job Offers	49
3.4. On-the-Job Search.....	50
4. Estimation Issues	50
4.1. Difference between Offered Wages and Observed Wages.....	50
4.1.1. Measurement Error	50
4.1.2. Selection Bias.....	51
4.2. Recovering Reservation Wages	51
4.3. The Likelihood Function	53
4.3.1. Remaining Non-Employed	53
4.3.2. Transitioning into Employment	53
4.3.3. Transitioning into a New Job.....	54
4.3.4. Keeping the Same Job.....	54
4.3.5. Transitioning Out of Employment	54
4.3.6. Searching for Employment.....	54
4.4. Identification	55
5. Empirical Analysis	56
5.1. Data	56
5.2. Descriptive Statistics.....	57
5.3. Fitting the Data to the Model	60
5.4. Homogenous Model.....	60
5.4.1. Wages	61
5.4.2. Household Utility and Reservation wages.....	61
5.4.3. Job Arrivals and Refusals.....	63
5.5. Heterogeneous Model.....	64
6. Conclusion.....	67
Appendix 2.....	68
Appendix 2.A: Additional Figures and Tables	68

Chapter 3: The Role of Unobserved Ability on Educational Attainment and Labour Market Outcomes	72
1. Introduction.....	72
2. Previous Work.....	73
2.1. Ability Bias.....	73
2.2. Instrumental Variable Approach.....	75
2.3. Structural Models.....	75
2.4. South African Literature.....	76
3. Dynamic School Participation Model.....	77
3.1. Model Structure.....	77
3.2. The Utility of Work and the Returns to Schooling.....	78
3.3. The Utility of Attending School.....	79
3.4. Grade Progression.....	80
3.5. Uncertainty and Information Updating.....	81
3.5.1. Updating Expectation about Schooling Ability.....	81
3.5.2. Updating Expectation about Market Ability.....	82
4. Estimation Issues.....	82
4.1. Bellman Equation.....	82
4.2. Identification.....	84
4.3. The Likelihood Function.....	85
5. Empirical Analysis.....	86
5.1. Data.....	86
5.2. Descriptive Statistics.....	86
5.3. Fitting the Data to the Model.....	89
5.4. Results.....	90
6. Simulations.....	93
6.1. Prospects among Young Black Males.....	93
6.2. Reducing the Noise in the Schooling System.....	94
6.2.1 The Direct Effect: The Probability of Passing.....	95
6.2.2. The Indirect Effect: The Information Revelation.....	96
6.2.3 The Total Effect: The Change in Education.....	96
7. Conclusion.....	98
Appendix 3.....	99
Appendix 3.A: Additional Figures and Tables.....	99
Appendix 3.B: Choosing whether to sit Matric.....	103
Conclusion	104
Bibliography	107

Introduction

Unemployment is notoriously high in South Africa. The current unemployment rate is estimated to be 25% according to Statistics South Africa (StatsSA, 2015a). This rate increases to 35% when using the broad definition of unemployment, which also includes those who are not actively looking for work. Youth unemployment in South Africa is even higher. The narrow unemployment level among the young (between the ages of 15 and 34) is estimated at 37%, while the narrow unemployment level for the adult population is estimated at 17%. (StatsSA, 2015b). Such high levels of unemployment is clearly an important issue that requires attention, both from a research and a policy perspective.

It is often argued that the increase in unemployment among the youth is due to an oversupply of labour relative to demand. Labour force participation has steadily increased since the political transition in 1994. Unfortunately, over the same period there has been a decline in the relative demand for unskilled labour (Burger & von Fintel, 2009). The surplus of unskilled labour caused by the increased participation has made it tougher for young entrants into the labour market to find work, since these individuals are less equipped to compete with older workers. (Mlatsheni & Rospabe, 1999; Bhorat & Oosthuizen, 2007). Alternatively, the higher levels of unemployment among the young could be due to higher reservation wages. Young adults may be financially supported by the rest of their household and therefore be less willing to accept low paying jobs (Rankin & Roberts, 2011). According to the job search literature, unemployment can also arise through imperfect information. If individuals have imperfect information about vacancies and future job offers this will add to the unemployment burden. Similarly, as both employees and employers have preferences, young entrants may take time to find a job that is suitable to their skills and tastes (McCall, 1970).

In this dissertation I plan to study the movement into and out of employment in order to investigate whether the slow absorption into employment among the young is driven by demand or supply side factors. I will do so using a mixture of standard reduced form and structural modelling approaches. In the first chapter I show that, despite slow absorption rates among new entrants into the labour market, the reported mobility levels between surveys are high. In the second chapter I go on to test whether these low absorption rates into employment among the young is due to a low job arrival rate or whether it is due to low acceptance rates. Finally, in the last chapter I allow for endogenous education by developing a structural model that explicitly models the sequential educational enrolment decision faced by students, who need to decide whether it would be better to continue their studies or not.

Throughout the thesis I will contrast the disadvantaged position of young black males relative to that of their white and older counterparts and the unfavourable position of the uneducated relative to that of the educated. All three chapters also explicitly considers the potentially confounding effect of unobservable individual heterogeneity on the behaviour and subsequent labour market outcomes of workers.

i. The Contribution of this Thesis to our Knowledge

Economic research has revealed a preference for simpler methods, partly since the results that are obtained from these methods are usually easier to convey to an audience, and particularly an audience of non-specialists. These descriptive statistics or multivariate regressions are often informative in telling us how the different variables in our model are correlated. Many economist, however, make the mistake of adding a causal interpretation to these estimated correlates or summary statistics. This is problematic since these estimates are usually not causal or economically interpretable. The results are usually not causal, since these models do not even attempt to address the endogeneity concerns that are inherent to most estimates. The results are not economically interpretable since underlying economic theory is required to give meaning to the effects one derive (Keane, 2010).

In this thesis, I contribute to the literature, by choosing three such descriptive results and exposing the inconsistencies that undermine their interpretation. In all three chapters, I look at the descriptive estimates, discuss why I think they may be biased and then attempt to find more reliable estimates that are better equipped in explaining the behavioural outcomes that we observe in the data.

In chapter one, I show that the employment transitions rates that we derive from transitions matrices are too high. If the probability to transition into employment is truly as high as the reported estimates claim, then young individuals should be absorbed into employment at a much quicker rate than what we are currently observing. It is unclear whether the divergence is driven by misclassification or heterogeneity, but what we can say with confidence is that conventional transition rates do not represent an accurate estimate of the probability of transitioning into employment.

In chapter two, I show that current reservation wages, which are usually derived from self-reported data, are inconsistent with the employment and wage data. Individuals often report reservation wages that are higher than their expected earnings and are seen accepting jobs that pay less than their reported reservation wages. In our model we look at the employment transitions between periods and the observed wage distribution to derive a plausible wage offer distribution and job offer arrival rate.

In chapter three, I show that the returns to education estimates are not aligned with the educational attainment figures. Descriptive results suggest that the returns to education in South Africa are high and convex. Yet, most people appear to drop out of school right before the largest payoff would occur. This behavior would be difficult to rationalize from the perspective of rational, well-informed agents that are not liquidity constrained. In our model we show that a large part of the difference in wages is attributed to ability differences. The actual gains to schooling are much lower than descriptive models suggest. Ignoring ability bias would lead one to overestimate the returns to education.

ii. Why Structural Modelling?

The Mincerian earnings model compares the wages across workers of different ages and levels of schooling. The model does well in capturing the underlying correlation between the covariates, but many applications of this model overlook two important caveats: that the model was meant to be descriptive and that it assumes full employment. The Mincer model may be less applicable to developing countries, like South Africa, than to the developed countries for which the model was first developed.

I hope to contribute to this literature by introducing a set of models that are more suitable to the labour landscape of a developing country with a high level of non-employment and a large degree of heterogeneity among workers. I develop a structural framework wherein individuals are allowed to be rational, utility-maximising and forward-looking. By adding microeconomic theory to the model I am able to endogenise variables that are usually considered to be determined outside the model (exogenous). Over the span of the three chapters, I will be modelling three such choices: the choice of how much education to attain, the choice to actively search for work, and the choice to accept or reject a job offer when one is offered.

Discrete choice structural models build on the static latent variable choice models of the 1960s by introducing an inter-temporal framework into the decision making process. This is important, since all three decisions I model have both instantaneous and future repercussions that need to be considered. The decisions to enrol in schooling and search for jobs require individuals to be dynamic in their considerations, since they are required to weight up current costs against delayed future benefits.

Keane (2010) criticizes non-structural studies for being agnostic about the role of unobservable characteristics. In doing so, they ignore the most interesting aspect of the data – since they are unable to tell us why seemingly similar people make different choices. In all three chapters I will introduce unobserved individual heterogeneity into the models through a finite mixture approach, which allows individuals to differ on unobservable traits.

Theoretical models rarely simplify to linear relationships that are recoverable through conventional ordinary least squares regression and other pre-packaged estimators. This is true for all three my models. In all three the chapters, the estimates that I am interested in are recovered through customised maximum likelihood rather than simpler, linear techniques. A maximum likelihood model was used in the first chapter, since it allowed us to fit the transition estimates on additional moments and not just the naive 1-period transition rates. The maximum likelihood approach enables us to explicitly test for the presence of misclassification and heterogeneity. In the second and third chapters, I model the decision making process faced by forward-looking utility-maximisers. The complexities involved in modelling the decisions of a forward-looking individual under uncertainty again requires making use of non-linear estimation techniques. In both cases the models are able to recover the set of parameters that were most likely to have produced the data that is observed, given the behavioural assumption about how optimizing individuals should behave.

iii. Why the Focus on Black Males?

The dissertation will predominantly focus on the labour market outcomes of black males. The black population was chosen since it is the largest and historically most disadvantaged population group, and hence the most interesting and important from a public policy perspective. The focus on males is convenient since it does not require one to model the simultaneous fertility and labour supply decisions. Lastly, given the very different problems and opportunities experienced by members of the different race groups, it may be over-ambitious to attempt to find an explanation for the behaviour and labour market outcomes for members of all the population groups. For this reason I decided to restrict the analysis a single race and gender.

iv. States, Decisions and Exogenous Transitions

The following three figures give a schematic representation of the underlying states that I am modelling in each chapter. Solid lines indicate exogenously determined transition probabilities, while the dashed lines indicate endogenous decisions.

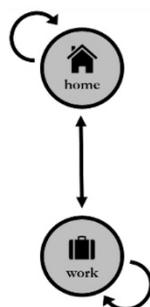


Figure 1:

In the first chapter I model the transition between employment and non-employment.

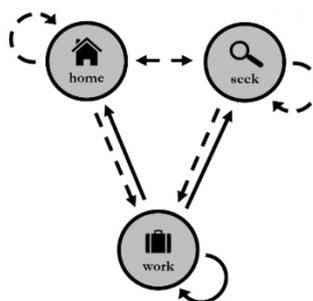


Figure 2:

In the second chapter I model the transition between home production, active search and employment.

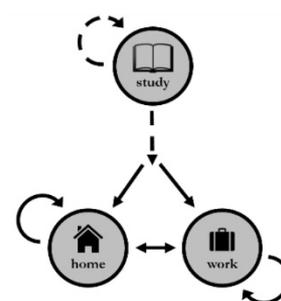


Figure 3:

In the third chapter I model the sequential schooling decision faced by students.

In chapter one I model the transition between employment and non-employment. The chapter is descriptive in nature and void of any causal interpretation. Unlike the descriptive models that merely look at the transitioning in and out of employment between two successive periods, I here adopt a more comprehensive model that allows me to fit on a richer set of empirical moments. Most importantly the model attempts to reconcile the high short term transition rates with the low absorption rates among new labour market entrants. The added structure that the additional data and assumptions imposes on the model allow me to test for the presence of misclassification and unobserved heterogeneity in current transition estimates.

In chapter two I expand the model that was introduced in chapter one, by introducing an active search decision and by adding microeconomic theory to the decision making process. In this chapter, the transition out of employment is still assumed to be exogenously determined, but the employment decision (conditional on being offered a job) and the job-search decision are assumed to be made by the individual. By modelling the process I hope to be able to uncover the extent of which the slow absorption rates among the young is actually a choice – whether unemployment is driven by low job arrival rates or low acceptance rates. Some of the recent literature argues that it is the latter. The methods used in the South African literature, however, rely on self-reported reservation wages that are prone to over-optimism. Job search theory provides an alternative. Rather than asking individuals what they think their reservation wages are, the chapter attempts to uncover reservation wages that are consistent with labour market behaviour.

In chapter three the focus shifts to the pre-labour market schooling decision, allowing individuals to use their knowledge of the labour market and the expected future payoffs of education to pick the level of education that they believe will maximise their expected discounted utility. Here, the educational attainment process is modelled as a sequential schooling decision, allowing individuals to use the information they gained about themselves at the end of each year to influence their decisions to remain in school or not. In modelling the schooling process I hope to develop a structural model that is capable of addressing the ability bias that confounds conventional wage regressions. Theory suggests that the returns will be overestimated if education is possibly correlated to motivation, family background or ability. Very few empirical models have, however, been able to recover lower education estimates. I develop a structural model that explicitly models the sequential educational attainment faced by young black males, whom I allow to differ in unobservable ways.

Chapter 1:

Employment Mobility in South Africa: A Reconsideration

The unemployment rate among the young is high in South Africa. In this chapter we attempt to uncover whether the high level of unemployment is a temporary state experienced by many individuals or a relatively permanent state experienced by only a few. Most of the current literature on South African labour market mobility suggests that it is the former – that the South African labour market is mobile.

We use panel data to investigate how individuals transition in and out of employment over time. Unlike previous studies we do not limit our analyses to simply comparing employment states over two successive periods. Instead, a more comprehensive maximum likelihood model that is able to also fit the age-employment profile and the multi-wave transition rates is developed. The added structure enables us to explicitly incorporate unobservable individual heterogeneity and measurement error into our mobility estimates.

1. Introduction

High unemployment rates among the young, as is experienced in South Africa, can result from either a slow inflow into employment or a rapid inflow which is offset by an equally rapid outflow. These two labour markets could produce the same aggregate level of non-employment, but would necessitate different policy interventions. In this paper we will argue that classic estimates overestimate the true level of mobility in the market. Most black males face a much lower entry and exit rate than was previously believed. According to our results, employment mobility may be up to 5 times lower than what conventional descriptive estimates would suggest.

In order to model the transitions between employment and non-employment we will be using the panel dimension of the Labour Force Survey (LFS) data that follow individuals over 6 waves. Unlike previous authors we will not be limiting our analysis to the transitions between successive periods. Instead, we develop a more comprehensive maximum likelihood model that allows us to exploit the additional waves

in our data to get a richer set of moments to fit our estimates on. The added structure and richer set of moments, have the added bonus of allowing us to incorporate unobservable individual heterogeneity and measurement error into our model.

In the following section we review the South African literature on employment transition. We also draw on international literature to show how different authors have dealt with the validity of the Markov assumption, the presence of individual level heterogeneity as well as the misclassification of employment. The data is introduced in section 3. In section 4 we look at some descriptives, comparing the mobility in employment of white and black males across some covariates. In section 5 we formally estimate the classic transition estimates using the conventional approach that is commonly adopted in the literature. In section 6 we develop and implement our own model that is capable of addressing misclassification and unobserved heterogeneity. Section 7 concludes.

2. Previous Work

Longitudinal labour data has only recently become available in South Africa. The first such dataset was the KwaZulu-Natal Income Dynamics Study (KIDS). Cichello et al. (2005) and Dinkelman (2004) use the first two waves of the panel to study the transition between labour market status in 1993 and 1998 among the black working-aged. While the data is not nationally representative it does provide researchers with an interesting first glimpse into the level of mobility that could be expected nationally. Similarly, Lam et al. (2008) use the Cape Area Panel Study to track the transition from education to employment among a group of young school-leavers in the Cape Town metropolitan area. As with KIDS, this data is not nationally representative. The first nationally representative longitudinal labour data was that of the Labour Force Survey (LFS), which has a rotating panel dimension between 2001 and 2004. Both Banerjee et al. (2006) and Ranchhod and Dinkelman (2008) look at the level of mobility within the data. More recently, Cichello et al (2014) make use of the National Income Dynamics Survey (NIDS).

2.1. Transitioning out of Employment

Dinkelman (2008) looks at the transition probabilities for a group of black males between the ages of 16 and 64. According to her study, the likelihood of still being employed in 1998 if a person was employed in 1993 was 71%. More recently, Cichello et al. (2014) predict that the probability of remaining employed over a two year period for males is 77.5%.

Most of the South African literature we have surveyed distinguishes between the formally and informally employed. Generally the formally employed are less likely to transition out of employment than the informally employed. Cichello et al. (2005) show that the formally employed had a 76% probability of remaining employed, while the informally employed had a 59% probability of still being employed in 5 years' time. Using the LFS data, Banerjee et al. (2006) find that the probability of remaining employed over

a 6 month period for black males is 88% if they were formally employed in the initial period and 63% if they were informally employed in the initial period. According to Ranchod and Dinkelman (2008) the probability of remaining employed over a six month period for a black male is 89% if he is formally employed and 68% if he is informally employed. The probability to remain employed is lower among the younger cohort of black males. Among the younger group the probability of remaining employed is 62% if formally employed and 35% if informally employed.

2.2. Transitioning into Employment

In most of the South African transition literature the non-employed are disaggregated into different groups. Of the five papers reviewed, Cichello et al. (2005) were the only to treat the non-employed as a single group. All four of the other papers distinguish between the unemployed and the not economically active; three of these papers further distinguish between the unemployed who were actively seeking and the unemployed who were discouraged.

Using the initial KIDS study, Cichello et al. (2005) find that the non-employed have a 40% probability of moving into employment. Using the same data Dinkelman (2004) shows that the probability of transitioning into employment differs slightly depending on whether an individual was classified as unemployed or not economically active. The unemployed have a 44% probability of transitioning into employment in 5 years' time. Individuals who were not-economically active have a 29% probability of becoming employed.

According to the results of Banerjee et al. (2006) there is a small 'advantage' to searching. Among the unemployed those who were actively seeking employment have a 17% probability of being employed 6 months later, while those who were not searching only have a 14% probability of being employed 6 months later. Only 5% of black males who were not economically active transitioned into employment over the same period. Although they used the same data, the results of Ranchod and Dinkelman (2008) differ slightly from those of Banerjee et al. (2006)¹. Ranchod and Dinkelman (2008) find that black males who were actively seeking employment are more likely to transition into employment than those who do not actively seek employment. The following job entry probabilities are obtained: 25% for the actively seeking unemployed, 20% for the discouraged unemployed, and 14% for the not economically active. There appears to be hardly any advantage to searching among the young. The respective entry rates among the youth are as follows: 12% if an individual was actively seeking, 11% if an individual was discouraged, and 3% if an individual was not economically active.

Most recently, Cichello et al. (2014) show that individuals who are discouraged are more likely to transition into employment than individuals who are actively seeking employment. Using the first two waves of NIDS

¹ The results differ for two reasons. Firstly, Ranchod and Dinkelman (2008) pooled all the transition periods together and looked at the average 6 month transition rates across all 6 periods, rather than merely looking at one transition period. Secondly, while Banerjee et al. (2006) sample all the working aged black males collectively, Ranchod and Dinkelman (2008) looked at the youth and adults separately.

they show that the probability of becoming employed in two years' time is 39% if an individual is actively seeking employment, 44% if an individual is discouraged, and 29% if an individual is not economically active. It is unclear whether these differences are statistically significant. Simple back-of-the-envelope calculations show that the average probability of transitioning from non-employment (the group as a whole) to employment is roughly 35%.

Lam et al. (2008) use the Cape Area Panel Study to track the transition from education to employment (or non-employment) among a group of young school-leavers in the Cape Town metropolitan area. They find a large racial difference in the likelihood of transitioning into employment. The authors' show 35% of coloured men were employed within one month from leaving school and 50% were employed after six months. Comparatively, only about 10% of black males reported being employed after one month, and only 25% of black males reported being employed after a year.

2.3. Validity of Markov Assumption

To our knowledge, Porteba and Summer (1986) and Magnac and Robin (1994) are the authors that have attempted to test whether the classic transition estimates conform to other moments in the data. Using US data, Porteba and Summer (1986) apply iterative proportion fitting or “raking” to adjust transition estimates to conform to the aggregate level estimates. Their unraked and raked results differ greatly, raising doubts over the validity of these assumptions, most notably the assumption that individuals are homogenous and that employment is correctly captured. Magnac and Robin (1994) follow a similar approach. They argue that if the Markov assumption is valid, then classic transition parameters should produce employment and non-employment spells that should be traceable over time. Unlike us they tested the validity of the Markov assumption using tenure data from the French Labour Force Survey rather than using repeated waves. Like us they, however, found that the Markov property fails in the absence of any misclassification or unobserved individual heterogeneity.

2.4. Misclassification

Employment status is reported with error. According to Meyer (1988) the misclassification in employment is much more detrimental to employment transition estimates than it is for employment estimates, since most of the misclassification cancels out in cross-sectional employment estimates, whereas in panel estimates the errors gets exacerbated. Misclassification errors will bias the employment transition estimates upward (Singh & Rao, 1995).

The most common approach in dealing with the bias that can arise out of misclassification is to make use of an auxiliary source of information that allows one to approximate the misclassification rate. In studies by Chua and Fuller (1987), Poterba and Summer (1986) and Magnac and Visser (1999) the auxiliary source came in the form of a “re-interview” that was conducted shortly after the actual survey. In the re-interview respondents were asked about their employment status a week before – when the actual survey was

conducted. If the answers conflicted with previous responses, respondents were triggered further until the researchers were confident that they had uncovered what their true employment status was a week before. Chau and Fuller (1987) look at the level of misclassification between three categories: the employed, unemployed and not in the labour force. They found that in the US the employed had a 2% probability of being misclassified non-employed. The unemployed had a 6% chance of wrongly being categorized as employed while the not economically active only had a 2% chance of being classified as employed.

Porteba and Summers (1986) use information from the Current Population Survey between 1977 and 1982 as well as the re-interview survey which followed the survey to calculate the degree of misclassification in the US. The authors discover monthly employment mobility rates that are far lower than conventional estimates would suggest. When the authors correct for reporting errors they find that the probability to leave employment drops from 5.0% to 1.9%, while the probability to transition into employment drops from 7.0% to 2.6%.

Biemer and Bushery (2000) also use the Current Population Survey to estimate the extent of misclassification. Using the three waves between 1993 and 1996 as well as the re-interview data, the authors compare the misclassification rates from using the three waves alone (without the re-interview data) to the estimated misclassification estimates from the more classical method that uses the re-interview data. They find that the results they derive using the two different methods are comparable. The probability of being misclassified if employed ranges between 0.4% and 1.3%. Interestingly, the probability of being misclassified as employed differs greatly depending on whether individuals are considered to be not economically active or unemployed. Among the not-economically active the probability to be incorrectly classified as employed ranges between 1.1% and 2.6%, while the probability to be incorrectly classified as employed among the unemployed ranges between 4.6% and 11%.

Using the Canadian Labour Force Survey of the 1989, Sing and Roa (1995) show that among the subsample of individuals that were re-interviewed the probability to have been wrongly classified as employed if non-employed was only 0.8%, while the probability to be wrongly classified as non-employed if employed was 0.9%.

2.4. Heterogeneity

Following a sample of inner city black youths in the United States, Ballen and Freeman (1986) find a high degree of state-dependence. Those individuals that have worked before and worked more recently were more likely to transition back into employment. Clark and Summers (1979) find that the bulk of unemployment is due to an important minority of workers who are unable to find and hold onto steady employment. These few individuals experience prolonged periods of non-employment. These findings lead the authors to question the validity of the Markov condition that stipulates that everyone has the same probability of transitioning into and out of employment.

3. The Data

The Labour Force Survey was collected bi-annually by Statistics South Africa between 2000 and 2007. In this paper we will be exploiting the panel portion of the survey to look at the transitions in and out of employment among black males. Although most of the individuals were matched using administrative records, Statistics South Africa (2006) also matched some of the individuals within household using information on age and gender. The panel runs for 6 waves, spanning from September 2001 to March 2004. A rotating panel was used with 20% of respondents being rotated out of the panel at each wave.

According to Statistics South Africa (2006), some of the variables were cleaned to make them consistent between the waves. The data appeared reliable, although we found some discrepancies between the reported tenure and time out of work values, not only with regard to other tenure and time out of work values, but also with the lagged employment values. Since our model relies on tenure and time out of work to trace out the employment history of individuals a cleaning algorithm was designed and implemented to ensure the accuracy of these variables².

For most of this paper we will be limiting our analysis to working-aged black males. The size of this subsample of black males ranges between 13,602 and 15,490. The subsample of working-aged white males that will act as a comparison group during our initial analysis is much smaller, ranging between 1,643 and 1,855. Table 1.A.1 and table 1.A.2 in the appendix 1.A show the exact sample sizes for each of these groups. The table also shows the level of attrition between waves.

The level of attrition is high and unlikely to be random. In the LFS no attempt was made to track individuals who moved. This is likely to bias our sample if individuals who moved in or out of employment are more likely to also relocate. Tentative evidence from Ranchod and Dinkelman (2008) supports this claim. The authors found that of all the groups, those individuals who were not-economically active were most likely to be matched over time in the LFS panel. The level of employment mobility among the subsample of individuals in our panel is therefore likely to be lower than the actual level of employment mobility of the larger population.

Fortunately, a large proportion of those individuals who were initially lost to follow-up are observed in later waves. Unlike Ranchod and Dinkelman (2008) and Banerjee et al. (2006) we do not restrict our analysis to those individuals that were observed in consecutive waves. In our analysis we retain individuals who were only observed once and individuals that dropped out of the sample but returned later. The structural method we employ uses the information of individuals that are observed only once to help fit derive the

² See Appendix 1.B for a complete discussion on how these variables were cleaned.

aggregate employment-age profile. Similarly, our model is able to extract longitudinal information from those individuals that leave the sample but return later, since we also fit on multiple period transition rates.

4. Descriptive analysis

4.1. Patterns of Employment

The age-employment profile is well understood and frequently reported. It shows how the probability of employment varies over the current pool of working-aged individuals. The following figure was derived using the larger pooled LFS dataset. It shows how the probability of employment varies by age for white and black males.



Figure 1.1: Probability of Employment among working-aged males.

The level of employment to which the two race groups converge differs greatly. For white males the stable level of employment is around 95%, while for black males it is around 70%. The speed of convergence also differs between the two race groups. White males reach their ‘steady state’ of employment much earlier than black males. The level of employment for white males plateaus around the age of 25, while the employment level for black males only plateaus around the age of 35.

This slow absorption into the labour market is the main driving force behind the high incidence of youth unemployment among black males. In appendix 1.F we show that the probability of being enrolled in education conditional on age is almost identical for black and white males. The slow absorption rates among black males, therefore, cannot be due to different drop-out rates.

4.2. Entry and Exit Rates

Entry and exit rates vary between individuals. In this section we show how the reported average transition rates differ by age, employment history and education.

4.2.1. Variation in Entry and Exit Rates by Age

First we compare how the probability to enter and exit employment over a six month period differ by age.

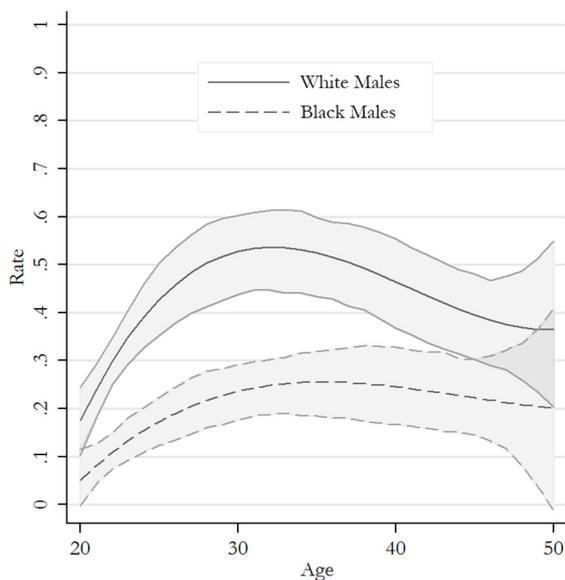


Figure 1.2: Job Entry by Age and Race

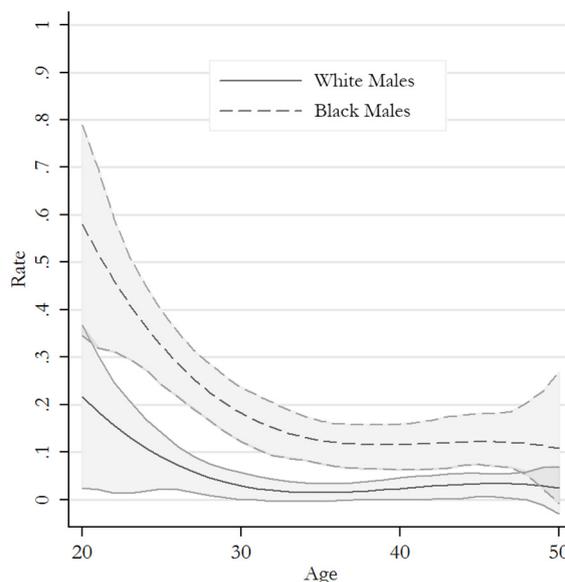


Figure 1.3: Job Exit by Age and Race

On average, white males are roughly two times more likely to transition into employment than black males. From figure 1.2 we can see that at age 20 white males have a 20% probability of becoming employed and black males have a 5% probability of becoming employed. For both groups the likelihood of obtaining work increases between the ages of 20 and 30. By age 30 white males have a 50% probability of moving into employment within the next 6 months. At the same age black males only have a 25% probability of transitioning into employment. From age 30 onwards we see a drop in the relative probability of becoming employed for white males, while for black males the probability of becoming employed remains stable³.

Figure 1.3 shows how the job exit rates vary over one's working age. The probability of exiting employment is highest among the youth. At age 20, employed black males have a 60% probability of becoming non-employed over the next 6 months, while white males have a 20% probability of exiting employment. The exit rates drop as workers become older. At age 30, black males have a 15% probability of becoming non-employed. At the same age white males only have a 3% probability of transitioning out of employment. From age 30 onwards these probabilities remain fairly constant.

Using longitudinal employee-employer data for over a million individuals in the US, Topel and Ward (1992) find that for the first ten years of labour force participation, job mobility was higher than the following

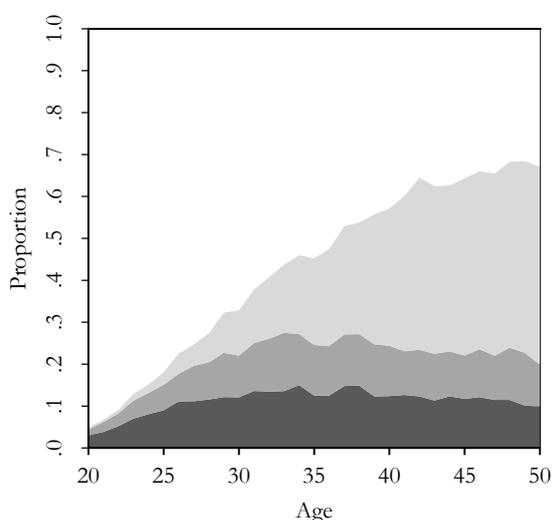
³The confidence interval is largest where the sample is smallest. For the job entry rates the confidence intervals around the conditional point estimates widens at higher ages where the proportion of the sample who is non-employed is smallest, while for the job exit model the confidence intervals are widest at the younger ages, where only a small portion of the sample is employed.

phase where employment appeared to be relatively stable. These results are consistent with a heterogeneous model where employees keep searching for work until they find an appropriate match. Ballen and Freeman (1986) argue that at early ages workers are still searching for employment that fits their talents and preferences; thereafter individuals hold down jobs for longer and mobility decreases. Since finding the correct match may take a while, one would expect the proportion of individuals who are in a “good match” to increase with age.

4.2.2. Variation in Entry and Exit Rates by Recent Employment History

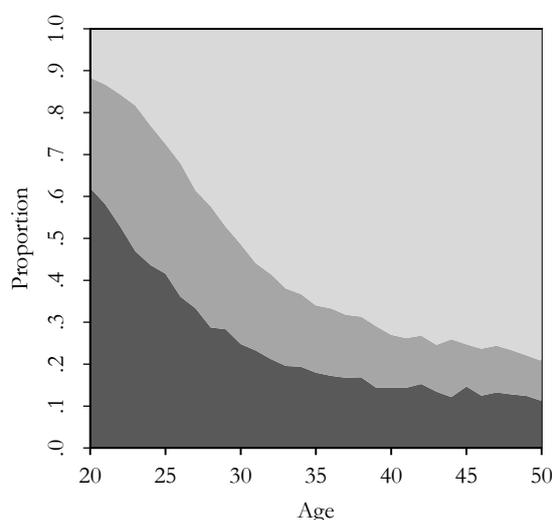
In the data we have some information on how long people have been in their current position. The employed were asked how long they have worked for their present employer while the non-employed were asked how long it has been since they were last employed. Throughout we will refer to the time spent with current employer as ‘tenure’ and the time since last employed as the ‘time gap’. Using this information the employed and non-employed are decomposed into three groups. The employed are split into those who have been employed for less than a year, those who have been employed in a specific position for between 1 and 3 years, and those who have been employed for more than 3 years. Analogously, the non-employed are split into those who have been non-employed for less than a year, those who have been out of work for between 1 and 3 years, and those who have not been employed during the last 3 years⁴.

The following two graphs show how the distribution of the time out of work among the non-employed and the distribution of tenure among the employed vary over a group of working-aged black males⁵.



Never worked before
 More than 3 years
 Between 1 and 3 years
 Less than 1 year

Figure 1.4: Time Out of Work by Age



More than 3 years
 Between 1 and 3 years
 Less than 1 year

Figure 1.5: Tenure by Age

⁴ In the graph below we distinguish between those who have not worked in the last 3 years and those who have never worked. Throughout the rest of the paper we group these two responses together.

⁵ Both graphs were drawn using a larger pooled LFS that runs from 2000 to 2007.

Figure 1.4 shows how the time out of work varies for non-employed black males over their working age. 95% of 20 year olds in our sample who are not employed have never worked before. The proportion of the non-employed who have never been employed decreases to 30% at age 50. This equates to roughly 10% of all 50 year-old black males, given that roughly two thirds of the cohort are employed at that age.

Figure 1.5 shows that the average level of tenure increases with age. Most of the difference in the distribution of tenure is between the ages of 20 and 30. The difference in tenure thereafter is less dramatic. A disproportionate number of the youth have low tenure. The probability of having tenure of less than one year drops from 60% at age 20 to 20% at age 30 and finally to around 10% at age 50.

Next the predicted transition rates are compared across the ‘time gap’ and ‘tenure’ states.

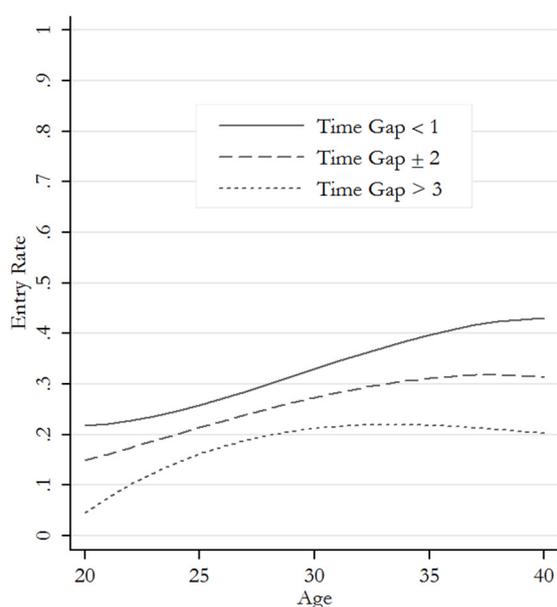


Figure 1.6: Entry Rate by Time Gap

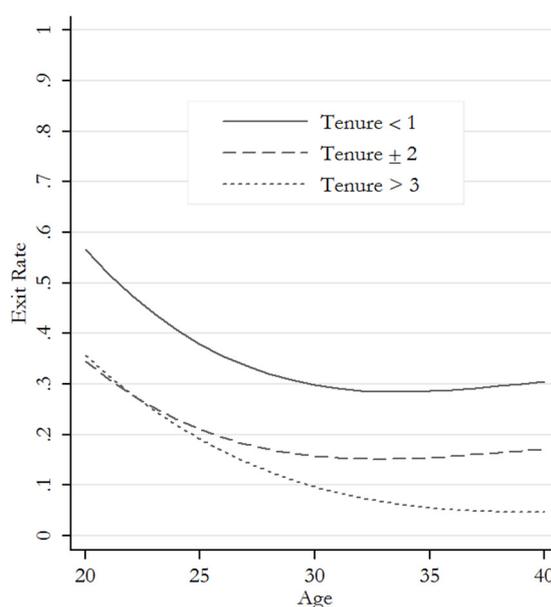


Figure 1.7: Exit Rate by Tenure

Figure 1.6 shows that those individuals who are furthest removed from the labour market (those who have either never worked or worked more than 3 years ago) are least likely to transition into employment. This holds regardless of age. Conversely, those individuals who are closest to the labour market (who have been out of employment for less than a year) are most likely to transition back into employment. Predictably, the transition rates for those who have been non-employed for between one year and three years lie wedged between the two other groups.

In the US, Ballen and Freeman (1986) find that a poor work history lessens the likelihood that one will find employment in the near future, while a good work history increases the probability that one will find future employment. This path-dependency of employment may help explain how some individuals get stuck in “vicious” non-employment cycles.

Figure 1.7 depicts the relationship between tenure and the probability of transitioning out of employment over age. Those individuals who have been in a job for less than a year are most likely to transition out of employment, while those who have been in their job for more than 3 years are least likely to transition out of employment. The probability of moving out of employment decreases with age, but the effect of having tenure remains positive throughout. Again, the transition rates for those who have been employed for between a year and three years lie wedged between the two other groups. The curves are far flatter than before (see figure 3). A large portion of the sharp initial decline in exit rate is driven by low tenure rather than age.

It is unclear through which channel tenure and time gap are affecting transition rates. One common interpretation regarding tenure is that it acts as a proxy for the job-specific experience that someone might have accrued in their current position. If so, workers who have been in a specific position for a while would be more costly to let go of. Conversely, the negative correlation between tenure and the exit rates could be driven by heterogeneity in employee-employer matching. According to Topel and Ward (1992) neither of these theories is likely to be completely accurate.

4.2.3. Variation in Entry and Exit Rates by Education Level

Education plays a central role in early labour market outcomes. Among the youth, that Lam et al. (2008), followed, those who had finished matric were 16 percentage points more likely to transition into employment during the first 4 years after school than those students who left school with less than grade 10. The effect drops off slightly to 10 percentage points when they control for the effect of ability by adding an earlier test-score to their regression.

The following two figures compare the likelihood of job entry and exit by age for individuals with less than matric, matric and more than matric using the constructed Labour Force Survey panel of before.

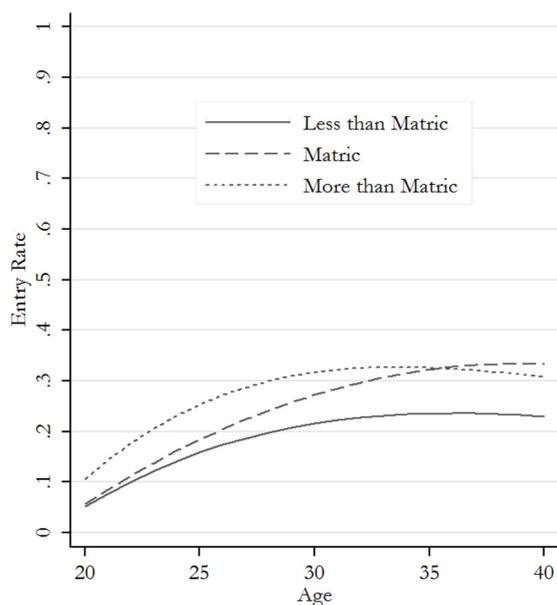


Figure 1.8: Entry Rate by Education Level

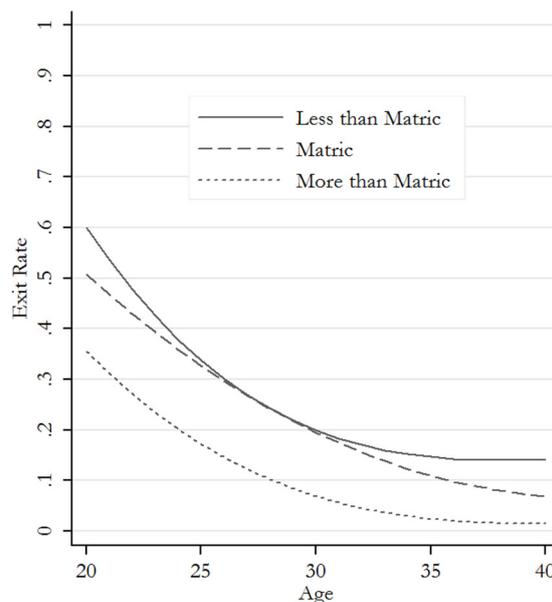


Figure 1.9: Exit Rate by Education Level

Generally speaking, workers with matric are more likely to transition into employment and less likely to transition out of employment. However, the difference in exit rates between these two groups is negligible at some ages. Having a post matric qualification increases the probability of moving into employment and decreases the probability of moving out of employment.

5. Classic Transition Estimates

In the following section we analyse the transition between employment and non-employment.

5.1. Justification for Only Having Two Groups

In our analysis we only distinguish between the employed and non-employed. This simple distinction is common in the international literature, but at odds with most of the previous work on South African labour mobility. Most of the South African literature disaggregates the non-employed into the unemployed who are searching, the unemployed who are discouraged and the not economically active.

While the distinction between the searching unemployed, discouraged and not economically active is clearly defined in the theoretical literature, the distinction is less conclusive in practice. Clark and Summers (1979) find that there is no meaningful distinction between being unemployed and not economically active. Most of the unemployment spells, in their data, end when individuals transitioned out of the labour force rather than back into employment. Many of these individuals who transitioned from unemployed to not economically active, re-emerged as unemployed within a month's time again.

Cichello et al. (2014) find that the differences in the job entry rates among the three different categories were not that sizeable: 22% for the not economically active, 28% for the discouraged and 32% for the actively searching. The level of churning between these 3 groups between the two periods was also high. Banerjee et al (2006) find that only about a third of those who were classified as discouraged were still classified as discouraged 6 months later. Similarly, roughly 25% of the not economically active remained in that category. Most moved to either discouraged or seeking.

Lastly, having more groups makes the model even more susceptible to measurement error, especially if the distinction between the groups is subjective. Chua & Fuller (1987) find that roughly 10% of unemployed misclassified themselves as being not economically active. Using more recent data, Biemer and Bushery (2000) estimate that the proportion of the unemployed who erroneously report to be not economically active ranges between 11% and 28%.

5.2. Classic Estimates

A transition matrix is constructed for black and white males. The matrices show how many individuals transitioned in and out of employment between two waves.

Table 1.1: Transition Matrix for Black and White Males

From	Black Males		From	White Males	
	To			To	
	Non-Empl	Empl		Non-Empl	Empl
Non-Empl	82.6%	17.4%	Non-Empl	64.4%	35.6%
Empl	14.1%	85.9%	Empl	3.7%	96.3%

The elements on the diagonal denote the probability of staying in the same employment state, while the elements off the diagonal denote the probability of transitioning in or out of employment. For both tables the transition probabilities were averaged over individuals and across time⁶.

Black males face an average six month job entry rate of 17.4% and an average six month job exit rate of 14.1%. White males, on the other hand, have a 35.6% probability of transitioning into employment and only a 3.7% probability of transitioning out of employment once employed. These results suggest that white males are not only quicker to be absorbed into the labour market, but are far less likely to transition out of employment once they find work.

The descriptive results that we discussed in the previous section suggest that the transition probabilities are not uniform. The entry and exit rates appear to differ among individuals and even among the same individuals over time. The easiest way to incorporate the observable heterogeneity would be through a linear probability model (LPM). In section 1.C of the appendix we show how linear probability models can be used to allow the transition probabilities to vary by age, working history and education.⁷

⁶ Issues regarding heterogeneity and misclassification were ignored.

⁷ Two LPM regressions were run: a job entry and a job exit model.

5.3. Concerns with Classic Estimates

If we are willing to assume that the matrices are Markovian (that the probability of switching states depends on an individual's current status and nothing else) then we are able to estimate the steady state of employment to which the population should converge.

$$Empl_{SS} = \frac{\phi_{ne}}{\phi_{ne} + \phi_{en}} \quad \text{where } \phi_{ne} \text{ denotes the job entry rate and } \phi_{en} \text{ denotes the job exit rate}$$

The predicted steady state is derived by substituting the transition probabilities from table 1 into the above equation. The average spell lengths is also reported⁸.

Table 1.2: Entry Rate, Exit Rate and Steady State of Employment by Race

	Black Males	White Males
Entry Rate	17.4%	35.6%
Exit Rate	14.1%	3.7%
Implied Steady State Employment Rate	55.2%	90.6%
Implied Average Employment Spell	3.5 years	12.7 years
Implied Average Non-Employment Spell	2.9 years	1.4 years

If these entry and exit rates are correct, black males will converge to an employment rate of roughly 55%. This is far lower than the steady state employment we observed from the cross-sectional data in figure 1.1, where the steady state was estimated at almost 70%. In table 1.A.5 (in the appendix) we show that the entry rate, the exit rate and the implied steady state of employment differ by age cohort.

The figure below shows the rate of convergence that is implied by the classic transition model and the actual rates of employment over different ages for black males⁹.

⁸ The transition probability estimates suggest that the average non-employment spell will last for roughly 2.9 years for black males and roughly 1.4 years for white males. Conversely, the average employment spell for a black male is 3.5 years, while the average employment spell for a white male is around 13 years.

⁹ The initial state assumption which determines the starting point of the curve is discussed in more depth in the next section and in section 1.F of the appendix.

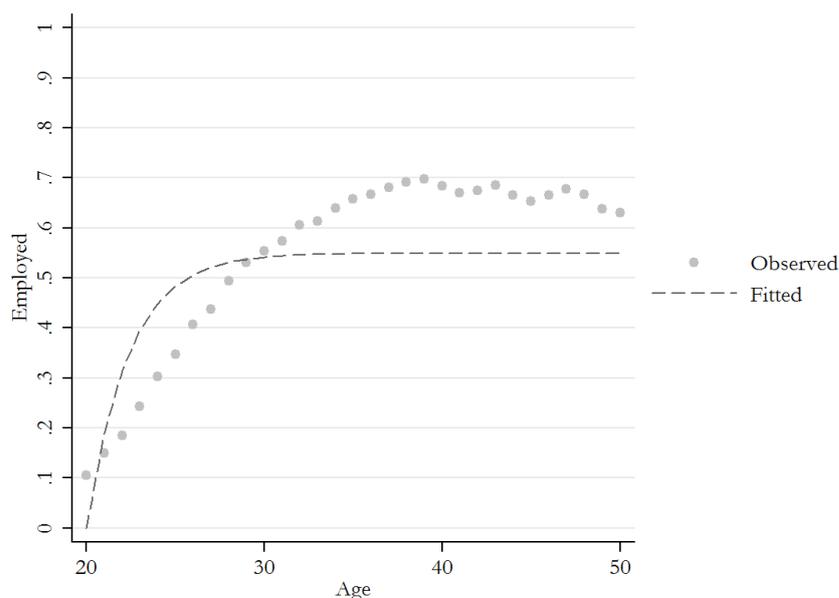


Figure 1.10: The Predicted and Reported Employment Rate by Age

The classic transition estimates not only converge to the wrong employment level, they also converge at the wrong rate. The implied rate of convergence of the classic model is much faster than the actual rate of convergence found in the data. In the appendix we provide further evidence against the validity of the unconditional Markov assumption. While the 6 month transition estimates are fairly comparable over time, the multiple period estimates implied by these 6 month transition periods are far larger than the actual multiple period estimates. In the next section we will test whether we are able to recover more agreeable results when we allow the entry and exit rates to vary by correlates.

To our knowledge, Porteba and Summer (1986) and Magnac and Robin (1994) are the only other authors that have attempted to test whether the classic transition estimates conform to other moments in the data. Both articles find that their observed 1-period transition rates are inconsistent with other moments in the data. Evidence that the Markov property fails to hold in the absence of any misclassification or unobserved individual heterogeneity.

6. Refined Model

In the previous section we showed that the Markov condition fails to hold when we do not allow for either unobserved individual heterogeneity or measurement error. The transition rates we derived from our descriptive model suggest that the economy should converge to the steady state of employment more quickly than it actually does. It also predicts multiple-wave transition rates that are much higher than the ones we observed by simply looking at the transitions rates between two successive periods in our panel.

In response to these discrepancies, we set out to develop a more comprehensive maximum likelihood model. Unlike conventional transition models, that only fit on the naive 1-period transition rates, our model incorporates additional empirical moments (the age-employment profile and multiple-period transition rates) into the likelihood estimate. The refined model allows us to introduce measurement error and heterogeneity into our model.

Before setting up the model, let us briefly introduce the key identifying assumptions that are required to hold for the estimates to be valid¹⁰.

Assumption 1: Stationarity Assumption

The LFS data we are using spans over 6 waves. In order to use the information more efficiently, we assume that the structure of the labour market does not change drastically over those three years. In section D1 in the appendix the transition probabilities across each pair of consecutive waves are compared for a group of black males. The transition rates between each set of waves is comparable suggesting that the stationarity assumption is plausible.

Assumption 2: Markovian Assumption

Under the first-order Markov condition the probability of each outcome at time $t + 1$ is fully described by an individual's outcome at time t . Put simply, the model has no memory. $P(S_{t+1}|S_t, S_{t-1}, \dots, S_1) = P(S_{t+1}|S_t)$. This assumption helps simplify analysis.

In the previous section we showed that the transition estimates that we derive through the conventional technique do not fit all the moments in the data. While this deficiency is sometimes taken as evidence against the Markov assumption, Langeheine (1988) argues that the difference between the implied and observed multiple period estimates could equally well be a symptom of heterogeneity. If a small subsample of individuals face high transition rates, while the rest of the population faces low transition rates it would make the multiple period transition more comparable to the classic 1-period estimates. According to Wolfe et al. (2003) and Biemer and Bushery (2000) the discord is also consistent with measurement error. If

¹⁰ The plausibility of each of these assumptions is discussed in section 1.D, section 1.E and section 1.F of the appendix.

individuals misclassify their employment status it would inflate short term relative to long term transition estimates. In our model we plan to explore both these avenues.

None of the five articles that estimated transition rates on South African labour data make any formal mention of their stance on the Markov assumption.

Assumption 3: Initial State Assumption

Conventional transition estimates ignore aggregate employment levels. In our model, however, we want to fit the data to the age-employment-profile. In order to do so we are required to make an assumption about when individuals enter the labour market for the first time. In section 1.F in the appendix it was shown that the average age was around 20. For the sake of simplicity the assumption was made that everyone enters the labour market at age 20 and that everyone enters the labour market as non-employed.

We start off by assuming that all individuals are identical (homogeneity) and that employment is measured without error (no misclassification), but will relax both these assumption shortly. Under homogeneity all workers face the same probability of job entry and exit. Therefore, the probability of being employed in period t is assumed to only depend on an individual's employment status in period $t - 1$.

The entry and exit rates can be written as follows:

$$\begin{aligned} P(S_t = 1) &= \phi_{ne} & \text{if } S_{t-1} = 0. \\ P(S_t = 0) &= \phi_{en} & \text{if } S_{t-1} = 1. \end{aligned}$$

The same set of probabilities can also be represented with a transition matrix¹¹ where the off diagonal elements represent the entry and exit rates.

$$T = \begin{bmatrix} \phi_{nn} & \phi_{ne} \\ \phi_{en} & \phi_{ee} \end{bmatrix} = \begin{bmatrix} 1 - \phi_{ne} & \phi_{ne} \\ \phi_{en} & 1 - \phi_{en} \end{bmatrix}$$

The employment transition model allows for two states. An individual's state at period t is denoted as $S_t = \begin{pmatrix} N \\ E \end{pmatrix}$, where S_t is a vector that takes the value $(1,0)'$ if an individual is non-employed in period t , and $(0,1)'$ if an individual is employed. Given a specific entry and exit rate, the likelihood of observing any two consecutive employment states for someone of a particular age is

$$\begin{aligned} L_i(\phi_{ne}, \phi_{en}; S_t, S_{t-1}, Age) &= P(S_t, S_{t-1}, Age) \\ &= P(S_t, S_{t-1} | Age) && \text{(Age is exogenously determined)} \\ &= P(S_t | S_{t-1}, Age) \times P(S_{t-1} | Age) && \text{(Markov Property)} \\ &= P(S_t | S_{t-1}) \times P(S_{t-1} | Age) && \text{(Transition rates is constant over all ages)} \end{aligned}$$

The first term $P(S_t | S_{t-1})$ captures the likelihood of being observed in a particular employment state in six months' time conditional on an individual's current employment state. These values can be read off the transition matrix.

¹¹ The transition matrix is conceptually identical to those shown in table 1.1.

$$P(S_t|S_{t-1}) = S_{t-1} \begin{bmatrix} \phi_{nn} & \phi_{ne} \\ \phi_{en} & \phi_{ee} \end{bmatrix} S_t'$$

The second term, $P(S_{t-1}|Age)$, captures the probability of employment for someone who has not been observed before, but for whom we know their age. Assuming everyone enters the labour market at age 20, and that employment evolves in a Markovian manner, we are able to trace out the predicted likelihood of being in employment at each age.

$$P(S_t|Age_t) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left(\begin{bmatrix} \phi_{nn} & \phi_{ne} \\ \phi_{en} & \phi_{ee} \end{bmatrix} \right)^{2(Age-20)}$$

Combining these two terms, the likelihood function for a homogenous model with no measurement error and only two waves can be rewritten as:

$$L_i(\phi_{ne}, \phi_{en}; S_{t_2}, S_{t_1}, Age) = S_{t-1} \begin{bmatrix} \phi_{nn} & \phi_{ne} \\ \phi_{en} & \phi_{ee} \end{bmatrix} S_t' \times \begin{pmatrix} 1 \\ 0 \end{pmatrix} \left(\begin{bmatrix} \phi_{nn} & \phi_{ne} \\ \phi_{en} & \phi_{ee} \end{bmatrix} \right)^{2(Age-20)}$$

The model is expanded to allow for multiple waves and missing observations. Letting N denote the number of waves and letting d capture the time lapse between consecutive observations, the likelihood function becomes:

$$L(\phi_{ne}, \phi_{en}; S_N, \dots, S_1, Age) = \prod P(S_t|S_{t-d}) \times P(S_1|Age)$$

$$\text{where } P(S_t|S_{t-d}) = S_{t-d} \begin{bmatrix} \phi_{nn} & \phi_{ne} \\ \phi_{en} & \phi_{ee} \end{bmatrix}^d S_t' \text{ and } d \text{ is the period between observations.}$$

This likelihood function derived above enables us to recover the transition parameters that were most likely to have produced the moments we observe in the actual data. As long as individuals are homogenous and employment is correctly captured, this would be adequate.

6.1. Constant Entry and Exit Rates

The following table contains the job entry and exit rates that best fit the likelihood function¹².

Table 3: Maximum Likelihood Estimates

(Model 2a)	
Job Entry	
Constant	0.104***
Job Exit	
Constant	0.094***
N	43013
Log-likelihood	-41216
Allow for Misclassification	No
Allow for Observed Heterogeneity	No
Allow for Unobserved Heterogeneity	No

An entry rate of 10.4 % and exit rate of 9.4% were obtained. These estimated transition rates are lower than the estimates we obtained in the previous section (17.4% and 14.1%), where we only fitted on transitional data. This model comes slightly closer to fitting the slow convergence to the steady state observed in the level data than model 1, where we only fitted on transition rates. However, the model does worse in fitting the 1-period transition data, since we now fit on other moments as well.

The current model, which assumes that the entry and exit rates are constant fails to adequately fit all the moments in our data. This should not be surprising, since this model still assumes homogeneity and does not allow for measurement error¹³. In the following two sections we expand the set of parameters used in our model to fit the moments in the data. The results show that the estimates obtained from fitting on subsequent period, fitting on multiple periods and fitting on the employment-age profile become consolable if we allow for misclassification or heterogeneity.

When we allow for misclassification and when we allow for heterogeneity we recover transition rates that are far lower than the observed 1-period transition rates which are commonly reported in the literature.

6.2. Misclassification

According to Meyer (1988) measurement error is particularly severe in transition data. In measuring the aggregate employment rate some of this bias cancels out – some individuals are erroneously classified as employed and some individuals are erroneously classified as non-employed. For flow data this is not the case; the errors tend to be multiplied. Generally, misclassification will inflate transition estimates (Pfeffemann, et al., 1988).

¹² Consult table 1.A.6 in the appendix to see the standard errors.

¹³ All three these models assume constant entry and exit rates. The only difference between the three models is the set of moments to which they are fitted.

There are a host of reasons why employment could be measured with error. Firstly, employment is not a clearly defined dichotomous state. Those on the margin, who are only loosely attached to the labour market, may be unsure as to whether they classify as employed or not.¹⁴ Secondly, questions are not always answered by the respondent themselves, but rather by whoever happened to be home during the survey. If different individuals respond differently, it will inflate the degree of misclassification. Lastly, measurement error can also be introduced through the way the data is captured, by the survey collectors, by the data coders or through the way the panel was merged.

In our model we allow the probability of misclassification to differ by true employment state, but assume that the probability of misclassification is independent of other covariates and time¹⁵. Letting $Empl$ denote the true employment status, and $Empl^*$ denote the reported state of employment, the probability of misclassification can be denoted as follows:¹⁶

$$P(Empl^* = 1 | Empl = 0) = \alpha_0$$

$$P(Empl^* = 0 | Empl = 1) = \alpha_1$$

The relationship between the reported and actual levels of employment can also be represented in matrix notation, where the misclassification model describes a projection of the true state and its corresponding probabilities onto the reported state and its corresponding probabilities.

$$P^* = MP \quad \text{where} \quad P = \begin{bmatrix} P_N \\ P_E \end{bmatrix} \text{ denotes the true proportion of employed and non-employed,}$$

$$P^* = \begin{bmatrix} P_N^* \\ P_E^* \end{bmatrix} \text{ denotes the reported proportions of employed and non-employed, and}$$

$$M = \begin{bmatrix} 1 - \alpha_0 & \alpha_0 \\ \alpha_1 & 1 - \alpha_1 \end{bmatrix} \text{ denotes the misclassification matrix } M.$$

Depending on one's perception of how the 'noise' that leads to misclassification in employment arises, one may be willing to make further assumptions about the relationship between α_0 and α_1 . Wolfe et al. (2003), for instance, assume that $\alpha_0 = \alpha_1$, in which case the probability of misclassification does not depend on one's actual employment status. Chua and Fuller (1987) show that the only case where the employment estimate will not be biased through misclassification would be if the model projects onto itself ($P = MP$). This will only occur when $P_E \alpha_1 = P_N \alpha_0$.

¹⁴ In the Labour Force Survey respondents were asked whether a specific household member had spent more than an hour during the last week on any of a list of work-related activities. Individuals were classified as employed if they responded 'yes' to any of the questions. While the survey was set up in such a manner as to minimise confusion, the questions remain open to interpretation and sensitive to when the survey was conducted. Theoretically, someone who only works a couple of days a month could in the same month be reported as either employed or non-employed.

¹⁵ In the literature, this independence condition is referred to as the Independent Classification Error (ICE) condition.

¹⁶ Lewbel (2000) also allows the misclassification rates to vary between the two groups.

In our model we are interested in recovering the transition rates rather than the level estimates. The relationship between the reported and true transition flows is denoted as follows:

$$T^* = MTM' \quad \text{where} \quad T = \begin{bmatrix} 1 - \phi_{ne} & \phi_{ne} \\ \phi_{en} & 1 - \phi_{ne} \end{bmatrix} \text{denotes the actual transition matrix,}$$

$$T^* = \begin{bmatrix} 1 - \phi_{ne}^* & \phi_{ne}^* \\ \phi_{en}^* & 1 - \phi_{ne}^* \end{bmatrix} \text{denotes the reported transition matrix, and}$$

$$M = \begin{bmatrix} 1 - \alpha_0 & \alpha_0 \\ \alpha_1 & 1 - \alpha_1 \end{bmatrix} \text{denotes the misclassification matrix } M.$$

According to the sandwich model above, each reported transition rate is a weighted sum of the four true transitions rates. To illustrate the point, let us consider the transition from non-employment to employment. There are 4 possible combinations of actual transitions and corresponding misclassifications that could be consistent with someone being reported as moving from non-employment into employment.

$$\phi_{ne}^* = (1 - \alpha_0)\alpha_0\phi_{nn} + (1 - \alpha_0)(1 - \alpha_1)\phi_{ne} + \alpha_1\alpha_0\phi_{en} + \alpha_1(1 - \alpha_1)\phi_{ee}$$

Intuitively, the probability of being reported as moving into employment should be highest among those who actually transitioned into employment. Conversely, the probability should be lowest among those who moved out of employment, since they would have to be misclassified in both waves.

The earlier equation can be transformed so that:

$$T = (M^{-1})'T^*(M^{-1})$$

We are interested in recovering T . We know T^* , since we are able to observe the reported transition rates directly from the data. In order to recover the true transition, T , we are, however, also required to identify the misclassification matrix M .

The most common approach used for identifying M requires an auxiliary source of information that allows one to find out what the actual misclassification rates are. In the case of Chua and Fuller (1987), Poterba and Summer (1986), and Magnac and Visser (1999), the auxiliary source comes in the form of a “re-interview” that was conducted shortly after the actual survey¹⁷. According to Dinkelman (2004) these methods are of little use to us, since none of the South African panels have any re-interview arms that we can exploit. She concludes that we are therefore unable to do much about the misclassification of employment in labour market transition data in South Africa.

We propose a different strategy. The method we use is comparable to the method adopted by Biemer and Bushery (2000). Rather than extracting information directly from an auxiliary source, we will use a maximum likelihood method that makes use of the repeated waves in the panel to estimate the misclassification rate.

¹⁷ In the re-interview respondents were asked about their employment status a week before – when the original survey was conducted. If their response differed to their previous response respondents were triggered even further until the researchers were confident that they had the true employment status.

Our method relies on the Markov assumption and stationarity assumption¹⁸ to identify the extent of the misclassification in our data.

Under our identifying assumptions the true k-period transition matrix can be rewritten as the sandwich estimate of the corresponding k-period reported transition matrix T_k^* :

$$T_k = (M^{-1})'T_k^*(M^{-1}) \text{ for any value of } k \text{ between } 1 \text{ and } 5.$$

The misclassification matrix M is now entrusted with the role of restoring the Markov assumption – fixing the model so that it fits the naïve 1-period transition rates as well as the multiple period-transition rates and the age-employment profile. We rerun our earlier models, this time allowing for misclassification¹⁹.

Table 1.4: Maximum Likelihood Estimates

	(Model 3a)
Job Entry	
Constant	0.041***
Job Exit	
Constant	0.017***
Misclassification1	0.067***
Misclassification2	0.077***
N	43013
Log-likelihood	-39659
Allow for Misclassification	Yes
Allow for Observed Heterogeneity	No
Allow for Unobserved Heterogeneity	No

The results from table 1.4 suggest that much of the transitioning between waves is due to misclassification rather than actual movements into or out of employment. If we compare model 3a to model 2a we find that once we control for misclassification the probability of transitioning into employment drops from to 10% to 4% and the probability of transitioning out of employment drops from 9% to 2%. Misclassification accounts for roughly two thirds of the variation in employment between waves under this specification. If we are willing to assume that all the deviation from the classic Markov assumption is through the misclassification alone, then our model predicts that the probability of being falsely classified as employed if an individual is non-employed is 7% and the probability of being falsely classified as non-employed if employed is 8%.

These misclassification rates are higher than the misclassification rates that Chau and Fuller (1987) found for the US. According to Chau and Fuller (1987) roughly 3% of the non-employed misclassified themselves as employed and roughly 2% of the employed misclassified themselves as non-employed.

¹⁸ The stationarity assumption (or local independence assumption) requires that the classification errors be independent between the 6 waves. Meyers (1988) and Singh and Rao (1995) find this assumption to be reasonable over short periods. Some preliminary tests on our data (see section 1.D in the appendix) support this claim.

¹⁹ The complete results are contained in the appendix.

6.3. Observed Heterogeneity

In the previous section we showed that the transition rates differ greatly by age, education and employment history. We develop a second less stringent version of the Markov assumption. Under this assumption the transition probabilities are allowed to vary between individuals and between periods. The state space is expanded to capture the progression of tenure and time gap as individuals advance through their respective working lives.

$$\text{The state-space is transformed from } s = \begin{bmatrix} \text{nonemployed} \\ \text{employed} \end{bmatrix} \text{ to } S = \begin{bmatrix} \text{nonemployed, with timegap} \in [3.0, \infty) \\ \text{nonemployed, with timegap} \in [2.5, 3.0) \\ \vdots \\ \text{nonemployed, with timegap} \in [0.0, 0.5) \\ \text{employed, with tenure} \in [0.0, 0.5) \\ \vdots \\ \text{employed, with tenure} \in [2.5, 3.0) \\ \text{employed, with tenure} \in [3.0, \infty) \end{bmatrix}$$

The state space now contains twelve states. Six employed and six non-employed states. Individuals are restricted in how they are able to transition between these different states. From one wave to the next, a non-employed individual can either remain non-employed and have their time gap increase by six months or they can transition into employment and obtain a tenure of between zero and six months. Similarly, employed individuals can either remain employed and have their tenure increase by six months or they can become non-employed with a time gap of between zero and six months. The transition probabilities that accompany the expanded state-space are allowed to vary by age, education and current state.

The likelihood of observing a specific sequence of states for someone of a particular age and education level given a certain set of state parameters²⁰:

$$L(\bar{\phi}_{ne}, \bar{\phi}_{en}; S_N, \dots, S_1, Age, Educ) = \prod P(S_t | S_{t-d}) \times P(S_1 | Age)$$

$$\text{where } P(S_t | S_{t-d}) = \begin{bmatrix} \bar{\phi}_{nn,it} & \bar{\phi}_{ne,it} \\ \bar{\phi}_{en,it} & \bar{\phi}_{ee,it} \end{bmatrix}^d$$

$$\bar{\phi}_{ne,it} = \gamma_{00} + \gamma_{01}Age20_{it} + \gamma_{02}Age30_{it} + \gamma_{03}TimeGap1_{it} + \gamma_{04}TimeGap2_{it} + \gamma_{04}Sec_i + \gamma_{04}Ter_i$$

$$\bar{\phi}_{en,it} = \gamma_{10} + \gamma_{11}Age20_{it} + \gamma_{12}Age30_{it} + \gamma_{13}Tenure1_{it} + \gamma_{14}Tenure2_{it} + \gamma_{15}Sec_i + \gamma_{16}Ter_i$$

$$\bar{\phi}_{nn,it} = 1 - \bar{\phi}_{en,it}$$

$$\bar{\phi}_{ee,it} = 1 - \bar{\phi}_{ne,it}$$

We allow the entry and exit rates to vary by age²¹, education and by recent employment history. The following in table 1.5 are obtained using this method²².

²⁰ $\bar{\phi}_{ne}$ and $\bar{\phi}_{en}$ now denote vectors.

²¹ Two sets of splines were added to the model to capture the non-linearity in age. One spline runs from age 20 to 30, the other from age 30 to 50. These splines were chosen since they are easily interpretable and fit the data well.

²² The complete results are contained in the appendix.

Table 1.5: Maximum Likelihood Estimates

(Model 2a)	
Job Entry	
Constant	0.050***
Age20	0.013***
Age30	0.000
TimeGap < 1	0.130***
TimeGap =± 2	0.078***
Matric	0.013***
More than Matric	0.072***
Job Exit	
Constant	0.234***
Age20	-0.017***
Age30	0.000***
Tenure < 1	0.254***
Tenure =± 2	0.091***
Matric	-0.028***
More than Matric	-0.053***
N	43013
Log-likelihood	-38573
Allow for Misclassification	No
Allow for Observed Heterogeneity	Yes
Allow for Unobserved Heterogeneity	No

In the table above, entry and exit rate are allowed to vary on observable covariates. The results are consistent with descriptive results. Every additional year of age in an individual's twenties increases the probability of finding employment by roughly 1.3 percentage point and reduces the probability of moving out of employment by 1.7 percentage points. Beyond 30, the entry or exit remain fairly stable.

How long ago someone has worked also has a significant effect on finding employment. Compared to someone who has worked within the last three years, someone who has worked within the last year would be 13 percentage points more likely to transition into employment, while someone who has worked between one and three years ago will be 8 percentage points more likely to gain employment.

How long someone has been in their current position also affects transition probabilities. Compared to someone who has been in their position for more than three years, someone who has only been employed for a year or less will be 25 percentage points more likely to transition out of employment, while someone who has been working for between one and three years will be 9 percentage points more likely to exit employment.

Having matric relative to less than matric increases the probability of finding employment by 1 percentage point, while having tertiary education relative to having less than matric increases the probability of finding employment by 7 percentage points. Having matric decreases the likelihood of becoming non-employed by 3 percentage points, while having tertiary education decreases the likelihood of becoming non-employed by 5 percentage points.

6.4. Unobserved Heterogeneity

Individuals may also differ by unobservable traits (such as motivation, family background or ability) that we do not have data on. A finite mixture model (FMM) is constructed, that allows for persistent unobserved individual heterogeneity²³. In our model we allow the job entry and job exit rates to vary between individuals depending on their unobserved latent type.

The likelihood function of before is modified in two ways: First, the transition probabilities are allowed to differ by type.

$$\phi_{ne}^k = \begin{cases} \phi_{ne}^1 & \text{if you are of type 1} \\ \phi_{ne}^2 & \text{if you are of type 2} \end{cases} \quad \text{and} \quad \phi_{en}^k = \begin{cases} \phi_{en}^1 & \text{if an individual is of type 1} \\ \phi_{en}^2 & \text{if an individual is of type 2} \end{cases}$$

The overall likelihood becomes the weighted sum of the likelihoods of the two types of individuals. For the heterogeneous model the likelihood function becomes

$$L_i(\phi_{ne}^1, \phi_{ne}^2, \phi_{en}^1, \phi_{en}^2, \pi_1; S_1, S_2, Age) = \sum_{k=1}^2 \pi_k L_i(\phi_{ne}^k, \phi_{en}^k; S_1, S_2, Age)$$

where each type is assumed to occur with probability π_k .

From the equation above we can see that the set of parameters that needs to be solved has increased. The model now has a second set of parameters that allows for a level shift in the entry and exit rates among the two groups, as well as a third parameter, π_1 , that will denote the proportion of type 1 individuals in the population.²⁴

If employers are able to observe traits like motivation, family background or ability (which are not observed by the econometrician) then we would expect that the group who have more desirable unobservable should be absorbed into employment more quickly and retain employment for longer (have higher entry rates and lower exit rates compared to the other group).

Conversely, one could also imagine a situation where one group is more mobile than the other: where one group regularly moves in or out of employment (have high entry and exit rates), while another group is more likely to remain employed or non-employed (have low entry and exit rates). In this case, it is unlikely that the differences would be driven by productive characteristics (such as ability and motivation). Topel and Ward (1992) distinguished between “movers” and “stayers” when they discuss unobserved person specific traits that may affect mobility.

Table 1.6 contains the estimates for a heterogeneous version of our constant exit and entry rate model²⁵.

²³ Persistence here implies that individuals will be of a specific type for the entire period that they are observed, but not necessarily for their entire life.

²⁴ Everyone who is not of type 1, will be of type 2. ($\pi_2 = 1 - \pi_1$)

²⁵ The standard errors are contained in table 1.A.7 in the appendix. The exhaustive version of this model, where we allow for covariates, is contained in table 1.A.8.

Table 1.6: Maximum Likelihood Estimates

(Model 4a)	
Job Entry	
Constant Type 1	0.038***
Constant Type 2	0.437***
Job Exit	
Constant Type 1	0.011***
Constant Type 2	0.508***
Proportion of Type 1	0.709***
N	43013
Log-likelihood	-39447
Allow for Misclassification	No
Allow for Observed Heterogeneity	No
Allow for Unobserved Heterogeneity	Yes

Two very different type of individuals emerge. Roughly 70% of the sample are of type 1 and roughly 30% are of type 2. Type 1 have a 96% probability of remaining non-employed and a 99% probability of remaining employed. Type 2 individuals, on the other hand, are very mobile. They have a 44% probability of transitioning into employment, and a 51% probability of transitioning out of employment. Their probability of employment therefore is less affected by their previous state – their employment state is close to being memory-less.

The results are consistent with the “movers” and “stayers” concept that was proposed by Topel and Ward (1992), where one group faces higher transition rates than the other. As with our previous estimates, which only allowed for misclassification, most individuals appear to face lower transition rates than are commonly observed in traditional transition models. In Model 3a, where all the deviation from the Markov assumption is driven by misclassification, roughly two thirds of the observed movement in and out of employment was estimated to be due to misclassification. The finite mixture model suggests that most of the variation in employment within the finite mixture model is driven by a small group, consisting of about one third of the population, who continuously transition in and out of employment.

Earlier we showed that each individual has a 71% prior probability of being of type 1 and a 29% prior of being of type 2. These prior probability denote the average probabilities over the entire sample – the unconditional probability of being of each type before look at any additional information. This probability of being of each type is however not constant over all observations.

A more accurate probability of type can be obtained if we use the information we have data on. To posterior probability of being of each type for each individual we have to compare the likelihood value for each individual to be of each type. The posterior probability compares the likelihood of a being of type 1 and type 2 given a set of covariates we label $x_{i1} \dots, x_{iT}$.

$$\text{Posterior class probability: } P(\psi = j | x_{i1} \dots, x_{iT}) = \frac{\psi_j \times f(y | \psi = j)}{\sum_q \psi_q \times f(y | \psi = q)}$$

Once we have obtained the posterior probability of being of each type, we are able to compare these probabilities across observed covariates in an attempt to uncover whether type is determined by supply side conditions, heterogeneity or by measurement error. The results are reported in table A8 in the appendix.

Interestingly the probability to be of each type does not change drastically by age or by having matric. This is somewhat surprising since the results earlier showed that the transition rates did vary by these covariates. We also compare the likelihood of being of each type across children, household-head and marriage (a set of household variables that are commonly believed to be correlated to labour supply), while all three variables were significant at a 10% level, none of the variables held any economic significance. Membership does not appear to be driven by labour supply.

How loosely people are tied to the labour is closely related to their type.²⁶ Individuals who earn less than the median wage are 8% and more likely to be in the second group, while individuals who are formally employed are 4% more likely to belong to the first group. Comparatively, individuals who report to be seeking employment when non-employed are 6% more likely to be of second group. Unfortunately, the level of misclassification is also higher among these set of individuals who operate at the margin – who have looser ties to employment.

A proxy for measurement error was constructed by calculating the standard deviation in the reported level of education across waves. The variable finds some traction. Individuals who drastically misreport their education are 4% more likely to belong the second group who experiences more employment mobility.

The descriptive results therefore suggest that most of the churning in employment is due to a relative small group of individuals who are (or at least report to be) continuously transitioning in and out of employment. This small group of individuals appear to be operating on the margin, where they are more likely to be informally employed in low paying occupations when employed and to be actively looking for work when non-employed.

7. Conclusion

In this paper we showed that classic employment transition, which is derived by looking at the reported transitioning between two successive periods, produce results that are at odds with multiple period transition estimates as well as the aggregate age-employment profile. In an attempt to figure out what could be driving these irregularities we developed our own maximum likelihood method which fits on an additional set of moments. Having more moments to fit on has the added bonus of allowing us to introduce more parameters to our model. In our analysis we are able to allow for differences in observable covariates, measurement error, and unobserved individuals specific heterogeneity.

²⁶ The sample is limited to those who are reported to have been employed at least once since we require information on job specifications.

Our results show that that the reported transition probabilities tend to overestimate employment mobility. Although we focused our analysis on black males, we are confident that this will also be true of for other race groups and for females.

When we allow for misclassification alone, we find that up to two thirds of the variation in transition rates may be driven misclassification. Similarly, when we allow for unobserved heterogeneity, we find that most of the variation is driven by a small group of individuals who continuously transition in and out of employment. The entry and exit rates faced by most individuals are far lower than were originally reported. The original model predicted an entry rate of 17% and an exit rate of 14%. The finite mixture model and the misclassification model find that the true entry rate that are faced by most individuals would be closer to 4%, while the true exit rate faced by most individuals would be closer to 1 or 2%.

Those individuals who are closest to the margin (have only recently lost or gained employment) appear to be most likely to lose or gain employment, while those further from the margin are less likely to transition in and out of employment.

Appendix 1

Appendix 1.A: Additional Tables

Table 1.A.1: Observations of Working-Aged Black Males per Wave

	t = 1	t = 2	t = 3	t = 4	t = 5	t = 6	Total
Observed only once	4,865	2,748	1,652	1,398	1,921	4,316	16,900
Observed more than once	10,379	12,742	12,318	12,396	11,681	9,669	69,185
Total	15,244	15,490	13,970	13,794	13,602	13,985	86,085
Percentage attrition between waves		54.5%	56.7%	56.5%	57.3%	59.0%	
Percentage remaining from initial wave		54.5%	65.0%	76.0%	85.1%	91.6%	

Table 1.A.2: Observations of Working-Aged White Males per Wave

	t = 1	t = 2	t = 3	t = 4	t = 5	t = 6	Total
Observed only once	577	291	212	197	288	576	2,141
Observed more than once	1,066	1,564	1,460	1,485	1,459	1,192	8,226
Total	1,643	1,855	1,672	1,682	1,747	1,768	10,367
Percentage attrition between waves		59.8%	54.0%	60.2%	54.1%	56.4%	
Percentage remaining from initial wave		59.8%	66.2%	78.1%	85.5%	90.9%	

Table 1.A.3: Transitions in and out of Employment for Black Males by Age Group

Aged 20 - 30			Aged 30 - 40			Aged 40 - 50		
From	To		From	To		From	To	
	N	E		N	E		N	E
N	85.9%	14.1%	N	75.5%	24.5%	N	77.6%	22.4%
E	29.1%	70.9%	E	13.4%	86.6%	E	11.8%	88.2%

**Tabulated as average six month transition probabilities.*

Table 1.A.4: Transitions in and out of Employment for White Males by Age Group

Aged 20 - 30			Aged 30 - 40			Aged 40 - 50		
From	To		From	To		From	To	
	N	E		N	E		N	E
N	69.9%	30.1%	N	50.0%	50.0%	N	59.3%	40.7%
E	8.3%	91.7%	E	2.0%	98.0%	E	2.9%	97.1%

**Tabulated as average six month transition probabilities.*

Table 1.A.5: Entry Rate, Exit Rate and Steady State of Employment by Race and Age Group

	Black				White			
	20-30	30-40	40-50	Overall	20-30	30-40	40-50	Overall
Entry Rate	14.1%	24.5%	22.4%	17.6%	30.1%	50.0%	40.7%	35.6%
Exit Rate	29.1%	13.4%	11.8%	16.8%	8.3%	2.0%	2.9%	3.7%
Steady State	32.6%	64.6%	65.6%	51.1%	78.3%	96.2%	93.4%	90.6%

Table 1.A.6: Maximum Likelihood Estimates

	(Model 2a)	(Model 3a)	(Model 4a)	(Model 5a)
Job Entry				
Constant1	0.104 (0.000)	0.041 (0.000)	0.038 (0.001)	0.031 (0.001)
Constant2			0.437 (0.001)	0.318 (0.001)
Job Exit				
Constant1	0.094 (0.000)	0.017 (0.000)	0.011 (0.001)	0.000 (0.001)
Constant2			0.508 (0.001)	0.495 (0.001)
Proportion of Type 1			0.709 (0.001)	0.651 (0.001)
Misclassification1		0.067 (0.000)		0.019 (0.001)
Misclassification 2		0.077 (0.000)		0.000 (0.001)
N	43013	43013	43013	43013
Log-likelihood	-41216	-39659	-39447	-39322
Allow for Misclassification	No	No	Yes	Yes
Allow for Observed Heterogeneity	No	No	No	No
Allow for Unobserved Heterogeneity	No	Yes	No	Yes

Table 1.A.7: Maximum Likelihood Estimates

	(Model 2b)	(Model 3b)	(Model 4b)	(Model 5b)
Job Entry				
Constant1	0.050 (0.000)	0.049 (0.000)	0.151 (0.000)	- -
Age20	0.013 (0.000)	0.009 (0.000)	0.008 (0.000)	- -
Age30	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)	- -
TimeGap < 1	0.130 (0.001)	0.143 (0.001)	0.138 (0.001)	- -
TimeGap = ± 2	0.078 (0.001)	0.079 (0.001)	0.086 (0.001)	- -
Matric	0.013 (0.001)	0.009 (0.001)	0.050 (0.001)	- -
More than Matric	0.072 (0.001)	0.063 (0.001)	0.081 (0.001)	- -
Constant2			0.017 (0.001)	- -
Job Exit				
Constant1	0.234 (0.001)	0.237 (0.001)	0.204 (0.001)	- -
Age20	-0.017 (0.000)	-0.018 (0.000)	0.000 (0.000)	- -
Age30	0.000 (0.000)	0.000 (0.000)	0.001 (0.000)	- -
Tenure < 1	0.254 (0.001)	0.238 (0.001)	0.251 (0.001)	- -
Tenure = ± 2	0.091 (0.001)	0.082 (0.001)	0.070 (0.001)	- -
Matric	-0.028 (0.001)	-0.029 (0.000)	-0.076 (0.001)	- -
More than Matric	-0.053 (0.001)	-0.049 (0.000)	-0.103 (0.001)	- -
Constant2			0.014 (0.001)	- -
Proportion of Type 1			0.402 (0.001)	- -
Misclassification1		0.024 (0.001)		- -
Misclassification 2		0.012 (0.001)		- -
N	43013	43013	43013	-
Log-likelihood	-38573	-38354	-38546	-
Allow for Misclassification	No	No	Yes	Yes
Allow for Observed Heterogeneity	Yes	Yes	Yes	Yes
Allow for Unobserved Heterogeneity	No	Yes	No	Yes

Table 1.A.8: Posterior Class Probabilities

	Observations	P(Type 1)	P(Type 2)
Age			
Between 20 and 30	20390	71.0%	29.0%
Between 30 and 40	13490	71.0%	29.0%
Older than 40	9132	70.6%	29.4%
Education			
Less than Matric	30943	70.6%	29.4%
More than Matric	11868	71.7%	28.3%
Head			
Not Household Head	21520	70.6%	29.4%
Household Head	21493	71.2%	28.8%
Marital Status			
Not Married	26105	70.4%	29.6%
Married	16908	71.8%	28.2%
Children			
No Children in Household	17797	70.7%	29.3%
Children in Household	25216	71.1%	28.9%
Proxy for Misreporting			
Standard Deviation of Education < 2	15794	71.30%	28.70%
Standard Deviation of Education > 2	4589	67.80%	32.20%
Type of Employment			
Informal	5822	63.74%	36.26%
Formal	15114	67.67%	32.33%
Wage			
Lower than Median	11556	63.49%	36.51%
Higher than Median	11556	71.22%	28.78%
Actively Looked for Work			
No	11906	68.90%	31.10%
Yes	11544	62.65%	37.35%

Appendix 1.B: Cleaning ‘Tenure’ and ‘Time Gap’

Our model uses both the ‘tenure’ and ‘time gap’ variables. Having repeated observation allows us to test whether the values of these two variables are consistent across waves. Irregular responses were corrected when we were confident about what the actual response was supposed to be and dropped when we were not. Two rules were followed.

Rule 1: “Tenure” cannot be higher than potential experience.

The assumption was made that tenure needed to be smaller or equal to potential experience (the maximum amount of experience a worker would have accrued if he started school at age 6, passed all grades in the allotted time and started working thereafter). To prevent us from wrongly disqualifying some individuals who might have started school early or skipped grades, we set the maximum at 2 years higher than potential experience.

Rule 2: Between two waves ‘tenure’ and ‘time gap’ can either reset to 0 or increase by half a year. Only one of two things can happen to tenure (and time gap) between waves; it can either increase by 6 months (when an individual stays in a job) or it is reset to zero (person quits or changes jobs). Since we have 5 transition periods between our 6 waves, there are 32 (2⁵) different permutations of how tenure (and ‘time gap’) could have evolved over the three years.

Appendix 1.C: Linear Probability Model

The linear probability model allows us to run a regression on the 1-period transition rates. In the two tables below we allowed the job entry and exit rates to vary by age²⁷, education and by recent employment history.

Table 1.C.1: Linear Probability Model Estimates

	(Model 1a)	(Model 1b)
Job Entry		
Constant	0.174*** (0.003)	0.050*** (0.006)
Age20		0.016*** (0.001)
Age30		-0.002* (0.001)
TimeGap < 1		0.148*** (0.011)
TimeGap = ± 2		0.070*** (0.011)
Matric		0.036*** (0.007)
More than Matric		0.092*** (0.017)
N	15135	15135
R²	0.000	0.047

Table 1.C.2: Linear Probability Model Estimates

	(Model 1a)	(Model 1b)
Job Exit		
Constant	0.141*** (0.003)	0.275*** (0.002)
Age20		-0.019** (0.002)
Age30		-0.002** (0.001)
Tenure < 1		0.228*** (0.008)
Tenure = ± 2		0.077*** (0.008)
Matric		-0.014* (0.007)
More than Matric		-0.092*** (0.010)
N	12746	12746
R²	0.000	0.115

²⁷ Two sets of splines were added to the model to capture the non-linearity in age. One spline runs from age 20 to 30, the other from age 30 to 50. The splines are easily interpretable and fit the data well.

The entry rates are estimated in table 1.C.1. Model 1a does not control on any covariates. The average entry rate is 17.4%.

In model 1b the probabilities of entering employment is allowed to vary on a set of covariates. The results show that a 20 year old black male without any work experience or a matric has a 5% probability of moving into employment in the next 6 months. The probability of transitioning into employment increases dramatically between the age of 20 and 30. The probability does not shift greatly after 30. Individuals who have not had work within the last three years are least likely to transition into employment. Individuals who have worked within the last 3 years are 13 percentage points more likely to transition into employment than those who have not worked within the last 3 years. Having worked within the last years increases the probability by a further 6 percentage points. Educated individuals are most likely to transition into employment. Relative to not having matric, having matric increases the probability of becoming employed by 3.7 percentage points, while having more than matric increases the probability by a further 5.5 percentage points.

The exit rates are contained in table 1.C.2. In model 1a we derive an average reported exit rate of 14.1%. In model 1b we see that the probability of exiting employment decreases drastically between the age of 20 and 30. Beyond the age of 30 however age does not appear to have much of an effect. Having been in a position for a while drastically decreases the likelihood of becoming non-employed. As soon as someone has been in a position for more than a year, their likelihood of exiting within the next 6 months drops by 15 percentage points. Individuals with higher education are significantly less likely to become non-employed.

Appendix 1.D: The Stationarity Assumption

We restrict our sample to the subsample of black males that was observed in all 6 waves. The table below shows each of the five 1-period transition rates as well as the averaged 1-period transition rate. The values should be comparable if stationarity holds.

Table 1.D.1: Average Transition Rates between waves

	T_{12}	T_{23}	T_{34}	T_{45}	T_{56}	Average
Job Entry Rate	22.8%	18.6%	22.9%	20.2%	23.0%	21.5%
Job Exit Rate	9.4%	15.0%	13.7%	12.8%	16.2%	13.5%
N	286	286	286	286	286	1430

The transition rates are generally of a similar magnitude. We conclude that the assumption of stationarity between the different waves is a plausible one.

Appendix 1.E: The Markov Assumption

If the Markov condition holds, any multiple period transition matrix should be equivalent to the product of two shorter period matrices ($T_{ik} \neq T_{ij}T_{jk}$). To test whether this condition holds we compare the 2-period observed transition matrices against what they would have been if we combined the two subsequent 1-period transition matrices. If the Markov assumption holds and our model is correct, then these two approaches should produce comparable estimates.

Table 1.E.1: Average Transition Rates between waves

	T_{13}	T_{24}	T_{35}	T_{46}	Average
Observed Entry Rate	25.2%	20.4%	25.4%	26.3%	24.4%
Observed Exit Rate	14.5%	13.9%	14.9%	18.0%	15.3%
	$T_{12}T_{23}$	$T_{23}T_{34}$	$T_{34}T_{45}$	$T_{45}T_{56}$	Average
Implied Entry Rate	33.7%	34.7%	35.5%	35.3%	35.5%
Implied Exit Rate	21.3%	23.2%	22.0%	24.0%	22.2%
N	286	286	286	286	1144

Generally, the 1-period transition matrices produce 2-period transition rates that are much higher than what were observed. Similarly, we can compare the 3-, 4- and 5-period transitions rates against what they should be if the Markov condition holds.

Table 1.E.2: Average Transition Rates between waves

	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$d = 5$
Observed Entry Rate	21.5%	24.4%	26.8%	26.3%	29.1%
Observed Exit Rate	13.5%	15.3%	16.0%	15.3%	15.7%
Implied Entry Rate	21.5%	35.5%	44.6%	50.5%	54.4%
Implied Exit Rate	13.5%	22.2%	27.9%	31.6%	34.1%
N	1430	1144	858	572	286

These values differ greatly. The Markov condition does not appear to hold. In the absence of misclassification or heterogeneity, the multiple period transition rates are much lower than 1-period transition rates would lead us to believe.

Appendix 1.F: The Initial State Assumption

Not everyone joins the labour force at the same time. We compare the school enrolment rates for black and white males to see whether there is a difference in their school-leaving patterns.

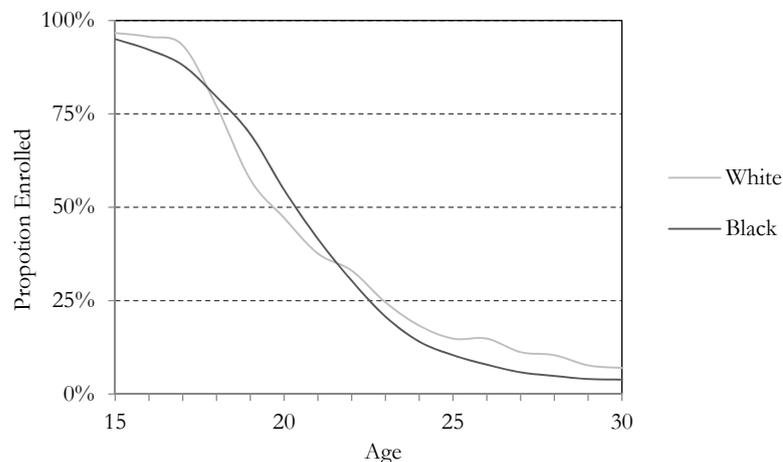


Figure 1.F.1: Probability of Being Enrolled by Age and Race

The difference in enrolment likelihood by age is negligible. Roughly half of black males and white males leave school by age 20. To keep our model as simple as possible we will assume that everyone joins the labour market at age 20.

Appendix 1.G: Identification

The Markov and stationarity assumption allow us to identify misclassification and heterogeneity. Earlier we showed that the Markov assumption does not hold under a simpler specification (when we had homogeneity and no misclassification). The following identity failed to hold: $T_k = (T_1)^k$ for any k .

In our classical model, where we assume that the transition rates are constant, we only have two parameters. These two parameters are unable to fit all the moments that we incorporated into our model. Once we allow for heterogeneity and measurement error, we free up some parameters that are entrusted with fitting the additional moments.

If we allow for misclassification we are able to restore these identities, since we have two more degrees of freedom to work with. To illustrate the point, we set $k = 2$ and rewrite the identity of earlier, this time allowing for misclassification matrices.

$$(M^{-1})'T_2^*(M^{-1}) \cong [(M^{-1})'T_1^*(M^{-1})]^2$$

$$T_2^* \cong (M^{-1})[T_1^*]^2M(M^{-1})'$$

We already know that $T_2^* \neq [T_1^*]^2$. Now we are able to set M so that T_2^* is comparable to $(M^{-1})[T_1^*]^2(M^{-1})'$.

Similarly, if we allow for heterogeneity, the three additional parameters allow us to restore the identities that were built into our likelihood function. Below we denote $T_{1,1}^*$ as the 1-period transition matrix for someone of type 1 and $T_{1,2}^*$ as the 1-period transition matrix for someone of type 2. π_1 and π_2 denote the proportion of type 1 and type 2 individuals.

$$T_2^* \cong \pi_1 [T_{1,1}^*]^2 + \pi_2 [T_{1,2}^*]^2$$

Unfortunately, once we allow for both misclassification and heterogeneity in our model, identification struggles as the model becomes over-specified. For the sake of completeness we added the results for model 5a (see table A8) that allows for misclassification and unobserved heterogeneity. Model 5b, which also controls for observed heterogeneity, did not converge.

Chapter 2:

Using Dynamic Programming to Model Job Search and Reservation Wages

The unemployment rate is high in South Africa, especially among the young. In this chapter we attempt to distinguish whether unemployment is predominantly driven by demand constraints or supply constraints – whether unemployment is due to low job arrival rates or high reservation wages. If we are willing to trust the self-reported responses in surveys, then reservation wages appear to be contributing to the unemployment problem. Unfortunately, these self-reported responses have been shown to be subjective and the reservation wage estimates derived from these self-reported upwardly bias. Job search theory provides us with an alternative. Rather than asking individuals what they think their reservation wages are we attempt to uncover what their reservation wages are based on their behaviour.

1. Introduction

The current unemployment rate is estimated at 25% if we follow the strict definition and 35% if we use the broad definition, which also includes those who have stopped actively looking for work (StatsSA, 2015a). Youth unemployment in South Africa is even higher. Unemployment is particularly severe among new entrants. The narrow unemployment rate among the young (between the ages of 15 and 34) is estimated at 36.9% (StatsSA, 2015b).

The increase in unemployment among the youth is generally believed to be due to an oversupply of labour relative to demand. Since the end of apartheid there has been a steady increase in labour force participation and a decline in the relative demand for unskilled labour (Burger & von Fintel, 2009). The surplus of unskilled labour caused by the increased participation has made it tougher for young entrants into the labour market to find work (Mlatsheni & Rospabe, 1999; Bhorat & Oosthuizen, 2007). According to Rankin and

Roberts (2011), the higher levels of unemployment among the young could be due to higher reservation wages. The young may be less willing to accept low-paying jobs, since many of them are still receiving financial assistance from their households. According to job search literature, unemployment can also arise through imperfect information. If individuals have imperfect information about vacancies and future job offers this will add to the unemployment burden. Similarly, as both employees and employers have preferences, young entrants may take time to find a job that is suitable to their set of skills and tastes (McCall, 1970).

In this chapter we will construct and estimate a discrete choice dynamic programming job search model that is more fitting to the South African labour market landscape than conventional OLS and sample selection models. In contrast to these commonly adopted models, our model allows for non-myopic decision-making under uncertainty and is capable of explaining unemployment. By modelling the process explicitly we hope to be able to uncover whether the slow absorption rates among the young is actually a choice – whether unemployment is driven by low job arrival rates and low reservation wages or high job arrival rates and high reservation wages. Thus far, South African literature has relied on self-reported reservation wages that are prone to over-reporting. The method we adopt provides us with an alternative. Rather than asking individuals what they think their reservation wages are, the chapter attempts to uncover reservation wages that are consistent with labour market behaviour. This is, to our knowledge, the first paper that attempts to recover reservation wages through a theory-consistent job search model for an African country.

In the following section we look at the previous South African attempts to link reservation wages and employment. We also review the international literature on job search, where structural models are used to recover plausible reservation wage estimates. In section 3 we formulate our job search model, by introducing the behavioural assumptions that guide the action of the optimising agents in our theoretical model. In section 4 we show how we plan to go about to solve the model. In section 5 we introduce the data, show some descriptive results and proceed to solve the parameters. In section 6 we conclude.

2. Previous Work

2.1. South African Literature

To our knowledge, there are five papers that have linked reservation wages to employment data in South Africa. Kingdon and Knight (2001) compares self-reported reservation wages to predicted wages for a group of unemployed individuals. According to the authors, the reservation wages are far higher than one would have expected. While it is possible that individuals may overestimate their earning potential, the authors believe that most of deviation is due to over-reporting. Rankin and Roberts (2011) also find evidence that reservation wages appear to be overestimated. The authors show that almost half of the unemployed have reservation wages that were higher than what individuals with similar observables are

currently earning in the labour market. Natrass and Walker (2005) compare the reported reservation wages to the predicted wages among a representative sample of Khayelitsha and Mitchells Plain residents. Their findings suggest that workers have realistic wage expectations. Following a job-search approach, Levensohn and Pugath (2010) use the self-reported reservation wages in the Cape Area Panel Study (CAPS) to examine the role of reservation wages on unemployment. According to their model job offers are frequent, but are low relative to reservation wages. The authors estimate that more than 70% of job offers are declined. According to Zoch (2014) individuals report differently when they are asked to report their lowest reservation wages compared to when they are asked whether they would accept specific reservation wages. Using CAPS data, the authors obtain an imputed reservation wage using a sequence of hypothetical job offer responses. The estimates they derive through this method are lower than the conventional reservation wage estimates. Their estimates also do better in predicting labour market outcomes.

Importantly, all six of these papers rely on self-reported reservation wages. Generally these measures have been found to be biased upwards (Brown & Taylor, 2011). Respondents tend to report what they believe a fair wage for them would be rather than the lowest wage they would be willing to work for. Similarly, individuals may mistake the surveying process as a wage negotiation process and try to bargain for a wage that is higher than their true reservation wage (Pettersson, 1997; Kingdon & Knight, 2001). It is therefore not uncommon to see respondents accept wages that are lower than their reported reservation wages (Krueger & Mueller, 2014).

In our analysis we will follow a structural approach that does not rely on the self-reported reservation wages responses. Instead, our model tries to recover a reservation wage that is consistent with the actions and choices we observe in the labour market, most notably the employment and job search decisions.

2.2. Job Search Literature

The first sequential job search models were developed by McCall (1970), Mortensen (1970) and Gronau (1971). These models build on the earlier human capital model of Becker (1964). In his earlier work Becker (1964) argues that a move out of non-employment into employment will only be worthwhile if the move is profitable – if the future income increases attached to accepting a job exceed the direct costs of moving. Job search theory builds on this model, allowing individuals to weigh up the costs and benefits of holding out for a better job offer when deciding whether to accept a current job offer.

The first wave of job search papers show that unemployment can be accurately modelled as a stochastic process. By relaxing the perfect knowledge assumption that guides neo-classical models, job search models are able to show how forward-looking individuals who would prefer to be employed may temporarily be out of work. Unemployment is introduced through the uncertainty and imperfect information regarding vacancies and future job offers.

The basic sequential job search models allow for two states: employed and searching. Job offers arrive randomly. An unemployed person is assumed to receive a job offer, drawn from a specific distribution function, in each period with fixed probability. Wage offers are assumed to be i.i.d. and there is no recall – once a job offer is refused an individual is unable to change their mind. The earlier models assume that the time horizon is infinite and that once a job is accepted it is held forever. None of the original models allow for on-the-job search or transitions out of employment (McCall, 1970; Mortensen, 1970; Gronau, 1971).

The crucial question within these earlier models was whether and when it would be worthwhile for an optimizing agent who is faced with a specific job offer to hold out for a better job offer. In deciding, an agent has to weigh up the possible future gains of a better job offer against the disutility of not working and the opportunity cost of not taking the offer. Generally, these early models were solved using a stopping rule - the minimum wage for which individuals would be willing to enter employment - a reservation wage²⁸.

A second wave of structural job search models followed, which relaxed some of the basic assumptions of the first set of job search models. The three assumptions that appear to have been most widely criticised are the constant reservation wage assumption (which assumes that the reservation wages are constant across individuals and over each individual's life-span), the infinite time horizon assumption and the assumption that individuals are not allowed to change jobs once they are employed.

Kiefer and Neumann (1979) examine the validity of the constant reservation wage (or stopping rule value) hypothesis within the original models. Following a group of comparable workers who were laid off when their plant closed down, they show that reservation wages vary greatly between individuals. Reservation wages are higher for workers with unemployment benefits and higher for those with greater earning potential. Reservation wages decrease the longer workers remain unemployed. Lastly, they find that reservation wages start high, drop to a minimum at age 26 and rise again thereafter. Jensen and Westergaard-Nielsen (1987) follow a similar approach to that of Kiefer and Neumann (1979). Using a sample of new law graduates Jensen and Westergaard-Nielsen (1987) show that there was a global reservation wage under which none of the applicants were willing to accept work. They also find that the intensity of the job search is positively correlated to the job arrival rate and to the reservation wage.

Flinn and Heckman (1982) developed a standard search model with an infinite time horizon, which they fit to a panel of young US males. According to their results, between 30% and 45% of individuals will be offered employment over a six month period and between 60% and 92% of those job offers will be accepted. Blau (1991) follows a group of white, male, high school graduates in the US. He finds that the arrival rate of job offers decreases with unemployment duration and estimate an acceptance rate that is closer to 100%. Unlike the classical models, Wolpin (1987) allows incorporates measurement error into his

²⁸ Since models assume that lifetimes were infinite, the reservation wages of these earlier models were not dependent on age.

model and allows for a finite rather than an infinite time-horizon. He also allows the probability of a job offer to vary depending on how long someone has been out of school: individuals have a 1.28% probability per week of receiving an offer, but the probability declines to 0.91% over time.

Early job search models assumed that jobs were held forever, there was thus no reasons to allow for on-the-job search. Recent papers have relaxed this assumption. Using a reduced form approach that controls for method of search, Blau and Robins (1990) show that the arrival rate of job offers was higher among the employed searching than the unemployed searching. Conversely, Holzer (1987) finds the opposite to be true. Controlling for some labour and demographic variables, he shows that the probability of receiving a job offer is higher for the unemployed that are searching than for the employed. Belzel (1996) uses a structural approach to look at relative efficiency of job search, more specifically, comparing the probability of receiving a job offer and the distribution of the job offers between unemployed search and employed search.

Most job search models assume that the cost of searching is immediate and the benefits are delayed. Paserman (2008) shows that in the face of hyperbolic time-preferences agents will underinvest in job search activities. For those agents with a present bias, the cost associated with searching will be weighed more heavily relative to the expected future gains. As a result, these agents will be less inclined to search and search with less intensity than what would have been optimal if they were trying to maximize lifetime utility.

3. A Dynamic Job Search Model

A discrete choice dynamic programming model is used to mimic the transition between work, job search and household production for a group of black males.

Each individual faces a finite decision horizon during their working life. For the sake of this model we will assume that the working age stretches between the age of 16 and 65, 16 being the minimum age at which one can legally be employed and 65 the minimum age at the time for which males would be eligible to receive old-age pensions²⁹. To bring the model in line with the bi-annual data that will be introduced in the second portion of the paper each period will be defined as lasting exactly six months. This leaves each individual with a maximum of 100 discrete decision periods over which to maximize their discounted net present lifetime utility.

$$U_i = \sum_{t=1}^{100} \beta^{t-1} u_{it} \quad (1)$$

²⁹ Since 2008 the age of eligibility for males to receive the old age grant has been gradually reduced from 65 to 60 (SASSA, 2008). The LFS data we will be using was collected between 2001 and 2004 when the age of eligibility for males was still 65.

Unlike some other dynamic programming models (see Keane & Wolpin, 1997; Belzil & Hanzen, 2002), the educational attainment choice is considered to be predetermined, a topic that will be revisited in the next chapter, where education is allowed to be endogenously determined.^{30,31}

For each period, an individual can pursue one of three alternatives: staying at home, actively looking for work or working. The observed choice set, d_i , that each individual faces, is denoted as follows:

$$d_{it} = \begin{cases} 1 & \text{household production} \\ 2 & \text{actively seeking work} \\ 3 & \text{working} \end{cases}$$

The instantaneous utility derived in each period will depend on the rewards (denoted as $r_{d,it}$) associated with each of the chosen activities.

$$u_{it} = I(d_{it} = 1)r_{1,it} + I(d_{it} = 2)r_{2,it} + I(d_{it} = 3)r_{3,it} \quad (2)$$

The discounted lifetime utility can be obtained by substituting equation (2) into equation (1):

$$U_i = \sum_{t=1}^{100} \beta^{t-1} \{I(d_{it} = 1)r_{1,it} + I(d_{it} = 2)r_{2,it} + I(d_{it} = 3)r_{3,it}\} \quad (3)$$

Unlike the Mincerian model and Heckman model, we allow agents to be forward-looking rather than myopic³². Individuals in our model are willing to forfeit some utility in the short run if they believe that it will increase future utility.

3.1. Home Production and Search Cost

The average level of home production is denoted by γ_H . Home production signifies the utility attached to the outside option. The utility from home production is allowed to vary by age in the following non-linear manner:

$$r_{1,it} = \gamma_H = \gamma_0 + \gamma_1 \text{Age} + \gamma_2 \text{Age}^2 \quad (4)$$

Individuals incur a cost in searching for work; this cost could be direct (e.g. travelling costs or phoning up potential employees) or indirect (e.g. less time for housework or leisure). The average level of utility when searching is allowed to differ from what it was under home production.

The utility under active search is captured by the equation below, where

³⁰ To derive potential experience we assume that everyone who attains less than a grade 11 education joins the labour market at age 16 and everyone who attains grade 11 or more finishes their education within the allotted time and joins the labour market the year thereafter.

³¹ Once someone is done studying they are not allowed to return to school.

³² Choosing to maximise present lifetime utility (equation 3) rather than instantaneous utility (equation 2).

$$r_{2,it} = \gamma_H - \gamma_S \quad (5)$$

The utility forfeited by searching is captured by γ_S , where γ_S is assumed to be i.i.d. normally distributed with mean u_S and deviation σ_S ³³.

Although our model does not strictly assume it, one would expect the instantaneous utility to be lower when searching. Therefore we would expect u_S to be larger than 0. Similar to Paserman (2008) we assume that the cost to searching is immediate, while the benefits are delayed. Those individuals that self-select into searching will need to be willing to forego some utility now, in order to increase their probability of securing higher future remuneration.

3.2. Reward for Working (Wage)

The current period reward associated with employment follows the conventional Mincerian shape, which combines time spent in school and time spent working. The expected wage offer conditional on one's level of education and experience can be expressed as follows

$$r_{3,it} = \bar{w}_i = \alpha_0 + a_1 educ_{it} + a_2 educ_{it}^2 + a_3 exp_{it} + a_4 exp_{it}^2 \quad (6)$$

where a_1 and a_2 measure the effect of schooling and a_3 and a_4 represent the effect of experience³⁴.

The true wage offer further depends on the remuneration draw.

$$w^* = \bar{w}_i + \eta_{i,t} \quad (7)$$

The remuneration draw, $\eta_{i,t}$, is assumed to be i.i.d. normally distributed with mean zero and standard deviation σ_d . Throughout we assume that individuals are aware of the distribution from which their wage offers will be drawn, but are not aware of the realisation of any of these draws before they occur. These values are only revealed when the job offers are made.

3.3. Job Offers

Not everyone gets offered a job during every period. The frequency of job offers (also referred to as the job arrival rate) is allowed to differ depending on whether an individual has searched in the previous period or not. The arrival rate for non-searching individuals is denoted as h_H and the arrival rate for the searching non-employed as h_S . Although we do not explicitly restrict h_S to be higher than h_H , we would expect this to be the case - looking for work should increase the likelihood of being offered work.

³³ Without the variation introduced through γ_S , the model is unable to explain the high degree of transitioning between home production and searching that is present in the data.

³⁴ We will model potential rather than actual experience. The DCDP model approach allows one to model the prior and then fit on the latter; this approach will however complicate the model dramatically and require us to fit on simulated moments. For this reason we decided to follow the more conventional approach.

Not all remuneration draws are accepted. After a job offer is made, individuals have to decide whether they want to accept the job offer or whether they want to remain non-employed. According to McCall (1970) prospective employees will keep on searching as long as the future benefits associated with waiting for a better remuneration draw outweigh the forfeited wage that one could currently be earning.

If agents decide to take the job offer the remuneration draw gets ‘locked in’ and remains ‘locked in’ until an employee exits employment or accepts a job from another employer. The probability of exiting employment will be denoted by q and is assumed to be constant across all individuals and all periods.

3.4. On-the-Job Search

The probability of receiving a new job offer when a worker is already employed is denoted by h_E . In our model we assume that employed workers will only accept a new job offer if the wage offer attached to that job offer exceeds their current job offer. The cost of searching while employed is assumed to be negligible. This assumption is made out of convenience rather than plausibility³⁵. Allowing on-the-job search decreases the reservation wage, since workers no longer have to hold out for high paying jobs before joining the labour market³⁶ (Cahuc & Zylberberg, 2004).

4. Estimation Issues

4.1. Difference between Offered Wages and Observed Wages

In our analysis we will be using the reported wages to recover a plausible distribution of wage offers. In identifying the wage offer distribution we need to be wary of the difference between the offered wages, the accepted wages and the observed wages. Selection drives a wedge between the first two, while measurement error drives a wedge between the last two.

4.1.1. Measurement Error

Job search models are more susceptible to measurement error than conventional wage regressions, since these models use the bottom tail of observed earnings distribution to identify reservation wages³⁷. Throughout we will assume that logged wages are reported with classical measurement error. We set $w = w^* + \varepsilon$, where w^* is the true wage and w is the observed wage. The error is assumed to be normally distributed with mean zero and standard deviation σ_ε . In order to simplify the notation we follow the approach of Wolpin (1987) in setting $\theta = \varepsilon + \eta$. Here, θ represents the difference between the expected and reported wage, which is a function of the measurement error and the remuneration draw. Since both ε

³⁵ Since we do not have data on on-the-job search, we are forced to assume that everyone partakes in on-the-job job-search to a similar extent.

³⁶ The extent to which on-the-job search will affect initial reservation wages will depend on the difference between the job offer rates in the three states.

³⁷ The original models did not allow for measurement error or heterogeneity. In these models the reservation wage was required to be lower than the lowest observed wage. These models would be even more susceptible to outliers.

and η are normally distributed it follows that θ will be normally distributed with mean of zero and standard deviation³⁸ $\sigma_\theta = \sqrt{\sigma_\varepsilon^2 + \sigma_\eta^2}$.

4.1.2. Selection Bias

A second reason why observed wages may differ from offered wages is due to positive selection into employment. Intuitively, we would expect agents, who abide to their reservation wages, to accept high offers and reject low offers. Therefore, the average accepted wage should be higher than the average offered wage³⁹ (Flinn & Heckman, 1982).

The relationship between the expected accepted wage offers, $E(w)$, and the expected wage offers, $x\beta$, can be recovered if we are willing to make some assumptions about the relationship between unobservables in the selection and wage equations. The expected observed wage can be written as

$$E(w) = x\beta + p\sigma_\varepsilon \left[\frac{\phi\left(\frac{x\beta - rw_i}{\sigma_\theta}\right)}{\Phi\left(\frac{x\beta - rw_i}{\sigma_\theta}\right)} \right]$$

The cumulative distribution function of the normal distribution is denoted by Φ and the density function is denoted by ϕ . $p = \frac{\sigma_\varepsilon}{\sigma_\theta}$ denotes the correlation between the two errors. In relative terms, p denotes the impact of the errors in the selection decision on the observed wage. The term in squared brackets denotes the selection correction term, commonly referred to as the inverse Mill's ratio (Heckman, 1979).

4.2. Recovering Reservation Wages

Individuals face two decisions. The first decision is whether or not to actively look for work when non-employed⁴⁰. The second decision is whether or not to accept a job offer when one is made.

The reservation wage for individual i at period t is the wage at which the individual would be indifferent between accepting and rejecting a job offer. Reservation wages are useful for our modelling purposes since they enable us to trace out the optimal employment choices for different individuals through their working lives.

Solving the optimal decision path requires agents to be capable of making decisions under uncertainty. Agents are not aware of what their future shocks will be, but are assumed to have perfect knowledge with regard to the probability distribution from which these future shocks will be drawn. Agents in our model incorporate all this information in their decision-making process. Finding the optimal path is complicated further by the fact that current decisions have implications for future rewards and choices. Due to these

³⁸ Technically the standard deviation is $\sigma_\theta = \sqrt{\sigma_\varepsilon^2 + \sigma_\eta^2 - 2\sigma_{\varepsilon\eta}}$ where $\sigma_{\varepsilon\eta}$ denotes covariance between the two error terms. Throughout we will however assume that the measurement error and the remuneration draws are uncorrelated and that $\sigma_{\varepsilon\eta} = 0$.

³⁹ Due to censoring at the lower end, the remuneration draws for observed wages is likely to be skewed to the left.

⁴⁰ Although our model allows for on-the-job search, we will assume that on-the-job search is automatic.

complexities econometricians make use of discrete choice dynamic programming (DCDP). DCDP models have the desirable attribute of being both complicated and flexible at the same time. DCDP models are intricate enough to allow a worker to weigh up all the future paths and consider all future shocks that may occur (their likelihood of occurring and how those would impact future returns) when they have to decide on the optimal current choice. Yet, these models are flexible enough to allow workers to renege on their optimal path when shocks get revealed in the future⁴¹.

To trace out the optimal decision path we introduce the state space Ω_t . The state space contains all the relevant information from the past that could influence the current-period decision d_t . The model is assumed to evolve in Markovian manner with a one-period memory. The current state is determined by education, experience, the previous state and the remuneration draw⁴²

Education is assumed to be constant and potential experience evolves in a predictable manner. The previous period's state determines whether the individual starts the period as employed or not. It also affects the job arrival rates, since job offer arrival rates differ depending on whether the individual have searched the previous period⁴³.

In deciding their optimal path, individuals compare the net discounted lifetime utility of the options available to them. We define $V(\Omega_a, a)$ as the value function at period a . It provides the maximum expected present value of lifetime utility at age a , given the state space Ω_a and the discount factor β .

$$V(\Omega_a, a) = \max_{\{d\}} E \left[\sum_{t=a}^{100} \beta^{t-a} [I(d_t = 1)r_{1,t} + I(d_t = 2)r_{2,t} + I(d_t = 3)r_{3,t}] | \Omega_t \right] \quad (8)$$

In its current form the expansion of decisions faced by our agents follows the specifications of a finite Bellman equation⁴⁴. Consequently, the value function can be rewritten as the highest of the 3 alternative value functions, where each of the subsequent value functions corresponds with one of three choices contained in our decision space.

$$V(\Omega_a, a) = \max_{\{d\}} [V_1(\Omega_a, a), V_2(\Omega_a, a), V_3(\Omega_a, a)] \quad (9)$$

Since each of these value functions (V_1, V_2 and V_3) obey the Bellman principles, they can each be rewritten as the sum of their respective reward plus the discounted sum of the value function for the following period.

⁴¹ In the presence of shocks what was previously envisioned to be the optimal decision path might turn out to be an inferior strategy once an array of shocks is realised.

⁴² Past realisations of the two productivity shocks (ε_t^w and ε_t^h) are not included since their realisations are assumed to be serially independent of future shocks.

⁴³ The remuneration draw is only important if an individual accepts a job – if the individual is employed.

⁴⁴ See Rust (1992) for a complete discussion of the principles that need to be satisfied in order to use a Finite Markov Bellman model.

$$V_i(\Omega_a, a) = r_{i,a} + \beta E[V(\Omega_{a+1}, a + 1) | \Omega_a, d_a = i] \quad (10)$$

Following this logic, the initial value function can be expanded all the way to the last period where $t = 100$. At this terminal period all errors would have been revealed and the optimal decision under each of the possible states can be recovered. Once we know the optimal set of choices for the state space for the final period we are able to use equation 10 to calculate what the optimal choice would be at $t = 99$. This process is repeated until we arrive at $t = 1$. Using the backward-recursion approach, we are to trace out the optimal non-employed reservation wage as well as the maximum value of γ_S for which someone would be willing to search. Both these values are allowed to vary by education, experience and time.

4.3. The Likelihood Function

Once we have obtained the reservation wage values that determine the optimal decision path for each agent under any set of parameters, we are able to model the likelihood function. The likelihood function will allow us to tell how well the model fits the data. In our model, the likelihood function captures the probability of transitioning between states⁴⁵.

4.3.1. Remaining Non-Employed

The likelihood of remaining non-employed can be modelled as follows.

$$L = (1 - h_j) + h_j \Phi\left(\frac{rw_i}{\sigma_u}\right) \quad \text{where } h_j = \begin{cases} h_H & \text{if } d_{t-1} = 0 \\ h_S & \text{if } d_{t-1} = 1 \end{cases}$$

The first term, $(1 - h_j)$, denotes the probability of not receiving a job offer and the second term denotes the probability of receiving and turning down an offer. h_j , the probability of receiving an offer, is allowed to differ depending on whether someone had searched during the previous period. $\Phi\left(\frac{rw_i}{\sigma_u}\right)$ denotes the probability of rejecting an offer.

4.3.2. Transitioning into Employment

The likelihood of transitioning into employment and earning a wage w_t .

$$L = h_j \left[1 - \Phi\left(\frac{rw_i - \rho \frac{\sigma_u}{\sigma_\theta} \left(\frac{w_t - x_t \beta}{\sigma_\theta}\right)}{\sigma_u \sqrt{1 - \rho^2}}\right) \right] \times \frac{1}{\sigma_\theta} \phi\left(\frac{w_t - x_t \beta}{\sigma_\theta}\right) \quad \text{where } h_j = \begin{cases} h_H & \text{if } d_{t-1} = 0 \\ h_S & \text{if } d_{t-1} = 1 \end{cases}$$

Again, h_j , denotes the probability of receiving a job offer. The term in the square brackets denotes the probability of accepting the offer⁴⁶. The final term denotes the probability of observing the specific wage offer.

⁴⁵ We will be exploiting the panel dimension in the LFS data, utilising data on employment, job search, age, education, tenure and wages.

⁴⁶ The cumulative density function within the square brackets shows that the probability of accepting a job offer is not a function of the wage offer, the reservation wage, the individual's characteristics and the correlation between the errors in the employment equation and the errors in the wage equation.

4.3.3. Transitioning into a New Job

Our model allows for on-the-job search. The likelihood of transitioning from a job that reportedly pays w_{t-1} to one that reportedly pays w_t is

$$L = h_E \left[1 - \Phi \left(\frac{(w_{t-1} - x_t \beta) - \rho \frac{\sigma_u}{\sigma_\theta} \left(\frac{w_t - x_t \beta}{\sigma_\theta} \right)}{\sigma_u \sqrt{1 - \rho^2}} \right) \right] \times \frac{1}{\sigma_\theta} \phi \left(\frac{w_t - x_t \beta}{\sigma_\theta} \right)$$

Employed individuals have a h_E probability of receiving a new job offer. The term in the square brackets denotes the probability of accepting the job offer. However, unlike before, the reservation wage is now a function of the reported current wage (which is observed by the econometrician) and not the expected reservation wage (which was recovered through dynamic programming)⁴⁷.

4.3.4. Keeping the Same Job

The likelihood of retaining a job and transitioning from a reported wage of w_{t-1} to w_t is

$$L = \left(1 - q - h_E \left[1 - \Phi \left(\frac{(w_{t-1} - x_{t-1} \beta) - \rho \frac{\sigma_u}{\sigma_\theta} \left(\frac{w_t - x_t \beta}{\sigma_\theta} \right)}{\sigma_u \sqrt{1 - \rho^2}} \right) \right] \right) \times \frac{1}{\sqrt{\sigma_\varepsilon^2 + \sigma_\xi^2}} \phi \left(\frac{(w_t - x_t \beta) - (w_{t-1} - x_{t-1} \beta)}{\sqrt{\sigma_\varepsilon^2 + \sigma_\xi^2}} \right)$$

The term on the left captures the probability of remaining in the current job⁴⁸. The term on the right captures the likelihood of observing the wage, w_t . Individuals with higher remuneration draws will be more likely to remain in their current position⁴⁹. The difference in the unobserved portion of the wage between the two periods is driven by measurement error.

4.3.5. Transitioning Out of Employment

At any time, any employed individual has probability q of moving out of employment. The probability of moving out of employment is exogenously determined.

4.3.6. Searching for Employment

The likelihood among the non-employed of actively looking for employment is:

$$L = \Phi \left(\frac{rw_i^S - u_S}{\sigma_S} \right)$$

Similar to the reservation wage, the minimum cost at which a non-employed worker would be willing to look for employment is backed out in a recursive manner. The value is denoted as rw_i^S and varies by education, experience and age. Once we have derived rw_i^S the probability of search can be computed through the cumulative distribution function since the search cost is normally distributed.

⁴⁷ Workers who hold jobs with lower remuneration draws will be more likely to transition into other jobs.

⁴⁸ Everyone excluding those who were fired and who accept a new job remain in their old jobs.

⁴⁹ The expected wage will be equal to the previous wage plus 1 period worth of experience.

For each individual, their overall likelihood will be the product of each pair of successive waves we were able to observe.

$$L_i(\theta) = \prod_{t=1}^{T_i} L_{it} \quad (11)$$

In order to simplify the optimization process the likelihood is logged.

4.4. Identification

Unlike cross-sectional models, that are only able to tell us about the level of unemployment, job search models have the potential of telling us whether unemployment is voluntary or not. Theoretically, these models are capable of distinguishing between labour markets with high arrival and rejection rates (high reservation wages) and labour markets with low arrival and rejection rates (low reservation wages). The identification of these models rely on the structural assumptions and the richness of the data⁵⁰.

In our homogenous model we assume that all individuals face the same job offer rates and remuneration draws. The homogenous model has 16 parameters that need to be recovered⁵¹: The parameters of the wage distribution (a_0, a_1, a_2, a_3, a_4), as well as the standard deviation for the remuneration draw and measurement error (σ_u, σ_{me}), the parameters for household production and search cost ($\gamma_0, \gamma_1, \gamma_2, u_s, \sigma_s$) and the parameters that determine the job arrival and exit rates (h_H, h_S, h_E, q). In order to identify the parameters, each of the parameters needs to have a distinctly different effect on the set of moments on which one plans to fit the model. For this reason identification requires a rich set of moments (Paserman 2008).

Unlike the conventional Mincerian model, that ignores non-employed, the job search model assumes that the attributes that determine wages also play a large role in determining employment. In this regard the coefficients in the wage regression are determined in a similar manner to those in the Heckman (1979) selection model⁵². In our model, household production is allowed to vary non-linearly over age. These two additional parameters allow the employment profile and wage profile to deviate from one another.

Wages are allowed to deviate from expected wages. In the absence of panel data we would not have been able to tell whether the variation is driven by the variation in the remuneration draws or by measurement error. Having repeated observations, however, allows us to distinguish between the part of the error in the wage that is persistent and the part that is idiosyncratic. Put simply, a large persistent variation from the

⁵⁰ Flinn and Heckman (1982) showed that in the absence of richer data the identification of job arrival rates, job refusal rates and the reservation wages rests on the distributional assumptions of the untruncated wage offer.

⁵¹ In the heterogeneous model the wage intercept and job offer rates are allowed to differ between the two types. The model requires an additional 3 parameters to be solved.

⁵² The Heckman model allows the coefficient in the second stage (wage equation) to be influenced by the relationship between the same set of characteristics and take up (employment) in the first stage. In the Heckman selection model individuals are considered to be myopic ($\beta = 0$).

expected wage offer would be evidence of a large remuneration draw effect, while a large idiosyncratic variation would be evidence of large measurement error effect⁵³.

The reservation wages and home production values are not directly observable. Job search models rely on a strong set of structural assumptions to identify the reservation wage and home production parameters. These assumptions allow us to use distribution of the observed wages to trace out the reservation wages and the expected utility under home production.

Once the reservation wages have been recovered and the distribution of job offers is known, we are able to use the structural model to estimate the proportion of jobs offers that would have been refused and accepted. The job arrival rate is then chosen so as to let the proportion of non-employed individuals who transition into employment correspond to the proportion in the data. The discount parameter is set to an annual discount rate of 5%^{54, 55}

In four of the six waves in our panel, the non-employed were asked whether they had turned down any job offers during the previous six months. We use this question as a robustness check to see whether the reservation wages that we recover from the distributional assumptions alone are adequate.

In the heterogeneous model we allow the wage intercept (the average offered wage) and job offer rates to differ between the two types. We also allow the proportion of individuals that are estimated to belong to each of the two types to vary. The finite mixture model (FMM) that we adopt requires an additional 3 parameters to be solved. Since we assume that individuals are unable to change their type, we are able to draw on the longitudinal information in our data to investigate whether different individuals face different labour market condition. The FMM allows us to see whether the same set of individuals who receive more job offers also receive higher wage offers.

5. Empirical Analysis

5.1. Data

Data is taken from six consecutive bi-annual Labour Force Surveys; the first was captured in the second half of 2001 and the last was captured in the first half of 2004. The sample is limited to working aged black males who were observed at least once in two consecutive waves. Using the rotating panel within the surveys, 22922 individuals were uniquely identified. The survey contains information on schooling, age, as well as employment status: whether individuals were working, seeking work or inactive. The survey also

⁵³ The average change in wages among those who did not switch jobs contributes further to the identification of the experience parameters (a_2 , a_3), while the variation from the expected trend helps to identify measurement error.

⁵⁴ The choice of the discount rate is by large arbitrary. Moore & Viscusi (1990) found that the estimated discount rates that are derived from labour data ranges between 1% 14%. We chose 5% since it lies comfortably within this range.

⁵⁵ A bi-annual discount rate of 2.53% ($1 - \sqrt{1 - 0.05}$).

contains further information about wages and tenure among those individuals who reported as being employed.

5.2. Descriptive Statistics

The choice between working, staying at home or actively seeking employment is central to our discussion throughout. The following graph shows how the likelihood of being in each state fluctuates over one's working-age.

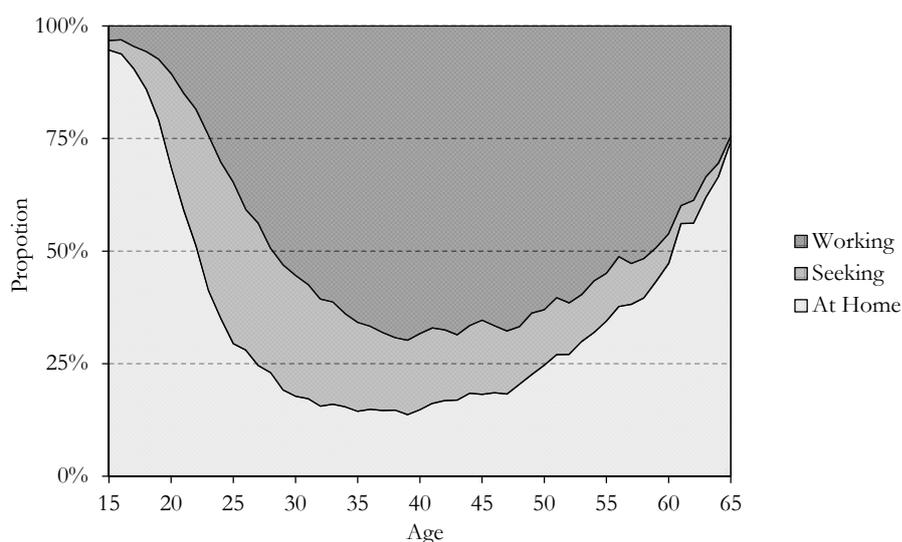


Figure 2.1: Choice Distribution by Age

According to figure 2.1, very few of those individuals who are younger than 21 (and not studying⁵⁶) are actively employed. Most are either inactive or searching for work. The probability of being employed increases steadily by age. The likelihood of being employed if a person is not enrolled is roughly 10% among 20 year olds, and roughly 70% among 40 year olds. Similarly, the likelihood of being at home (non-employed and non-searching) decreases from 70% to 20% over the same age range. The probability of being an active job seeker remains relatively stable, at around 20%, between the ages of 20 and 40. From 40 onwards there is a sharp decline in both the probability of being employed and the probability of seeking employment. It is unclear whether the decrease is driven by early retirement or discouragement.

The likelihood of being in either of the three states also differs by educational attainment.

⁵⁶ Our model ignores all those individuals that are considered to be full-time students. According to figure 3.1 school attendance drops off quickly after the age of 16.

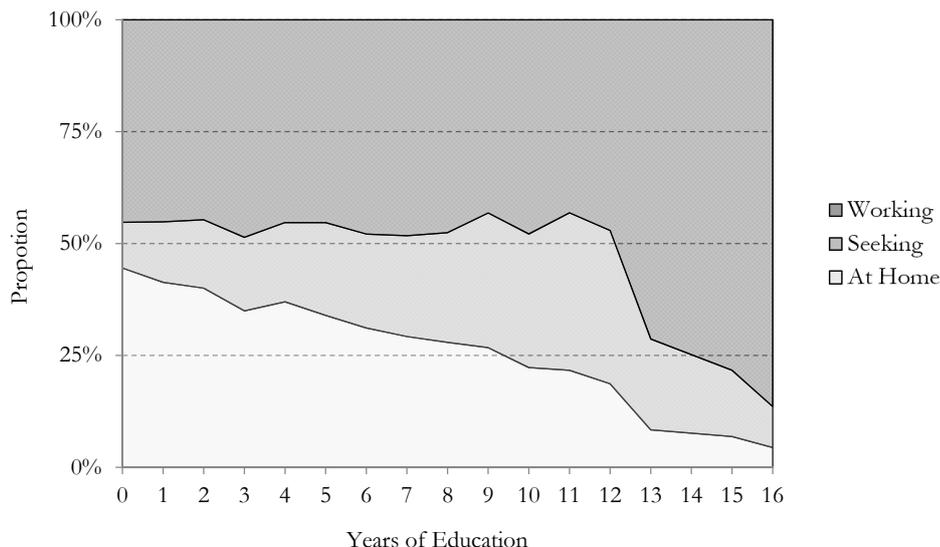


Figure 2.2: Choice Distribution by Education Level

Up until grade 12 education appears to have a very small effect on employability. Beyond matric, the probability of being employed increases, while the probability of being at home or searching dwindles. Interestingly, the probability to be actively searching for employment is highest among those with moderate levels of education.

In our formal model we are interested in the level of churning between the three decision states and the factors that give rise to these changes. Table 2.1 shows the average transition rates between two waves among the three states for the subsample of black males who were observed in two consecutive periods⁵⁷,
58.

Table 2.1: Transition between States

		Future State			Total
		At home	Seeking	Working	
Current State	At home	59% (6065)	24% (2458)	17% (1728)	100% (10251)
	Seeking	25% (1995)	52% (4144)	23% (1885)	100% (8445)
	Working	8% (1582)	9% (1843)	83% (16860)	100% (20285)
	Total	25% (9642)	22% (8445)	53% (20473)	100% (38560)

⁵⁷ Only those individuals for whom we had consecutive observations are included in this table. Some people were not observed in consecutive periods, while others were observed multiple times.

⁵⁸ In chapter 1 we showed that the level of persistence is likely to be understated.

Individuals are most likely to stay in the state they are in. Persistence is highest among the employed: 83% of those who were employed in the previous period remained employed, while 59% of those who were at home were also at home in the next period. Transition is the highest among the actively seeking. An individual that was actively looking for employment 6 months ago has a 52% chance of still searching, a 25% chance of having quit searching but remaining non-employed, and a 23% chance of becoming employed.

Among the non-employed, those who are actively looking for employment are more likely to become employed. The difference is, however, not as large as one would have expected. Those who are actively looking for employment had a 23% chance of being employed the next period, while those who are not looking have a 17% chance of finding work. From these descriptives it appears as though actively seeking work only increases the likelihood of becoming employed within the next 6 months by 6 percentage points or roughly 40%. However, drawing a causal interpretation from these tentative results is problematic since it is unclear whether the group that are searching and the group that are non-searching are composed of comparable individuals. For instance, if the searching are more educated than the non-searching (as suggested in figure 3), this would bias the returns to searching upwards.

In four of the six waves, the non-employed were asked whether they had turned down any job offers during the previous six months. While this provides us with a valuable source of information we are wary that these self-reported measures may be bias. In the case of job offers one would expect that the self-reported estimates are susceptible to underreporting. Like, Levinsohn and Pugath (2010) we are concerned that some of the job offers that people reject might be implicit. For this reason we decided not to incorporate these moment into our model of choice, but to rather include it as a robustness check.

The table below shows the proportion of respondents who report to have declined work within the last 6 months.

Table 2.2: Proportion of Respondents who Declined Work

	Decline	Not Decline
At home	0.8% (13970)	99.2% (277)
Seeking	1.9% (27047)	98.1% (220)
Total	1.2% (41017)	98.8% (497)

The proportion is low. Only 1.2% of black males who remained non-employed reported to have received and turned down a job offer during the previous 6 months. Interestingly, the proportion of individuals who received and declined job offers were roughly two times higher among the group who were actively looking for employment over the last six months, compared to those who were not actively looking for

employment. While it is possible that some of the difference is driven by variation in the rejection rates it is also likely that a large portion of the difference could be due to differences in job arrival rates.

The descriptive results show that individuals who actively look for employment are more likely to transition into employment and also more likely to decline employment relative to those individuals who are not actively looking for employment. Combined, these two measures provide us with a rough indication of what the job arrival rate could be for both groups. Some crude back of the envelope calculations suggest that the job arrival rate for those who actively search for work is around 25% and the job arrival rates for those who do not actively search is around 18%.

5.3. Fitting the Data to the Model

Following a maximum likelihood approach, we set out to find the set of parameters that are most likely to have produced the moments observed in the South African labour market. At each iteration the likelihood function compares the actual probabilities of real world outcomes from the data to the predicted moments from our model, until it finds the set of parameters that are deemed to be most adequate. The optimal set of parameters are derived using the Powell-algorithm and non-linear-simplex-and-simulated-annealing-algorithm. The Powell-algorithm is the more efficient of the two algorithms and does well in finding local maximums, but is sensitive to discontinuities in the likelihood estimate⁵⁹ (Brent, 1973). Conversely, the non-linear-simplex-and-simulate-annealing-algorithm is slower, but more cautious of accidentally converging to a local rather than a global maximum⁶⁰ (Cardoso et al., 1996).

5.4. Homogenous Model

The conventional Mincerian regression results are reported in column 1, while column 2 and 3 contain the set of parameters that best fit the dynamic job search model that we have developed in this paper. Model 2a uses the distributional assumptions alone to fit the reservation wages, while model 2b also fits the self-reported rejection data. In our analysis we will focus on model 2a. All the results contained in table 3 are for a sample of working-aged black males in South Africa.

Table 2.3: Parameter Results for Homogenous Models

	OLS Model (Model 1)	Dynamic Model (Model 2a)	Dynamic Model (Model 2b)
Wage			
Constant	0.499***	0.285***	0.616***
Education	-0.045***	0.050***	0.027***
Education Squared	0.014***	0.006***	0.004***
Experience	0.082***	0.106***	0.079***
Experience Squared	-0.001***	-0.002***	-0.002***
S.D. of Remuneration Draw		1.159***	1.048***
S.D. of Measurement Error		0.780***	0.777***

⁵⁹ The matlab file was written by Argimiro R. Secchi and can be downloaded from:

<http://www.mathworks.com/matlabcentral/fileexchange/15072-unconstrained-optimization-using-powell/content/powell.m>

⁶⁰ Repeated testing, by the author, has shown that this method is less sensitive to the choice of initial values.

Household Production			
Constant		2.241***	1.940***
Age		-0.117***	-0.120***
Age Squared		0.014***	0.015***
Search Cost		0.719***	0.891***
S.D. of Search Cost		0.309***	0.311***
Offer Probability			
Default Arrival Rate		0.149***	0.122***
Searching Multiplier		1.785***	1.950***
Employed Multiplier		0.646***	0.754***
Job Exit Rate		0.180***	0.182***
N	18823	22922	22922
R²	0.314		
Log Likelihood		-63029.4	-63913.1

Note: * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level

5.4.1. Wages

All prices were set to 2010 values. The average hourly wage among the subsample of black males in our sample who were employed was R2.57, which translates to a median monthly income of roughly R3000. Throughout we allow the expected wage offer to be influenced by education and experience. Education was introduced quadratically, since the returns to schooling are believed to be convex in South Africa (Keswell & Poswell, 2004). The returns to education in the structural model are slightly lower and less convex than in the OLS model (see figure 2.A.2 in the appendix). The expected wage-experience-profile follows the conventional inverse U-relationship (see figure 2.A.3 in the appendix). The expected wage profiles follows a similar shape to that found by Imai and Keane (2004) for the US, who used structural models to evaluate the role of human capital accumulation over the life-cycle path.

5.4.2. Household Utility and Reservation wages

The reservation wage represents the minimum logged hourly rate at which someone will accept employment, while the household utility denotes the utility attached to the outside option. Neither reservation wages nor home utility are directly observed. Instead, in our model, we attempt to uncover what the reservation wages and household utility could be, based on labour market behaviour.

In the figure below we show how the estimated utility under home production and the reservation wage evolve over the life-cycle⁶¹.

⁶¹ The curve was constructed for a group of black males with 8 years of education. The average level of education for this sample is 8.2 years.

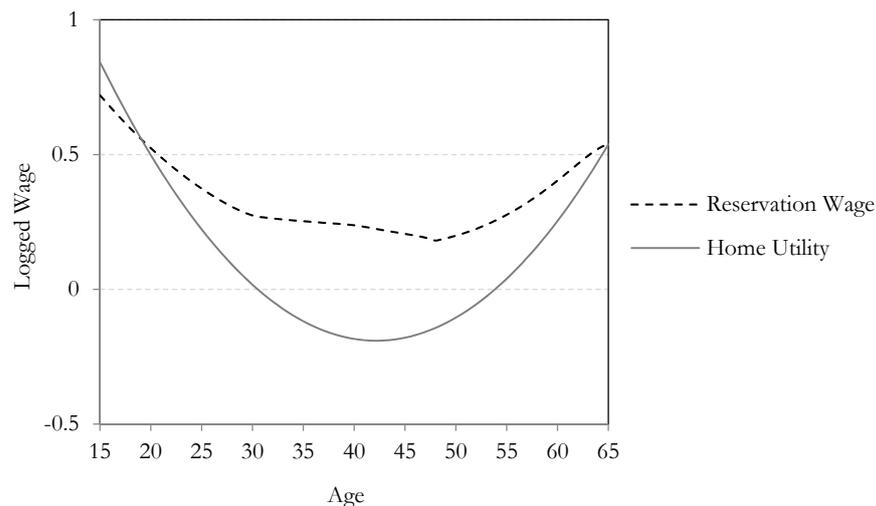


Figure 2.3: Home production utility and the reservation wage over age for an individual with 8 years of education.

The reservation wage follows a similar path as the home utility curve. This should not be surprising. Under a myopic models the reservation wage and the outside options are considered to be the same thing. In our model, unlike conventional models, agents are allowed to be forward-looking. Therefore, agents in our model may turn down job offers even if the wages exceed their current household utility.

Figure 2.3 shows that the largest deviation between the reservation wage and home utility occurs where the expected wage is at its highest round age 40. Conversely, we find that as individuals get closer to the end of their working life (and run out of additional periods over which to optimize), their willingness to hold out for higher wages diminishes and the reservation wage converges to the utility under home production. For the final period, when there are no further periods over which to optimize, the employment choice simplifies to a myopic decision, since individuals only have to compare their offered wage for that period to their expected utility under home production.

Like Kiefer and Neumann (1979) we also recovered a U-shaped reservation wage curve. The utility under home production starts high, reaches a minimum at age 42 and rises thereafter. The logged hourly reservation wage for the non-employed ranges between 0.105 and 0.992. The equivalent monthly values range between R215 and R522, respectively⁶². The reservation wage is lowest for 44 year old individuals with 0 years of education and highest for 25 year old individuals with 16 years of education and no experience. These estimated values are far smaller than previous estimates have suggested⁶³.

⁶² Transformed: $e^{0.105} \times 4.3 \times 45 = R215$; $e^{0.992} \times 4.3 \times 45 = R522$

⁶³ Using the National Income and Dynamics Study (NIDS) data, we derive a median self-reported reservation wage for a similarly aged group of black males at R2000.

The logged hourly utility to home production ranges between -0.191 and 0.806. This equates to a monthly value of roughly R160 and R435, respectively⁶⁴.

Interestingly, Rankin and Roberts (2011), who used self-reported reservation wages for a group of young black individuals in South Africa, also uncovered a negative relationship between age and reservation wages between the ages of 20 and 30. According to them, young people start off with inflated expectations about their own earnings potential. As individuals update their expectations about their own ability, their reservation wages come down to more reasonable levels. In our model individuals are assumed to be aware of their earnings potential. Instead, in our model, the drop in reservation wages is driven by a decrease in the utility under home production (or a decrease in the distaste for working) over that age range. One would expect that the pressure on young males to leave the household or to find a job and to contribute to the income pool should increase as they grow older.

5.4.3. Job Arrivals and Refusals

The job arrival rate is allowed to differ depending on whether an individual was non-employed and did not search, was non-employed and searched or was employed. The job acceptance rate, on the other hand, depends on the reservation wage, which in turn depends on an individual's age and level of education.

The figure below shows how the job offer rate and average acceptance rates differ depending on an individual's previous state⁶⁵.

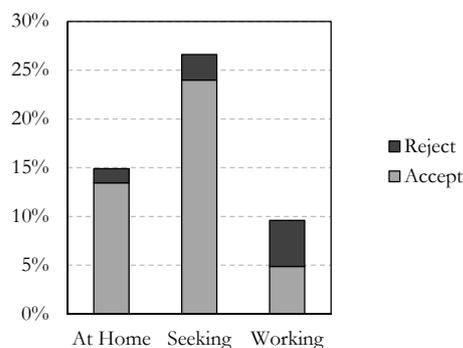


Figure 2.4: Job arrival and acceptance rate by state

Individuals who were non-employed and did not actively look for work have a 15% chance of being offered employment over a six month period, while those who did look for work had a 27% chance of being offered employment. Active search increases the probability of obtaining a job offer by 78% (13 percentage points).

⁶⁴ Transformed: $e^{-0.191} \times 4.3 \times 45 = R160$; $e^{0.806} \times 4.3 \times 45 = R435$ (work 45 hours in a week and 4.3 weeks in a month)

⁶⁵ The rejection rates vary depending on age and education. The results in figure 2.4 represent the average rejection rates for a simulated sample of black males with a similar educational distribution to our sample.

The search cost follows a log-normal distribution a median monthly rand value of R465 per month⁶⁶. On average, only 10% of the non-employed in the model who were offered work turned it down. The degree of selection is relatively low. The sub-sample of numeration draws that were accepted (by the non-employed) is therefore not much different from the distribution of all the numeration draws.

The model also allows for on-the-job search. 10% of employed individuals are estimated to receive a job offer each period. Interestingly, the probability of receiving a job offer is lower when employed than when non-employed, regardless of search. Our simulated (using the parameters that we obtained from model 2a) results showed that slightly more than half of the on-the-job job-offers that were received were declined. This makes sense. The employed will only change jobs if they receive a better job offer than their current one. If there were no selection into employment and individuals were only allowed to change jobs once we would have expected the acceptance rate to be 50%, but since there is a minor degree of positive selection into employment we expect the current distribution of wages to be slightly higher than the average wage draw and the acceptance rate to be slightly lower than 50%. The estimated reservation wages as well as the estimated job arrival rates and rejection rates are lower in our model than in the other South Africa studies (Kingdon & Knight, 2001; Rankin & Roberts, 2011).

The model predicts an exit rate of roughly 18%.

5.5. Heterogeneous Model

The baseline model does not account for unobserved heterogeneity. Unobserved heterogeneity can arise in several places – through the wage distributions, through the job arrival or exit rates, or even through the utility attached to the outside option or through the disutility attached to active search. According to Wolpin (1987), the choice as to where to introduce heterogeneity is a largely discretionary, yet necessary, decision, since the model would not be identifiable if all the parameters are allowed to vary by type.

A finite mixture model (FMM) is constructed. In our model we allow for two types of workers, type 1 and type 2. The two types differ with respect to the arrival rate of their wage offer as well as their wage distribution. We loosely refer to these unobserved characteristics that may shift the wage distribution and arrival rate as ability, but it could easily be seen as school quality, motivation, or anything else the model does not currently account for.

In identifying type, we will examine whether some one group of individuals receive job offers at a more frequent underlying rate than another group of individuals and whether one group receives higher remuneration draws than the other group of individuals (after already controlling for education and experience). Having repeated observations for the same individuals allow us to exploit the longitudinal

⁶⁶ The median is reported rather than the average cost, since the term follows a log-normal distribution.

dimension of our data. For the sake of our model, we will be assuming that individuals are unable to change type – they remain of the same type throughout the survey⁶⁷.

Two finite mixture models were estimated. Model 3a uses the distributional assumptions alone to fit the reservation wages, while model 3b also fits on the self-reported rejection data. Both models improve their fit when we allow for a second type, supporting the claim that there might be some unobserved heterogeneity in the model. In our analysis we will focus on model 3a, the first of the two heterogeneous models.

Table 2.4: Results for Finite Mixture Models

	Heterogeneous Model (Model 3a)	Heterogeneous Model (Model 3b)
Wage		
Constant (Type 1)	0.083***	0.067***
Constant (Type 2)	-0.813***	0.249***
Education	-0.019***	0.003***
Education Squared	0.009***	0.006***
Experience	0.149***	0.096***
Experience Squared	-0.003***	-0.003***
S.D. of Remuneration Draw	1.109***	1.023***
S.D. of Measurement Error	0.780***	0.778***
Costs		
Constant	2.148***	1.419***
Age	-0.117***	-0.095***
Age Squared	0.001***	0.001***
Cost of Searching	0.223***	0.248***
Cost of Searching Shock	0.097	0.075
Arrival Rate of Job Offers		
Default Arrival Rate (Type 1)	0.257**	0.200***
Default Arrival Rate (Type 2)	0.134***	0.111***
Searching Multiplier	1.230***	1.242***
Employed Multiplier	0.501***	0.618***
Job Exit Rate	0.179***	0.181***
Proportion	0.490***	0.446***
N	22922	22922
Log Likelihood	-62553.8	-63490.7

Note: * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level

Roughly half of individuals in our finite mixture model are considered to be of type 1 and roughly half are considered to be of type 2. Type 1 individuals face more favourable labour market prospects. Type 1 individuals are two times more likely to receive a job offer than type 2 individuals and receive job offers they do receive are, on average, higher than the offers that type 2 individuals receive. Conceptually, we could imagine that type 1 individuals possess more desirable unobservable traits than type 2 individuals, which may make them more productive employees.

⁶⁷ The relative advantage to searching and being employed relative to being non-employed and non-searching is assumed to be constant across types.

The graph below compares the expected wages, household utility and reservation wages for the two types.

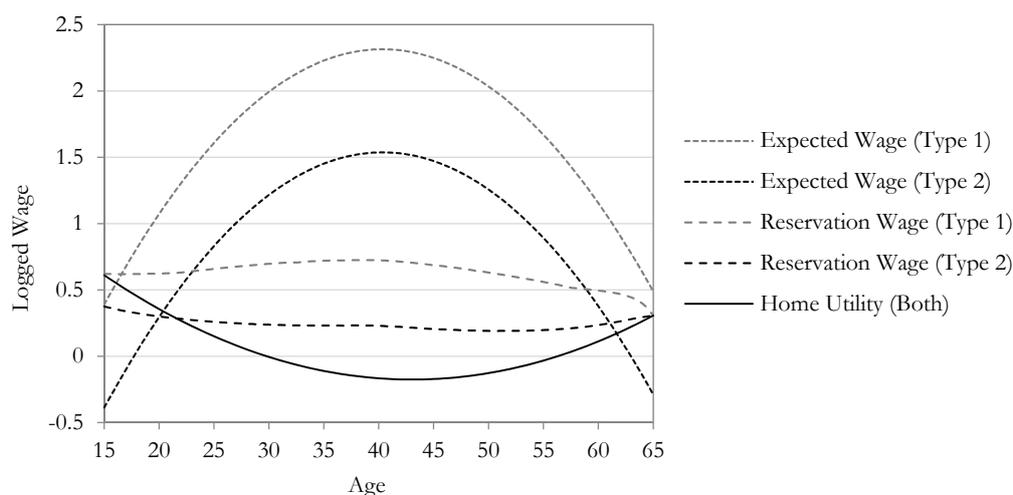


Figure 2.5: The expected wages, household utility and reservation wages by type. The graph was drawn for two individuals with 8 years of education.

Figure 2.5 shows the gap between the expected wage offers for type 1 and type 2 individuals. Despite having the same household utility, type 1 and type 2 individuals face very different reservation wages. Type 1 individuals are more inclined to turn down low job offers than type 2 individuals, since type 1 individuals know that they face a more favourable job offer rate and wage draw than type 2 individuals. The probability of receiving a more lucrative job offer in the near future is therefore far higher for this group⁶⁸.

Interestingly, we see that the benefit and cost to search decreases when we allow for heterogeneity. In model 3a, active search increases the probability of receiving a job offer by 23%. This is far less than what was estimated under the homogenous model (see model 2a), where searching increases the probability of receiving a job offer almost 80%. The expected median monthly disutility from searching decreased from roughly R400 to R240 when we allowed for unobservable heterogeneity.

According to our results much of the increase in the transitioning out of non-employment that we perceive to be through active search is actually through positive selection. Type 1 individuals are more likely to self-select into active search, since for them the relative gains to search are larger. While the percentage change in the job offer rate is the same for both types, the increase is from a larger base for type 1 individuals. Searching increases the probability of a job offer by 6 percentage points for type 1 individuals and 3 percentage points for type 2 individuals. The gains of receiving a job offer are also higher for type 1 individuals, since their wage offer distribution is higher than for type 2 individuals.

⁶⁸ Individuals are assumed to be aware of their type and able to adjust their reservation wages accordingly,

6. Conclusion

A structural job search model was developed and estimated. The structural model relies on a rich set of assumptions to help explain how utility maximizing working-aged agents may operate in the face of uncertainty. Parameters were recovered for the wage coefficients, the distribution of measurement error, the remuneration draw, the outside option and the cost of search, and the job arrival and exit rates.

Reservations wages were backed out of the model and allowed to vary by education and over one's working life. Unlike other authors we rely on the behaviour of labour market participants rather than on their self-reported responses to recover the reservation wage. The median reservation wages for non-employed black males are estimated at roughly R400 per month when we do not fit on the reported rejection rate and at roughly R300 when we do. These reservation wage values are far lower than been previously reported. Our results reconfirm the earlier suspicion of Kingdon and Knight (2001) and Rankin and Roberts (2011) that reservation wages are over-reported. Like Rankin and Roberts (2011) we also uncovered a negative relationship between age and reservation wages between the ages of 20 and 30, despite using a very different technique. Unlike Rankin and Roberts (2011) our model assumes that individuals are aware of their expected earnings potential and that the decline in reservation wages is not due to a revision of these expectations. Instead, in our model, the drop in reservation wages is driven by a decrease in the utility under home production (or a decrease in the distaste for working) over that age range. This decrease in the utility of home production with age for young males is likely to be a function of their bargaining position within the household. One would expect that the pressure on young males to find a job and to contribute to the income pool should increase as they grow older.

Our model suggests that much of the increase in job arrival rates that we see from searching is through selection. Actively looking for work has a much smaller effect on employment than the descriptive results would suggest. The low proportion of job refusals suggests that the main constraint to employment is not reservation wages, but rather a shortage of job offers.

Appendix 2

Appendix 2.A: Additional Figures and Tables

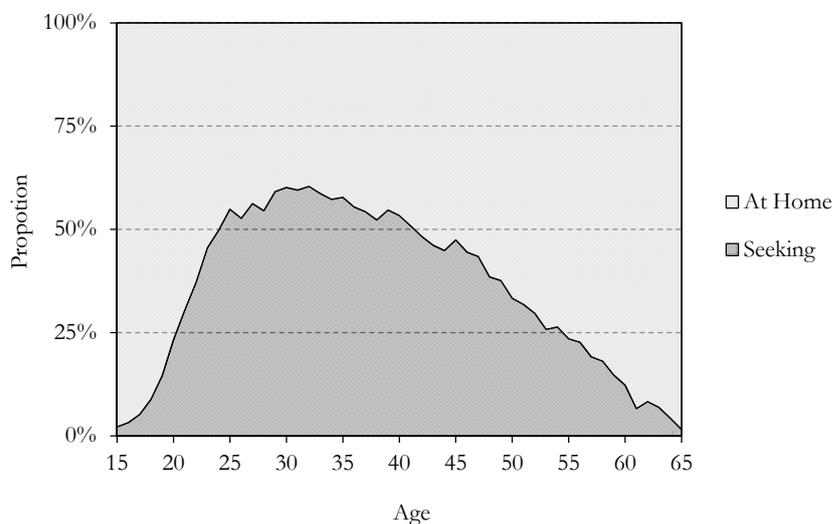


Figure 2.A.1: Decision to Search by Age

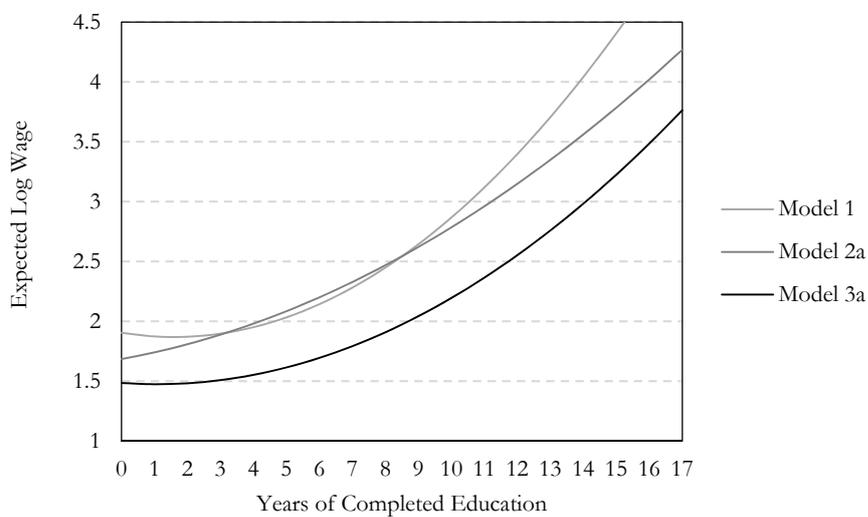


Figure 2.A.2: Returns to education profile.

Note: The graph represents the expected logged hourly wage for someone who is aged 28 and who is not aware of their type.

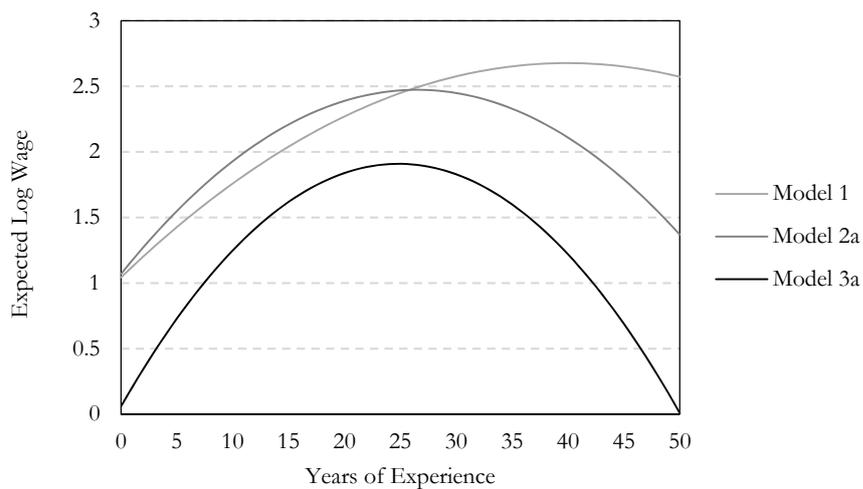


Figure 2.A.3: Returns to experience profile.

Note: The graph represents the expected logged hourly wage for someone who has 8 years of education and who is not aware of their type.

Table 2.A.1: Comparing the parameters obtained from the Static and Dynamic Models

	OLS Model (Model 1)	Homogenous Model (Model 2a)	Homogenous Model (Model 2b)
Wage			
Constant	0.499 (0.037)	0.285 (0.033)	0.616 (0.030)
Education	-0.045 (0.006)	0.050 (0.003)	0.027 (0.003)
Education Squared	0.014 (0.000)	0.006 (0.000)	0.004 (0.000)
Experience	0.082 (0.002)	0.106 (0.000)	0.079 (0.000)
Experience Squared	-0.001 (0.000)	-0.002 (0.000)	-0.002 (0.000)
S.D. of Remuneration Draw		1.159 (0.030)	1.048 (0.022)
S.D. of Measurement Error		0.780 (0.003)	0.777 (0.003)
Costs			
Constant		2.241 (0.044)	1.940 (0.047)
Age		-0.117 (0.001)	-0.120 (0.001)
Age Squared ($\times 100$)		0.014 (0.000)	0.015 (0.000)
Age Quadratic ($\times 10000$)			
Age Quartic ($\times 1000000$)			
Cost of Searching		0.719 (0.025)	0.891 (0.025)
Cost of Searching Shock		0.309 (0.016)	0.311 (0.013)
Arrival Rate of Job Offers			
Arrival Rate		0.149 (0.006)	0.122 (0.003)
Searching Multiplier		1.785 (0.043)	1.950 (0.039)
Employed Multiplier		0.646 (0.033)	0.754 (0.035)
Job Exit		0.180 (0.003)	0.182 (0.003)
N	18823	22922	22922
R²	0.314		
Log Likelihood		-63029.4	-63913.1

Table 2.A.2: Comparing the parameters obtained from the Static and Dynamic Models

	Heterogeneous Model (Model 3a)	Heterogeneous Model (Model 3b)
Wage		
Constant (Type 1)	0.083 (0.017)	0.067 (0.022)
Constant (Type 2)	-0.813 (0.188)	0.249 (0.005)
Education	-0.019 (0.008)	0.003 (0.001)
Education Squared	0.009 (0.001)	0.006 (0.000)
Experience	0.149 (0.028)	0.096 (0.003)
Experience Squared	-0.003 (0.005)	-0.003 (0.001)
S.D. of Remuneration Draw	1.109 (0.038)	1.023 (0.016)
S.D. of Measurement Error	0.780 (0.033)	0.778 (0.007)
Costs		
Constant	2.148 (0.041)	1.419 (0.068)
Age	-0.117 (0.002)	-0.095 (0.001)
Age Squared	0.001 (0.000)	0.001 (0.000)
Cost of Searching	0.223 (0.017)	0.248 (0.014)
Cost of Searching Shock	0.097 (0.194)	0.075 (0.158)
Arrival Rate of Job Offers		
Arrival Rate (Type 1)	0.257 (0.116)	0.200 (0.045)
Arrival Rate (Type 2)	0.134 (0.019)	0.111 (0.009)
Searching Multiplier	1.230 (0.213)	1.242 (0.179)
Employed Multiplier	0.501 (0.017)	0.618 (0.008)
Job Exit	0.179 (0.013)	0.181 (0.008)
Proportion	0.490 (0.019)	0.446 (0.016)
N	22922	22922
R²		
Log Likelihood	-62553.8	-63490.7

Chapter 3:

The Role of Unobserved Ability on Educational Attainment and Labour Market Outcomes

Educational attainment is not random. Individuals who stay in school longer and attain more years of education are different from those individuals who drop out earlier. If these attributes are correlated to labour market outcomes, they could bias our estimates on the returns to education.

While we are able to control for the selection bias on observables we are unable to control for the portion of the bias that is introduced through the selection on unobservable individual heterogeneity.

One way in dealing with the endogeneity of education is through the experimentalist approach which attempts to introduce an exogenous source of variation into the model. In our model we do the opposite. Instead of eliminating the endogenous portion of the variation in education that is correlated with the unobservables, we attempt to estimate the extent of the bias directly by explicitly modelling the educational process that gave rise to the bias.

1. Introduction

Few South Africans finish high school and even fewer continue on to tertiary education. The crowding at low levels of education is somewhat surprising since the returns to education in South Africa are believed to be high and convex, while the cost of schooling is low. The convexity in the returns to education suggests that the gains to an additional year of schooling is increasing – every additional year of schooling increases one's expected wage by more than the year that preceded it. Following this logic, in the absence of any constraints and assuming the costs are constant, it would be illogical for a student who finished the previous year of schooling to not further their studies.

While some individuals may be unaware of the high returns to education, it is doubtful that there would be a systematic tendency to undervalue the returns to schooling, and more specifically to undervalue the returns at the top end relative to the bottom. Instead, it is commonly acknowledged that these estimates may be susceptible to selection bias. If more capable, more productive, more driven individuals select into higher levels of education, it will bias conventional OLS estimates of education returns upwards.

Theoretically, the instrumental variable approach is capable of recovering the true returns to education, by introducing an exogenous source of variation into the model. This approach, however, requires one to have access to a valid instrument that is correlated to education, but not to the unobserved correlates that could bias education. Ironically, most of the instrumental variable models that have attempted to correct for ability bias have recovered estimates with higher, rather than lower, returns to education (Card, 1999).

Instead we develop a dynamic programming model that mimics the schooling decision faced by young black males in South Africa. Educational attainment is modelled as a sequential process, where utility maximizing agents have to weigh up the cost and benefits of remaining enrolled under uncertainty. Agents do so by comparing the discounted expected future utility of both choices. Unobserved ability is introduced through a finite mixture model, where we allow individuals to have different schooling and labour market abilities. In our paper we will assume that individuals are aware of the distribution of ability in the population and how ability affects schooling and labour outcomes. They are assumed to know the returns to ability, but are not fully aware of their own ability type. Similar to econometricians, agents learn about their ability as they progress through the schooling system, updating their expectations of their ability at the end of each year.

In the following section we look at the South African literature on the returns to education, and more specifically the attempts that have been made to address the endogeneity that may be introduced through ability bias. We also look at international attempts at using structural modelling to correct for the endogeneity that is induced through ability bias. In section 3 we formulate the sequential schooling model, by introducing the behavioural assumptions that guide the action of the optimising student as they progress through the schooling system. In section 4 we show how optimising individuals in our model use the limited information about their expected probability of passing and their expected gains to passing to inform their decision to remain enrolled. In section 5 we introduce the data, show some descriptive results, before continuing to solve our parameters. In section 6 we compare the schooling and labour market prospects of high and low ability students under the current schooling system and show how these outcomes would be affected if there was less noise in the distinction between who fails and passes in black schools. Section 7 concludes.

2. Previous Work

2.1. Ability Bias

The Mincer model compares wages across age profiles and different levels of schooling (Mincer, 1974). The model does well in capturing the underlying correlation between the covariates. Inferring causality, however, requires that individuals be identical on all attributes apart from education. This assumption – that individuals with different levels of education share the same unobservables – is commonly referred to as the exogeneity, orthogonality or homogeneity assumption.

The validity of the exogeneity assumption has been questioned from two sides. Conceptually, Rosen (1977) argues that if the identifying assumption is valid and all individuals are truly homogenous, then there would be no reason for individuals to not all have the same level of schooling. Unlike Rosen, Becker (1967) believes that people are inherently different and that the optimal level of capital investment will differ by individual. He maintains that the human capital investment would be lower for students with lower innate ability and with limited access to funding (since these individuals are forced to draw on more expensive sources of financing). Self-selection into education on such unobserved attributes will bias the returns to education, since we are not able to simply compare the average wages across different educational levels any more. Despite these reservations, most labour economists have persisted with the Mincerian model.

Most of the classic models, like the Mincerian model, assume that the schooling decision is exogenous. Conventional wisdom, however, suggests that individuals with greater unobservable labour market attributes may select into higher levels of education. In the presence of such heterogeneity, simple comparisons or multivariate regressions will deliver biased results. Econometricians cannot just compare wage and employment outcomes between those with higher and lower levels of education, since those who select into higher and lower levels may also differ with regard to other attributes that could impact labour outcomes. Often in the literature, researchers try to circumvent these endogeneity concerns about ability by adding a cognitive measure of ability to the set of controls. Schultz (1988) found that controlling ability reduced the returns to education estimate, but only by about 10% in the US. Unfortunately, even with the most elaborate set of controls, there are still likely to be some unobservable attributes that could impact both education and labour outcomes.

The discussion about the extent of ability bias within the returns to education is analogous to the more general discussion about whether education actually enhances productivity or whether it merely serves as a signal of innate ability. According to the human capital theory education helps to increase an individual capabilities and skills. Proponents of this theory therefore believe that there is a causal link between an education, increased productivity and increased returns in the labour market.

According to the sorting hypothesis the primary objective of education is not to accumulate human capital, but rather to signal to potential workers that an individual possesses desired unobserved characteristics. Generally it is believed that more abled individuals self-select into higher levels of education. In which case, these desirable unobserved attributes are positively correlated to education. In an imperfect market, where employers are unable to readily observe all the productive characteristics of potential employees, employers are able to infer productivity from education. (Arrow, 1973; Stiglitz, 1975).

Both the human capital theory and the sorting hypothesis predict that there would be a positive correlation between education and labour market outcomes – that the returns to education would be positive. In the absence of richer data, these two theories deliver observationally equivalent outcomes. Both Psacharopoulos (1979) and Wolpin (1977) have devised methods to distinguish between the validity of

these two theories. Psacharopoulos (1979) compares the returns to education between the public and private sector, while Wolpin (1977) compares the returns among the employed and self-employed. The key identifying assumption in both these methods are that sorting takes place in one group and not the other. Applying these methods to South African data, Koch and Ntege (2006) found that education appears to contribute to earnings through both the proposed channels. Education increases productivity and serves as a signal.

2.2. Instrumental Variable Approach

The instrumental variable approach provides a potential solution to the problem of causal inference, one that avoids the necessity to control on all possible unobservables. Theoretically, the instrumental variable approach allows one to recover the true returns to education, by introducing an exogenous source of variation into the model. The IV method, however, hinges on the existence of a valid instrument – a covariate that affects schooling but is not correlated with any of the unobservable attributes that may affect labour outcomes. Much of the focus in the recent literature has been in finding eccentric policy rules that can act as “natural experiments”. Somewhat surprisingly, studies like that of Ashenfelter and Krueger (1994) found estimates that were 30% or more higher than the OLS estimates. This is at odds with what one would have expected, since it implies that the original models might have been biased downwards rather than upwards.

Hypothetically it is possible that the increase in the returns to education under the IV approach could be due to heterogeneous marginal effects or measurement error corrections. Returns to education would increase rather than decrease if the group that is induced into schooling by the instrument have higher returns than the rest of the population. Similarly, the instrument may have neutralised measurement error, which was biasing the initial results downwards. A more feasible explanation for the counterintuitive results may be that the instruments were not valid, in which case the IV method may be introducing ability bias into the estimates rather than removing it. (Card, 1999).

2.3. Structural Models

Keane and Wolpin (1997) use a discrete choice dynamic programming (DCDP) model to explain schooling and employment decisions in the US by tracking the labour market outcomes of a sub-sample of young males from the 1979 cohort of the National Longitudinal Study. While they do not technically model the schooling for earlier grades (it only follows students from age 16 onwards), it does allow for differences in initial schooling endowments. Unlike any previous studies they allow the returns to schooling to differ by occupation and the choice of schooling and occupation to be endogenous – determined by the optimizing individual.

Belzil and Hansen (2002) use a dynamic programming to determine whether the return to schooling in United States are convex or concave. The authors adopt a finite mixture model which allow for 6 types of

individuals. In their model they introduce unobserved individual-level heterogeneity by letting the schooling ability and market ability vary by type. Their results shows that the correlation between unobserved market and schooling ability is positive and large - the correlation between the ability in school and the ability in the labour market is estimated to be 0.95. This finding is consistent with the classical assumption of ability bias, which is commonly believed to inflate the returns to schooling estimates.

Attanasio et al. (2012) use the data from a randomized conditional cash transfer policy named PROGRESA to investigate the impact of conditional cash transfers on school enrolment among the poor in Mexico. The authors use a structural model, to interpret the impact of the program. They conclude that the conditional cash transfers were more successful in improving high school enrolment, than improving primary school enrolment.

Magnac and Thesmar (2002) look at the change in schooling attainment in France between 1980 and 1993⁶⁹. The authors argue that the increase in educational attainment could either be due to increased returns in education or increase probability of employment, reduced cost of education (both objective or subjective) or a reduction in the probability of failing (an increase in the success ration). Using survey data that has information on educational history and on employment, the authors estimate what the optimal educational level would be using a structural model that uses an optimal stopping rule. Interestingly, Magnac and Thesmar (2002) find that the largest determinate of the increase in educational attainment in France during the period under study was due to the decrease in the relative difficulty of attaining education.

2.4. South African Literature

Most of the South African literature on returns to education is descriptive in nature, adopting either ordinary least squares or quantile regressions to estimate the return to education. The prevailing consensus is that an additional year of schooling increases one's earnings at a varying rate. The returns to education are low initially, moderate during high school and high from grade 11 onwards. Mwabu and Schultz (1996) estimate the returns to primary schooling at 8%, the returns to secondary schooling at 16% and the returns to tertiary schooling at 24%. Increasing returns to education, have been found in multiple datasets and studies (Moll, 1996; Mwabu & Schultz, 1996; Keswell & Poswell, 2004; Anderson et al. 2001).

Only three authors have attempted to correct for the possible ability bias in South African education estimates. Moll (1998) finds that including a proxy for cognitive ability does not dramatically change the returns to education estimate or the convexity of the original model. Hertz (2003) corrects for some of the omitted variable bias by running a within-family fixed effects model. Unlike Moll (1998), Hertz (2003) finds that the OLS results were inflated. The estimated returns to education falls from 12%, as was measured using the ordinary least squared estimate, to roughly 5% when he only allows for within-family variation.

⁶⁹ As discussed by Belzil (2006).

Similarly, Burger & van der Berg (2011) used a quasi-structural model to show how the differences in schooling performance may impact educational returns. The authors demonstrate that a large portion of the differences in the returns to earnings and the returns to education between race groups can be explained by the differences in expected matric performances between race groups, which in turn, is a function of schooling quality and latent ability.

Interestingly, the non-linearity in the returns to education has strong implications on the educational choice, since it affects the option value of an additional year of schooling. Burger and Teal (2013) show that forward looking students would be willing to endure lower initial returns to schooling if it provides them with the option of accessing higher returns at higher levels of education. The authors show that under convex returns the choice to remain enrolled does not only revolve around the gains of the next year, but also the possible gain of the years to follow.

Lam et al. (2011) found that a self-administered numeracy score was a strong predictor of passing grade 12 exam for all races, but only a strong predictor of passing lower grades (grade 8 to 11) for white and coloured students. The numeracy test was only a weak predictor of passing lower grades for black learners. They conclude that there is a larger stochastic component in black schools.

3. Dynamic School Participation Model

3.1. Model Structure

We use a discrete choice dynamic programming model to mimic the participation decisions faced by school-going individuals as they proceed through school.

Individuals enter the model at age 6^{70} , the age at which children are supposed to start schooling. Each year students have the option of remaining enrolled or dropping out⁷¹. For the sake of simplicity we assume that once individuals leave the education system they do so permanently and are unable to re-enter the schooling system thereafter. Similarly, we will also assume that students can only study up until the age of 30. After the age of 30 they get forced into the labour market. This leaves each individual with a maximum of 24 discrete decision periods over which to choose the optimal action that would maximize their discounted net present lifetime utility.

Lifetime utility, denoted as U_i , is the discounted additive sum of the utility one derives over one's lifetime. Individuals are assumed to work up to the age of T_i .

⁷⁰ This assumption is not critical, since we do not fit on age in our model.

⁷¹ This school choice could also be taken by parents. As long as the parents act in the best interest of their children the model will remain unaffected.

$$U_i = \sum_{t=1}^{T_i-6} \beta^{t-1} u_{it} \quad (1)$$

In the above equation u_{it} denotes the instantaneous utility for individual i at period t and β denotes the discounting factor – how much a person weighs the future relative to the present. β is assumed to be fixed not only over time but also across individuals.

In each period, each school-going individual needs to decide whether to continue their studies or not. The choice set, d_{it} , that these individuals face is denoted as follows:

$$d_{it} = \begin{cases} 0 & \text{study} \\ 1 & \text{drop out} \end{cases}$$

The two choices are mutually exclusive; an individual can only pursue one of the two options in any one period. In our model we will assume that everyone who is enrolled is enrolled full-time⁷².

The instantaneous utility, u_{it} , at any specific period will depend on whether an individual is enrolled or not. We denote the instantaneous utility while still in school as u_s and the instantaneous utility when someone is out of school as u_w .

$$u_{it} = I(d_{it} = 0)u_s + I(d_{it} = 1)u_w \quad (2)$$

The discounted lifetime utility can be obtained by substituting equation (2) into equation (1):

$$U_i = \sum_{t=1}^{T_i-6} \beta^{t-1} \{I(d_{it} = 0)u_s + I(d_{it} = 1)u_w\} \quad (3)$$

We will discuss the relative rewards under each of the alternatives in turn.

3.2. The Utility of Work and the Returns to Schooling

The actual utility derived outside of school will depend on whether someone finds employment and on what they earn if they do. Both the employment probability and the expected wage are likely to be affected by schooling, experience and age. To keep the model as simple as possible we will assume that education, experience and ability affect both these labour market outcomes in a similar manner.

We define m_{it}^* , as an individual's latent labour market productivity. Both wages and employment are assumed to be a function of m_{it}^* . We allow the variable to be affected by the three determinants through the following relationship.

⁷² This is different from Eckstein and Wolpin (1999). In their paper, students are allowed to do work part-time. They find that financially constrained students are able to alleviate the high financial burden of tertiary education by working part-time and during holidays.

$$m_{it}^* = \alpha_0 + \alpha_1 educ_{it} + \alpha_2 educ_{it}^2 + \alpha_3 exp_{it} + \alpha_4 exp_{it}^2 + v_i^m \quad (4)$$

Here, α_1 and α_2 captures the effect of schooling and α_3 and α_4 the effect of experience⁷³. The ability draw, v_i^m , captures the unobserved heterogeneity in skills endowment. The ability draw, v_i^m , is assumed to be person specific, but time-invariant. The ability draw follows a standard normal distribution.

Both wages and employment are assumed to be a function of m_{it}^* . The following two equations describe the logged wage rate, w_{it} , and the employment probit function⁷⁴:

$$w_{it} = m_{it}^* + \varepsilon_{it}^w \quad (5)$$

$$empl_{it} = 1(m_{it}^* + \varepsilon_{it}^E \geq \tau) \quad (6)$$

Here, ε_{it}^E captures the idiosyncratic productivity shock in employment, while ε_{it}^w captures the idiosyncratic productivity shock in wages. The error terms are assumed to be i.i.d. normally distributed with mean zero and standard deviation of $\sigma_{\varepsilon w}$ and $\sigma_{\varepsilon E}$, respectively. The threshold parameter τ allows employment to deviate from 50%.

The expected instantaneous utility for someone who has joined the labour market can now be formulated in terms of the probability of being employed and the subsequent expected payoffs and the probability and utility under non-employment.

$$E(u_w) = P(Empl_{it})E(w_{it}) + (1 - P(Empl_{it}))u_h \quad (7)$$

Agents are assumed to have rational expectations about their future labour market prospects⁷⁵. Like Attanasio et al. (2012), we will be making use of a terminal value function (TVF). This represents the expected discounted present value of lifetime future earnings if someone was to leave the schooling system.

$$EV^w = E \left[\sum_{t=1}^T \beta^{t-1} u_w \right] \quad (8)$$

Similar to the reservation wage in the previous chapter, the TVF is calculated recursively.

3.3. The Utility of Attending School

The overall expected instantaneous utility to schooling is denoted by u_s , which encapsulates both the monetary and non-monetary utility to schooling. We allow the cost to schooling, κ_i , to vary depending on

⁷³ The education-squared term is introduced to capture the convexity in the returns to education found in South Africa and other developing countries.

⁷⁴ Individuals in our model are not allowed to choose between employment and non-employment. Our model is agnostic with regard to whether non-employment is a demand or supply side constraint.

⁷⁵ While our model does not explicitly model the choices agents face once they exit the schooling system (and enter the labour market), it does use the actual employment and wage data to ensure that the expectations in our model coincide with the actual outcomes.

whether students are in school or in higher education⁷⁶. The non-monetary gains, u_s^* , are assumed to be constant across each year of schooling and across individuals.

The overall instantaneous utility when studying is allowed to vary as follows:

$$u_s = \begin{cases} u_s^* - \kappa_1 & \text{if in school} \\ u_s^* - \kappa_2 & \text{if in tertiary institution} \end{cases}$$

The non-monetary utility of schooling, u_s^* , is assumed to be i.i.d. normally distributed with mean \bar{u}_s^* and standard deviation of σ_{ε_s} .

3.4. Grade Progression

Grade attainment is not guaranteed. An additional year of studying does not automatically lead to an additional year of education. As in real life, individuals will only accrue the benefits of a year of studying if they pass⁷⁷,

The probability of passing depends on schooling ability (v_i^s), an idiosyncratic shock (ε_{it}^s) that is loosely referred to as “school noise”, and the relative difficulty the grade the student is currently enrolled in (λ_i).

Passing is denoted by $pass_i$ and expressed in the following form:

$$pass_i(v_i^s, educ_i) = 1(v_i^s + \varepsilon_{it}^s > \lambda_i(educ_i)) \quad (9)$$

Given that the error, ε_{it}^s , is distributed normally the probability of passing can be modelled as follows⁷⁸:

$$P(pass_i(v_i^s, educ_i)) = \Phi\left(\frac{v_i^s - \lambda_i(educ_i)}{\sigma_{\varepsilon_s}}\right) \quad (10)$$

The difficulty parameter, λ_i , is allowed to differ between the latter and earlier phases of schooling as well as between normal schooling and higher education.

$$\lambda_i(educ) = \begin{cases} \lambda_1 & \text{if you are between grade 1 and grade 9} \\ \lambda_2 & \text{if you are in grade 10 or grade 11} \\ \lambda_3 & \text{if you are in grade 12 or tertiary institution} \end{cases}$$

The idiosyncratic error, ε_{it}^s , is assumed to be i.i.d. normally distributed with mean 0 and deviation σ_{ε_s} . We will refer to σ_{ε_s} as the noise parameter⁷⁹. We allow σ_{ε_s} to vary between the grades prior to matric, and the years following and including matric, since we believe that the standardised senior certificate exam provides students with a more reliable signal of their performance.

⁷⁶ In our model we assume that the monetary disutility to schooling, κ_i , captures the school fees, the school clothes and the travel cost to schooling. In Figure A2 in the appendix we show that the cost to schooling increases once individuals enroll into higher education.

⁷⁷ Belzil and Hansen (2002) assume that a year of studying and a year of education are the same thing. Anderson et al. (2001) find that black students progressed through school at a rate of 0.8 grades per year.

⁷⁸ The grade you attempt to pass will be the one that follows your immediate education level.

⁷⁹ In Item Response Theory (IRT) this is called the discrimination parameter, since it determines the model’s ability to discriminate between higher and lower ability individuals.

$$\sigma_{\varepsilon s} = \begin{cases} \sigma_{\varepsilon s}^s & \text{for someone between grade 1 and grade 11} \\ \sigma_{\varepsilon s}^m & \text{for someone in grade 12 or tertiary institution} \end{cases}$$

According to Lam et al. (2011) the strength of the feedback mechanism is not homogenous over all schools. They find that the stochastic component, that we refer to as ‘noise’ is larger in previously black schools than in previously white and coloured schools. Since apartheid there has been an extensive shift in resources towards predominantly black schools, to bring the spending on those schools in-line with former white schools. Despite this shift there has been no real convergence in the performance or observable change in quality within previously black schools (van der Berg, 2007).

The noise distorts the link between ability and achievement and weakens the quality of the information provided through this feedback mechanism. Consequently, those students who are in schools with noisy signals would be less informed about their true ability and likelihood of passing future grades when they have to decide whether to remain enrolled or to drop out.

3.5. Uncertainty and Information Updating

In our homogenous model all students have the same ability and are aware of their ability. Student therefore do not require any updating of their true ability. In the heterogeneous model, however, we allow students to have different schooling and labour market ability draws. Students in our model have imperfect information regarding these abilities. Students start off not knowing anything about their ability draw, but are able to update their expectations about their ability draw as they progress through school.

3.5.1. Updating Expectation about Schooling Ability

Throughout we will assume that students know the difficulty of each grade as well as the distribution from which their schooling ability and idiosyncratic shocks are drawn. Students are also assumed to be aware of the relationship between these factors and passing (as described in equation 9). Applying this knowledge, rational students are able to make use of their academic history to assess their true ability and probability of passing future grades.

The expected probability of passing, conditional on one’s expectation regarding one’s own ability is represented by:

$$P[Pass_i(educ)] = 1(E(v_i^s | educ, repeats) + \varepsilon_{it}^s > \lambda_i(educ_i)) \quad (11)$$

As students proceed through school they are able to update their expectation of their own ability, v_i^s . If they pass they will update their expectations of their ability upwards. If they fail they will update their expectations downwards. This is analogous to Bayesian updating (West, 1993), where the updated posterior probability is a function of the prior probability distribution and the information gained from either passing or failing the following year. The magnitude of these adjustments will depend on the likelihood of the event – if someone fails an earlier grade it will be more informative than if they fail a higher grade. Similarly, the

information contained in a pass and fail will be less informative about ability if idiosyncratic components within the schooling system are large⁸⁰.

At each grade, each type has a fixed probability of passing. A Bernoulli distribution is constructed to obtain the probability of a sequence of observed pass and fails conditional on type. For instance, the probability of observing a certain sequence of repeats and fails over the first 9 or fewer years (which are considered to be of a comparable difficulty) can be modelled as follows⁸¹.

$$P(\text{fails} = x) = \binom{N}{x} [1 - P(\text{pass}_i(\text{educ}))]^x [P(\text{pass}_i(\text{educ}))]^{N-x} \quad (12)$$

where N denotes the number of years that an individual has been enrolled in that phase of schooling, $P(\text{pass}_i(\text{educ}))$ denotes the probability of passing and x denotes the number of fails a specific student has accumulated.

3.5.2. Updating Expectation about Market Ability

Rational agents are able to use knowledge about their expected schooling ability to infer what their expected market ability would be. Students are assumed to be aware of the underlying relationship between school and market ability. Drawing on this information, any updates about schooling ability can also help inform expectations about labour market ability.

The expected terminal value function (TVF) can be rewritten as a function an individual's education, but also the number of times an individual has repeated.

$$E(V^w) = f(E(v_i^w | \text{Repeats}_i), \text{Educ}_i, \text{Age}_i) \quad (13)$$

4. Estimation Issues

4.1. Bellman Equation

Current decisions have implications for future rewards and choices. In deciding the optimal path agents need to be aware of not only the current payoffs but also the future consequences of their current actions. Uncertainties and expectations play a crucial role in how these future returns are factored into current decisions⁸².

⁸⁰ According to Lam et al. (2011) this signal is weaker in former black schools than in former white or coloured schools.

⁸¹ A different Bernoulli distribution that only counts the occurrences within that next phase of schooling where the likelihood of passing (an occurrence) is different.

⁸² Agents in our model are ignorant about what their future shocks will be, but have knowledge with regard to the probability distribution from which these future shocks will be drawn. What was previously envisioned to be the optimal decision path might turn out to be an inferior strategy once an array of shocks are realised.

In order to keep track of the sequential decisions faced by the individual we introduce the state space, denoted as Ω_t , that contains all the relevant information from the past that could influence the current-period decision d_t . The current state $\Omega_t = (d_{it-1}, ed_{it}, rep_{it})$ is determined by a combination of current endowments (education and times repeated already) as well as an indication of whether a person is still enrolled or not (d_{it-1}).

The maximum expected present value of lifetime utility at period t , given the state space Ω_a and the discount factor, is defined as follows.

$$V_a(\Omega_a, t) = \max_{\{d\}} E \left[\sum_{t=t_0}^{T_i} \beta^{t-1} [I(d_t = 0)u_s + I(d_t = 1)u_w] | \Omega_t \right] \quad (14)$$

The choice above follows the specifications of a finite Bellman equation⁸³. The value function above is equal to the higher of the two alternative value functions in our decision space – the value if a person chooses to remain enrolled and the value if he or she drops out.

$$V_t(\Omega_t, s_t) = \max_{\{d\}} [V_t^s(\Omega_t), V_t^w(\Omega_t)] \quad (15)$$

Since both of these value functions also obey the Bellman principles, they can each be rewritten as the sum of the respective instantaneous returns under both choices plus the discounted sum of another value function from the next period. The state space faced in the following period differs depending on what decision an individual makes at present (whether they set $d_t = 0$ or 1).

The value of studying for someone who has s_t years of schooling at age t is

$$V_t^s(\Omega_t, s_t) = u_s + \beta \left(pass_i(s_t) V_t(\Omega_t, s_t + 1) + (1 - pass_i(s_t)) V_t(\Omega_t, s_t) \right) \quad (16)$$

$$\text{where } V_t(\Omega_t, s_t) = EMax[V_{t+1}^s(\Omega_t, s_t), V_{t+1}^w(\Omega_t, s_t)]$$

The value of work for someone who has s_t years of schooling at age t is

$$V_t^w(\Omega_t, s_{it}) = u_w + \beta V_{t+1}^w(\Omega_t, s_{it}) \quad (17)$$

The difference between u_s and u_w reflects the current cost of schooling relative to the current payoff of working, while the difference between the two second terms in equation 16 and equation 17 describes the expected future payoffs under both options. In equation 16 we see that if students pass they get to pick between $V_{t+1}^s(\Omega_t, s_t + 1)$ and $V_{t+1}^w(\Omega_t, s_t + 1)$ in the next period, while if they fail they are forced to pick between $V_{t+1}^s(\Omega_t, s_t)$ and $V_{t+1}^w(\Omega_t, s_t)$. There are no further choices in equation 17, since individuals in our model are not allowed to re-enter the schooling system once they drop out.

⁸³ See Rust (1992) for a complete discussion of the principles that need to be satisfied in order to use a Finite Markov Bellman model

The model is solved recursively. Agents in our model first ask themselves whether, depending on a range of characteristics and factors, it would be worthwhile to go on to do a PhD if they had already acquired a master's degree. Once that choice has been accurately modelled, we go on to model the decision to enrol for a master's if one had completed a bachelor's degree. This process is repeated backwards up until the first year of education.

The recursive manner in which the model is solved allows agents to consider not only the direct benefit of the following year but also the option value of an additional year of education. Burger and Teal (2013) show that not all the gains to an additional year of education are through the marginal benefit of the next year alone. Due to the convexity in the returns to education, individuals may be willing to endure lower returns on earlier grades if it provides them access to later grades that have higher returns. Similarly, being in the schooling system provides individuals with more information about their true ability. Agents in our model take both these option values into consideration when they choose whether to remain enrolled.

4.2. Identification

Great caution should be taken when constructing structural models to ensure that there is enough variation to estimate each of the individual parameters – that the model is not over-specified.

The homogenous model has 15 parameters that need to be recovered: The parameters of the wage distribution ($\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4$), as well as the standard deviation for the idiosyncratic shocks in the wage function and the standard deviation and threshold parameter for the employment equation ($\sigma_{\varepsilon_{it}^E}, \sigma_{\varepsilon_{it}^W}, \tau$), the expected utility under home production, the expected utility under standard and higher education and the standard deviation of the utility of schooling ($u_h, u_s^* - \kappa_1, u_s^* - \kappa_2, \sigma_{\varepsilon_S}$), and the three difficulty parameters ($\lambda_1, \lambda_2, \lambda_3$)⁸⁴.

In the heterogeneous model we allow the labour market and schooling outcomes to be correlated to unobserved traits. The finite mixture model (FMM) we use distinguishes between two latent classes. Structurally, this requires us solve a further four parameters: the relative noise ratio ($\sigma_{\varepsilon_S^s}/\sigma_{\varepsilon_S^m}$), a schooling ability parameter, a market ability parameter and an estimate of the relative proportion of the groups.

The parameters in the wage and employment equation are identified in a similar manner as in the conventional Mincerian wage equation and Probit employment equation. Having repeated labour market observations for the same set of individuals allows us to ascertain how much of the variation around the expected wage and expected probability of employment is random (driven by ε_{it}^W) and how much is persistent over time (driven by ν_i^W)⁸⁵.

⁸⁴ In the homogenous model, where we do not have any schooling ability variation, the noise parameters are irrelevant.

⁸⁵ In our homogenous model, where $\nu_i^W = 0$, we assume that all the variation in employment and wages relative to the expected wages and expected probability of employment is random – driven by ε_{it}^W .

Ideally we would have wanted to identify u_S , κ_1 and κ_2 . Unfortunately in our model we do not have enough exogenous variation to uniquely identify the non-monetary and monetary cost to each grade. Instead we identify $u_S - \kappa_1$ and $u_S - \kappa_2$ ⁸⁶.

In our model, whether someone passes or fails a grade is determined by their own ability, the difficulty of the grade and some noise. The approach followed in modelling the role of each of these determinants is similar to the Item Response Theory (IRT) techniques that have been commonly used in psychology and education testing since the sixties (Rasch, 1980). Generally speaking, if many people repeat the same grade then that grade is considered to be difficult (high λ_i). Conversely, if the same individuals either repeatedly pass or fail then that would be evidence of a strong ability (v_i^w) effect⁸⁷. If we see different people failing different grades that would signal that there is a strong noise (ε_{it}^s) effect.

4.3. The Likelihood Function

In order to derive the structural estimates that best fit our data, we require a likelihood function. The likelihood function tell us how probable our model is to have generated the observed moments in our data.

For each individual, their overall likelihood will be the product of the respective likelihoods for each period.

$$L_i(\theta) = \prod_{t=1}^{T_i} L_{it} \quad (18)$$

During any period an individual can only be enrolled as a student or be in the labour market. The likelihood for those who are enrolled describes the probability of passing and dropping out, while the likelihood for those who are not enrolled represents the probability of being employed and the density of observed wages⁸⁸.

$$L_i(\theta) = \prod_{t=1}^{T_i} [P(d_{it}|pass_{it}, \Omega_{it})P(pass_{it}|\Omega_{it})]^{1(d_{it-1}=0)} [P(w_{it}|empl_{it}, \Omega_{it})P(empl_{it}|\Omega_{it})]^{1(d_{it-1}=1)} \quad (19)$$

The model above is augmented to allow for heterogeneity. The finite mixture model allows us to introduce persistent individual heterogeneity. In our model we will allow schooling ability (v_i^s) and market ability (v_i^m) to differ by type. Each type is assumed to occur with probability p_k . The above model needs to be modified slightly when we set up the finite mixture model, in which case the overall likelihood becomes the weighted sum of the likelihoods of each of the K types.

⁸⁶ In our data we do have information on the costs to normal and tertiary schooling. Potentially, if we are willing to believe these self-reported figures we would be able to ‘pin down’ what u_S could be.

⁸⁷ In our homogenous model $v_i^s = 0$. All the variation in school progression is therefore driven by grade difficulty and noise.

⁸⁸ As before, Ω_{it} denotes the entire set of variables known to the individual at period t that can affect preferences and expectations. Generally $\Omega_{it} = (rep_{it-1}, ed_{it-1}, d_{it-1})$, but for new entrants $\Omega_{it} = (0,0,0)$.

$$\log L = \log \sum_{k=1}^K p_k L_k \quad (20)$$

where p_k represents the population proportion and L_k represents the likelihood of type k .

Heckman and Singer (1984) propose that amount of types, K , be increased incrementally until increasing K has no more impact on the likelihood L . Aguirregabiria and Mira (2010), however, warn that if K is set larger than its true value then the model will not be identified. The computational cost also becomes a constraint with multiple types. Increasing K not only increases the parameter space that needs to be solved, it also increases the number of iterations to solve. For this reason most finite mixture models only allow for two or three types.

5. Empirical Analysis

5.1. Data

The National Income Dynamics Survey is a bi-annual survey that was conducted between 2008 and 2012 (SALDRU, 2015a-c). The survey follows the same set of individuals for three waves. Although we will be making use of the panel dimension of the survey, the largest portion of our school information in our panel is derived from the retrospective questions about schooling that were asked during the first survey⁸⁹. The baseline questionnaire contained a list of questions on repetitions that allowed us to construct a historical account of how each individual proceeded through school. If individuals are still in school in the 2nd and 3rd waves of NIDS then the subsequent waves provide us with further schooling information. If student are however out of school, then the 2nd and 3rd waves provide us with additional data on employment and observed wages. Since the retrospective questions about schooling were only asked to those individuals who were aged between 6 and 30 during the first survey we decided to restrict our sample to only contain black males from that age cohort.

5.2. Descriptive Statistics

Among a subsample of black males aged between 25 and 35 the distribution of education looks as follows:

Table 3.1: Distribution of Educational Attainment

	Probability
Primary School or Less	18.3%
Commenced High School	45.2%
Finished High School	26.0%
Commenced Tertiary Education	10.5%
Total	100%

Most young black males in our sample did attend some high school, but only 37% of the cohort went on to finish high school and only 10.5% went on to further their studies beyond school. The low school

⁸⁹ Individuals between the ages of 15 and 30 were asked whether they had repeated any grades and if they had, which grades these were.

completion rate is especially surprising given that the cost of schooling is relatively low (and sometimes free⁹⁰) and the returns are judged to be high and convex. The opportunity cost of employment for this group is also low. In chapter 1 we showed that black males are unlikely to be immediately absorbed into employment upon leaving school.

Although we will not be making use of the self-reported reasons for dropping out in our analysis, it is still worthwhile to have a look at these responses. The following table summarises the most common responses among young black males who dropped out prior to finishing high school (i.e. matric).

Table 3.2: Reported Reasons for Not Finishing High School

	Males
Worked	14%
Looked for Work	30%
Too Expensive	29%
Poor Marks	4%
Other	23%
Total	100%

Nearly half of the black males that drop out before finishing school claim to have done so because of labour market opportunities (either employment or the prospect of employment), while more than a quarter of the students claimed to have dropped out because school was too expensive. Only 4% of students report to have dropped out due to their own poor ability or poor marks⁹¹.

The following graph shows the likelihood of being enrolled by age.

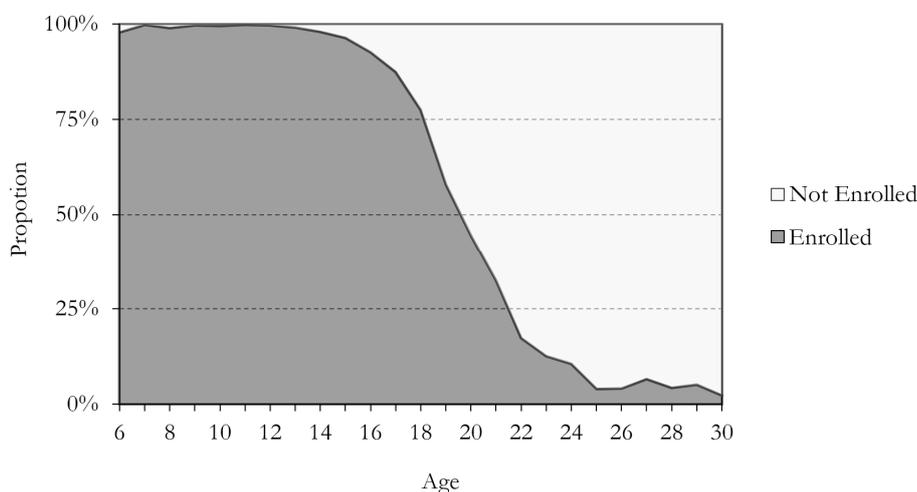


Figure 3.1: Probability of being enrolled among black males

⁹⁰ Schools that are judged to be serving students from the bottom three quintiles in South Africa are no longer allowed to charge schools fees.

⁹¹ Interestingly, 46% of the females that dropped out before finishing school claim to have done so due to pregnancy or birth. This is partly why we restricted our analysis to black males, since the educational attainment choice for females is intrinsically connected to the fertility decision.

Between the ages of 6 and 15 almost 100% of the population appears to be enrolled. This is not surprising, since attendance is supposed to be compulsory for all students between the ages of 7 and 15 (SASA, 1996). Between the ages of 15 and 18 we see that enrolment drops by roughly 25%. Most of these students would not have attempted the final school leavers' examinations. Over the following three years, from age 18 to age 21, we see an even sharper decline – enrolment drops by roughly 50 percentage points. Most of those students who are still enrolled after the age of 21 would be busy with tertiary schooling.

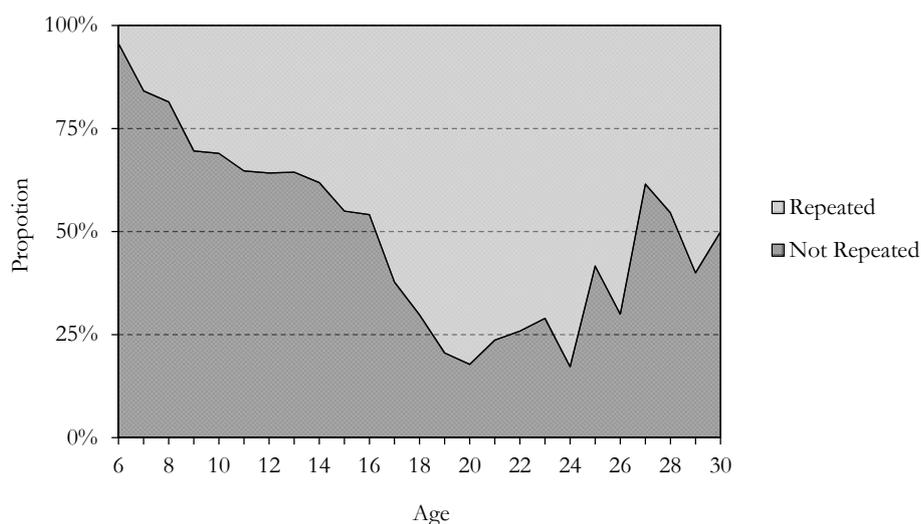


Figure 3.2: Probability of having repeated among black males who are currently studying

Repeating is prevalent among black males, especially among older students who have been in the schooling system for longer periods. Among those who are still enrolled, the likelihood of repeating increases up until the age of 20, whereafter it decreases again. There are two conflicting forces at play here. As a cohort ages, the probability of failing increases. However, there is also a selection process since we are not looking at the entire cohort across all the ages. Students who stay on to complete tertiary education are probably of higher academic ability and less likely to have repeated any grades while in school⁹².

Grade repetition does not only vary among individuals, it also varies between grades. The following table compares the matric pass rate to that of the three grades that preceded it.

Table 3.3: Probability of Passing by Prior Repeats

	Grade 9-11	Grade 12
Having Never Repeated	88.3%	74.5%
Having Repeated Once	82.2%	57.6%
Having Repeated Twice	75.4%	52.5%
Having Repeated Three or more times	60.0%	51.9%
Unconditional Probability	82.9%	64.5%

⁹² The spikes that are observed at higher age levels is due to the decreased sample size, which in turn was caused by low levels of enrolment at higher ages.

The table shows that the average unconditional probability of passing grades 9, 10 or 11 is 83%, while the unconditional probability of passing grade 12 is 65%. These reported matric pass rates in our data compare favourably to the average black matric pass rate of recent years, which was estimated at 62.9% in 2010 (SAIRR, 2012).

If we condition on a student's academic record up to that point (whether they have failed and how often), we get a more accurate estimate of a student's probability of passing the next grade. Someone who has not failed before has an 88% chance of passing grade 9, 10 and 11. The likelihood drops to 82% if they have failed once, 75% if they have failed twice, and 60% if they have failed more than two times. Those who have not repeated any grades up to grade 12 have a 75% chance of passing matric. The probability drops to between 50% and 60% if an individual has repeated⁹³.

Repeats are also correlated to labour market outcomes. The table below shows the expected probability of employment and the average hourly wage among a set of black males with exactly 12 years of education, conditional on the number of times they repeated while in school⁹⁴.

Table 3.4: Labour Market Prospects by Repeats

	Empl	Wage
Having Never Repeated	61.5%	28.36
Having Repeated Once	53.1%	20.96
Having Repeated Twice	46.5%	17.67
Having Repeated Three or more times	31.6%	8.71
Unconditional Wage	55.1%	23.77

On average, black males who have repeated are less likely to be employed and more likely to earn lower wages when employed than those who have not repeated. Similarly, black males who have only repeated once face better labour market prospects than those who have repeated multiple times. In our structural model, we assume that repetition in itself does not affect earnings and employability, but simply reveals that someone may be of a lower labour ability – with a weaker labour market draw. This hypothesis will be tested in our finite mixture model.

5.3. Fitting the Data to the Model

Following a maximum likelihood approach, we proceed to find the set of parameters that are most likely to have produced the education and labour market outcomes that we observe in the data. For each iteration, the likelihood function compares the actual outcomes from the data to the predicted moments from our model, until our numerical optimization algorithm finds the set of parameters that are deemed to be most adequate.

⁹³ The actual matric pass rate would have been lower if weaker students had not failed or dropped out along the way.

⁹⁴ The results were obtained from table A2 in the appendix.

The sample contains black males aged between 6 and 35. Unfortunately, the retrospective portion of our panel, from where most of the educational data is derived, does not contain any data on pass or fails, but only on repetitions. The distinction between failing and repeating is trivial at lower grades, where most people who fail remain enrolled and become repeaters. The distinction is however problematic at higher grades, since we are unable to distinguish between those students who failed a grade and dropped out and those students who decided to drop out straight after finishing the previous grade.

As with the job-search model, the optimal set of parameters were derived using the Powell and Fminsearch algorithm. Both these algorithms were obtained from the iFit library in Matlab (Farhi, 2011; Farhi et al., 2013).

5.4. Results

The following table contains the set of parameters of the DCDP model that best fit the observed data for our sample of black males. Model 1 represents the original OLS estimates. Model 2 is a DCDP Model, where all individuals are assumed to have identical ability draws. The DCDP model is expanded further to allow for unobserved heterogeneity in model 3.

Table 3.5: Model Results

	OLS Model (Model 1)	Homogenous Model (Model 2)	Heterogeneous Model (Model 3)
Wage/Employment			
Constant	1.120	-1.588***	-1.592***
Education	-0.083**	-0.122***	-0.125***
Education Squared x 100	0.013***	0.013***	0.008***
Age	0.036	0.276***	0.275***
Age Squared x 100	-0.007	-0.005***	-0.004***
S.D. of Wage Shock		0.902***	0.833***
S.D. of Employment Shock		4.593***	4.981***
Employment Threshold		2.269***	1.422***
Non-Working Utility			
Utility at home		-0.580*	-1.378***
Utility at school		1.172***	0.145
Utility at tertiary institution		-1.566***	-1.684***
S.D. of School Shock		9.739***	4.551***
Passing			
Difficulty parameter for Gr1-9		-1.411***	-3.794***
Difficulty parameter for Gr10-11		-1.006***	-2.039***
Difficulty parameter for Gr12+		0.085***	-0.041
Relative Noise Ratio ($\sigma_{\epsilon S}^S / \sigma_{\epsilon S}^m$)			2.642***
Ability			
Proportion Type 1			0.157***
Schooling Ability Type 1			1.392***
Market Ability Type 1			0.906***
N	1617	22408	22408
R²	0.166		
Log Likelihood		-22300.4	-22420.5

Note: * significant at the 10% level, ** significant at the 5% level, *** significant at the 1% level

The returns to education are convex – every additional year of education is worth more than the one that preceded it. In the figure below we show how the returns to education vary over the different years of attainment for the three different models⁹⁵.

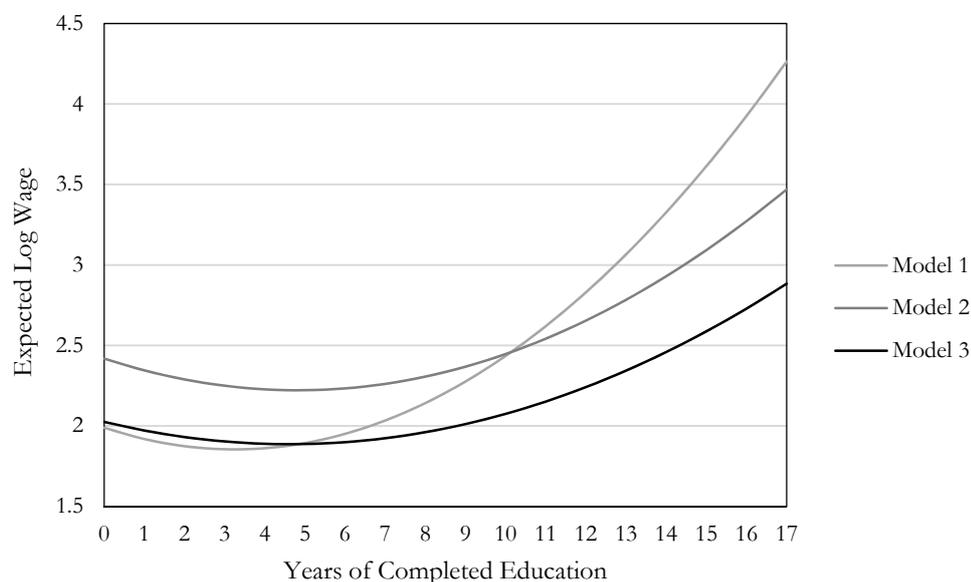


Figure 3.3: Returns to education profile⁹⁶.

The returns are highest and most convex in the descriptive OLS model (model 1). The returns are lower and less convex for model 2. Model 2 assumes homogeneity. The reduction in the returns between model 1 and 2 is therefore not driven by the difference in the educational attainment between types, but rather by the general educational enrolment decisions. The returns to education are the flattest for Model 3, where we introduce heterogeneity and attempt to correct for ability bias.

The instantaneous utility of being at home is slightly lower in this model than in the previous chapter. The overall instantaneous utility was lower during higher education than during schooling. Theoretically we assume that individuals gain the same non-monetary utility out of both, but that the cost of schooling would be higher for higher education. Individuals in our dataset were also asked about the cost of schooling. The median self-reported annual cost to an ordinary year of schooling was reported to be R400, while the median cost of a year of tertiary schooling was roughly R10000 (see figure A2 in the appendix).

⁹⁵ The OLS regression results that we obtain from model 1 was fitted to the subsample of employed males for who we had wage data. Not many individuals in our sample were employed. This is partly why we chose to let the education and age effect employment and wages in the same manner, since it allows us to use both employment and wage outcomes to model the returns to education.

⁹⁶ The graph represents the expected wage for someone who is aged 28 and is of average ability.

The difficulty estimates for the three phases differ greatly. Our results show that it becomes incrementally tougher to pass a grade as we proceed through the schooling system. The probability of passing grade 1 through 9 is estimated at 97% in the homogenous model. The probability drops to 78% for grade 10 and grade 11, and to 45% for matric and higher education.

The finite mixture allows us to recover the proportions, and market and schooling ability values for each type. The estimated probabilities of each type and their corresponding school and market abilities are given in the table below⁹⁷.

In the homogenous model everyone was of the same ability and the probability of passing a specific grade. In the heterogeneous model we allow for two different ability types and allow the probability of passing a grade to vary by type. Conceptually, if we see that previous fails have a large impact on future fails this would be evidence of a significant level of heterogeneity in schooling ability. Individuals were also allowed to vary in their labour market abilities. Using the repeated waves we are able to exploit employment and wage outcomes to see whether an individual is doing surprisingly poorly or surprisingly well conditional on his level of education and experience. These two sets of ability draws are then compared to see whether they are correlated – whether there is an ability bias.

Table 3.6: Distribution of School and Market Ability

	Probability	School Ability	Market Ability
Type 1	15.7%	1.391	0.906
Type 2	84.3%	-0.260	-0.169
Overall	100.0%	0.000	0.000

A positive correlation was uncovered. The person types who have the highest schooling ability draws also have the highest market ability. The results are similar to that of Belzil and Hansen (2002), who also find a positive correlation between schooling and market ability. Figure 3.A.3 in the appendix illustrates the positive relationship more explicitly. The results are consistent with the premise of ability bias, where individuals with greater unobservable labour market attributes may select into higher levels of education. In the presence of ability bias conventional OLS regressions that attempt to recover the returns would be biased, since some of the positive correlation between education and wages would work through ability rather than education. This also explains why the returns to education were lower than was originally predicted in the descriptive OLS model.

⁹⁷ The proportion of type 2 individuals as well as their respective school and ability draws can be backed out once we know the proportion of type 1 individuals, and the school and ability draws for type 1 individuals.

6. Simulations

6.1. Prospects among Young Black Males

Using the parameters from the finite mixture model (model 3) we proceed to simulate the transitions from childhood to early adulthood for the different ability types.

The table compares the average level of education, the average number of fails and the average years enrolled between the two types:

Table 3.7: Education, Years of Schooling and Repetitions by Type

	Years Passed	Years Failed	Years Enrolled
Type 1	10.44	0.52	10.96
Type 2	9.81	1.29	11.10
Overall	9.91	1.16	11.08

These simulations provide us with an insight as to how students progress through school, when they drop out and how quickly they get absorbed into the labour market. The following two figures show the expected likelihood of studying and working during a black male's early life. Each year, an individual can either be enrolled or not. Individuals who were enrolled were either classified as "Pass" if they studied and passed that year or as "Fail" if they studied and failed that year. Similarly non-enrolled individuals were classified as "Work" if they were employed or as "Home" if they were not.

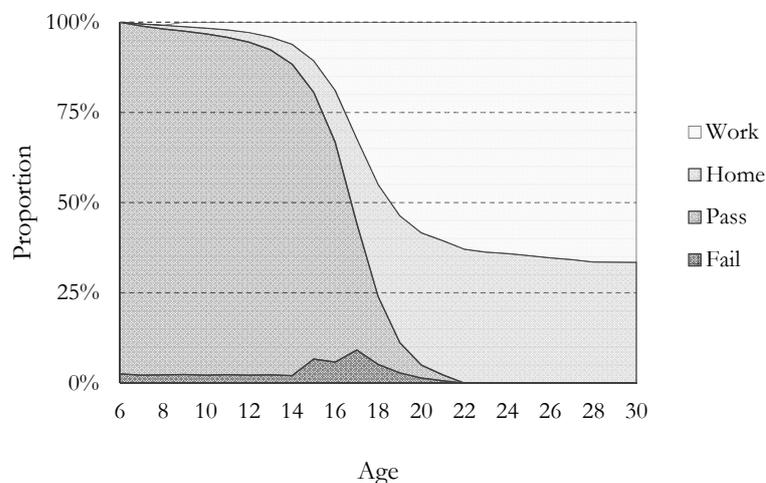


Figure 3.4: Activities by age for type 1 individuals.

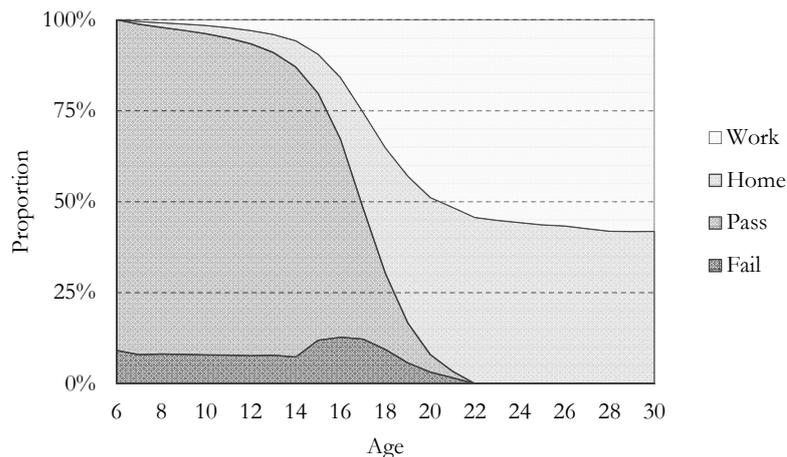


Figure 3.5: Activities by age for type 2 individuals.

The differences are subtle. Generally the two groups drop out at a similar rate and at similar ages. Type 1 individuals have a 2.5% chance of failing an early grade, compared to type 2 individuals who have 8.7% probability of failing an early grade. The probability of failing jumps to almost 10% for type 1 individuals and almost 25% for type 2 individuals when they reach grade 10. The jump is not that dramatic on either of the figures, since not all students make it to grade 10 and even less in the allotted years. Type 1 individuals have a 6.6% probability of failing matric, compared to type 2 individuals who have a 56% probability of failing matric.

By age 25 we see that nearly 65% of type 1 individuals are employed. At the same age, roughly 55% of type 2 individuals have found employment. This gap in employment likelihood among the two types remains fairly constant from age 25 onwards.

6.2. Reducing the Noise in the Schooling System

Lam et al. (2011) point out that the additional noise in the feedback signal that students have about their performance in formerly black schools distorts the link between ability and achievement. Our model provides a theoretical landscape wherein we are able to explicitly test this hypothesis. We proceed to shock the feedback mechanism to see how it will affect different ability types. In the original model the grades before matric had a standard deviation that was 2.64 times higher than matric. In the simulated model we set the standard deviation of the earlier grades equal to the standard deviation of matric.

Our simulations show that when the idiosyncratic error within the schooling system is set to the same level as matric it will effect educational attainment through two channels. Firstly, it will affect the progression through schools directly by affecting the probability of passing. Secondly, it will strengthen the feedback mechanism, which will help students make more informed decisions.

6.2.1 The Direct Effect: The Probability of Passing

The figures below illustrate how the actual probabilities of passing each grade will be altered if the noise in the earlier grades were the same as in matric. In our simulation we set $\varepsilon_{it}^S = \varepsilon_{it}^{M98}$. Figure 5a shows the relationship between ability and passing each grade in the original model. Figure 5b shows the relationship between ability and passing for the simulated model, where the noise is decreased.

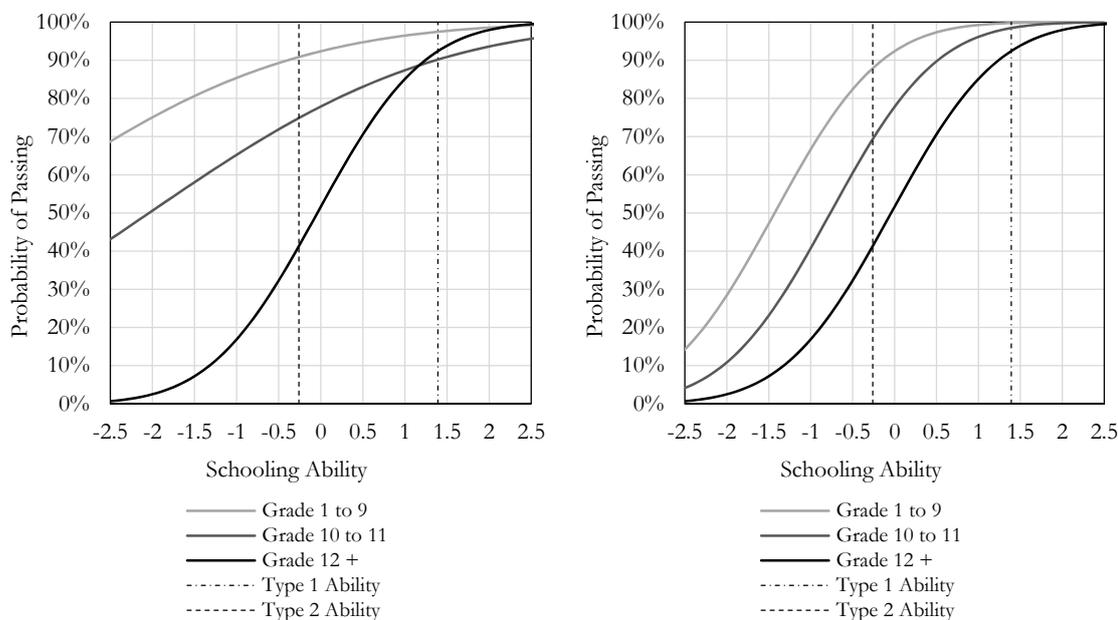


Figure 3.6.a: Actual probability of passing

Figure 3.6.b: Simulated probability of passing

The slope of each grade curve is determined by the ‘noise’ parameter, while the position is determined by the difficulty parameters. The dashed lines indicate the ability levels of both types. The probability of passing, for each type and under both scenarios, can be obtained from the intersect where the curved and dashed lines meet.

Figure 5a shows the results for our baseline model, where we used the parameters from model 3. Type 1 individuals have a 2.5% chance of failing early grades, a 9.5% chance of failing grade 10 or 11 and a 6.7% chance of failing thereafter. Despite grade 10 and 11 being easier than grade 12, type 1 individuals have a higher probability of passing the latter. This is because the noise is larger for the earlier grades.

Figure 5b was constructed using the simulated for the scenario where the noise of earlier grades is reduced to the matric level. Type 1 individuals have less than 1% probability of failing early grades, a 1.4% probability of failing grade 10 and 11 and a 6.6% probability of failing a year thereafter. Type 1 students will be less likely to fail when the noise in the schooling system is decreased. Type 2 individuals have an 8.7% chance of failing an early grade, a 24.5% chance of failing grade 10 or 11 and a 56% of failing grade thereafter.

⁹⁸ The probability of passing for someone with average ability ($v_i^S = 0$) is fixed. Each curve is pivoted around this point.

Under the simulation these probabilities increase to 11%, 29% and 56%, respectively. Thus, type 2 students will be more likely to fail if the noise is decreased.

6.2.2. The Indirect Effect: The Information Revelation

The indirect effect works through information updating. As the level of noise in the feedback mechanism reduces, the signal becomes less distorted and passes and fails become a more informative signal of underlying schooling ability. The figure below compares the speed at which students learn about their true abilities under the two scenarios⁹⁹. In the “baseline” we keep the noise element to early grades at its estimated level, while in the “simulation” we reduce the noise to the matric level.

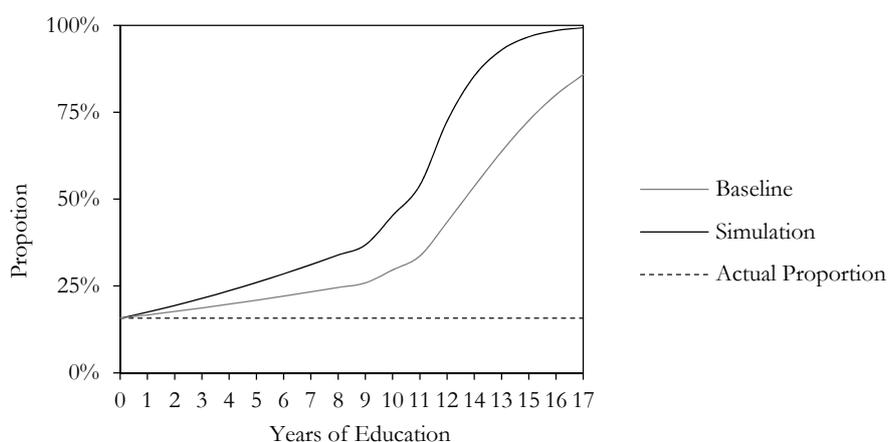


Figure 3.7: Probability of being type 1 conditional on never having failed

Updating at early grades is slow. For the early grades the failure rate is low and the noise is high. Consequently, we find that most students pass, and learn very little about their ability. Due to the noise in the schooling system, it is often not the least able students who end up failing. The signal becomes more informative at higher grades, where grades are more difficult and the signal is less distorted. When we reduce the noise, both type 1 and type 2 students are quicker to recognise their true type.

6.2.3 The Total Effect: The Change in Education

If schools are less noisy, weaker students will be more likely to fail, while better students will be more likely to pass. Put simply, when the noise in the schooling system decreases, good students will learn that they are good students more quickly, but will also be more likely to pass relative to weaker students than in a noisier school system. Similarly, weaker students will be quicker to learn about their true type, but will also be more disadvantaged in that they will be less likely to ‘sneak through’ grades than before. Both effects work in the same direction.

⁹⁹ It depicts the probability of being of type 1 if someone has never failed a grade.

The figure below shows the combined effect of reducing the noise component. In the simulated model the standard deviation of the earlier grades was set to be equal to be noise in matrix.

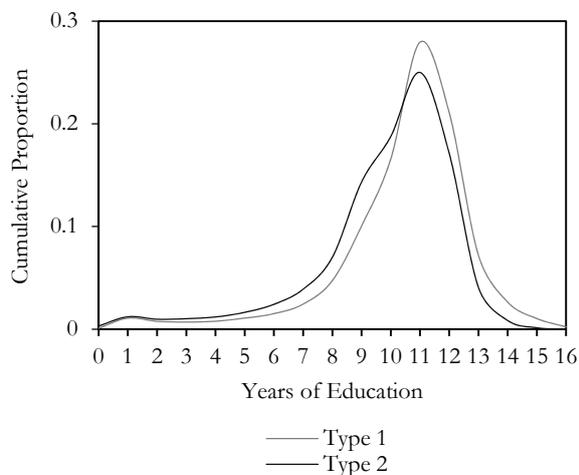


Figure 3.8.a: Distribution of education in baseline model

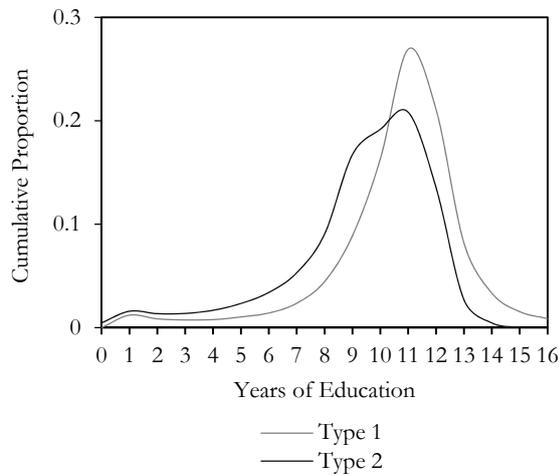


Figure 3.8.b: Distribution of education in simulated model

Reducing the noise within earlier grades causes an increase in the gap in educational attainment between the two ability types. On average, type 1 individuals will attain more education, while type 2 individuals will attain less education than is the current situation.

The table below compares the years in schooling and educational attainment for each ability type.

Table 3.8: Education, Years of Schooling and Repeats by Type

	Baseline			Simulation		
	Passed	Failed	Enrolled	Passed	Failed	Enrolled
Type 1	10.44	0.52	10.96	10.56	0.09	10.65
Type 2	9.81	1.29	11.10	9.32	1.49	10.82
Overall	9.91	1.16	11.08	9.52	1.27	10.79

The simulated results are compared to original results in the baseline model. Compared to the baseline scenario, type 1 individuals spend slightly less time in school, but attain more education since they are less likely to fail than before. Their average level of education increases from 10.44 to 10.56. Type 2 individual spend less time in school and spend more time failing grades. Their average level of education among type 2 individuals decreased from 9.8 years to 9.3 years. Our simulated results suggest that the gap in schooling between the two groups will double from just more than half a year of education to more than a whole year if the level of noise in the schooling system is reduced.

Conceptually a reduction of the noise within schools that will allow people to make more informed decisions should be a desired outcome. In doing so government can help people increase their utility.

However, from a policy perspective this seems unfortunate that any policy aimed at reducing noise (like standardized national testing at earlier grades) will cause some students to drop out sooner than they currently will.

7. Conclusion

In this chapter we attempt to formulate a structural model that is capable of explaining the sequence of schooling decision faced by utility maximizing agents in the face of uncertainties regarding future shocks and their own true ability. Individuals were allowed to be heterogeneous in their labour and schooling ability. Using dynamic programming we derive the set of parameter that best reflects the actual school and labour market outcomes for a group of young black males.

The strong behavioural assumptions and the structure of our data allowed us to test whether labour market and schooling ability are in fact correlated. We find a strong positive correlation between the school and market ability, leading us to reject the exogeneity assumption. Our results are consistent with the existence of a positive ability bias, where more able students proceed through school more quickly, attain more schooling, find work earlier and earn higher wages, while less equipped students proceed through school more slowly, attain less schooling, take longer to find work and earn lower wages.

Having solved the structural parameters of our model we proceeded to test how low and high ability individuals would react to a decrease in the idiosyncratic component within earlier grades – how they would have fared if there was less noise in their schools. Our simulated results show that high ability students would have been less likely to fail and would have learned and obtained even more education, whereas low ability students would have been more likely to fail and would have obtained less schooling. The ability bias would, therefore, have been higher if the level of noise in the schooling system was lower.

Appendix 3

Appendix 3.A: Additional Figures and Tables

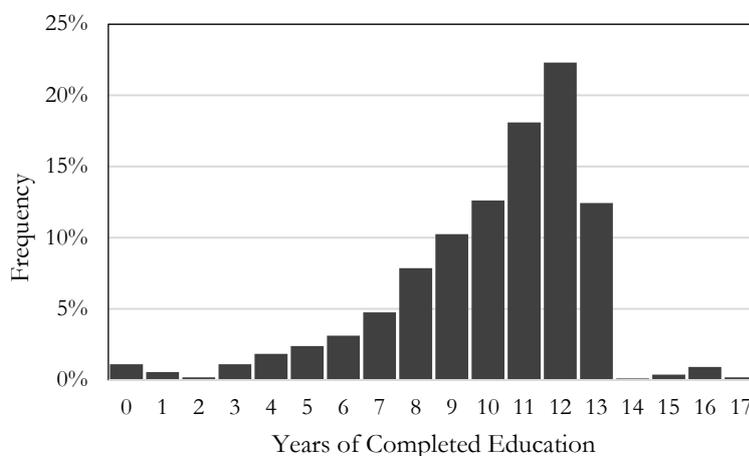


Figure 3.A.1: Histogram of years of schooling among young black males

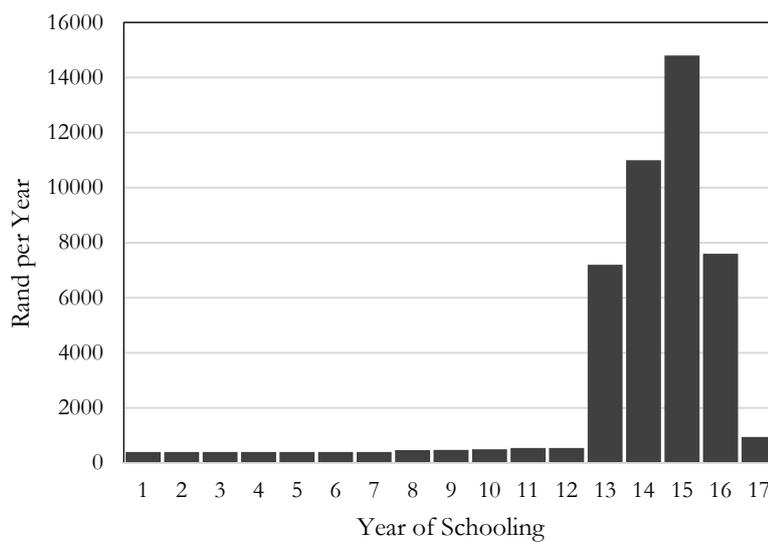


Figure 3.A.2: Median self-reported cost of schooling by years of schooling.

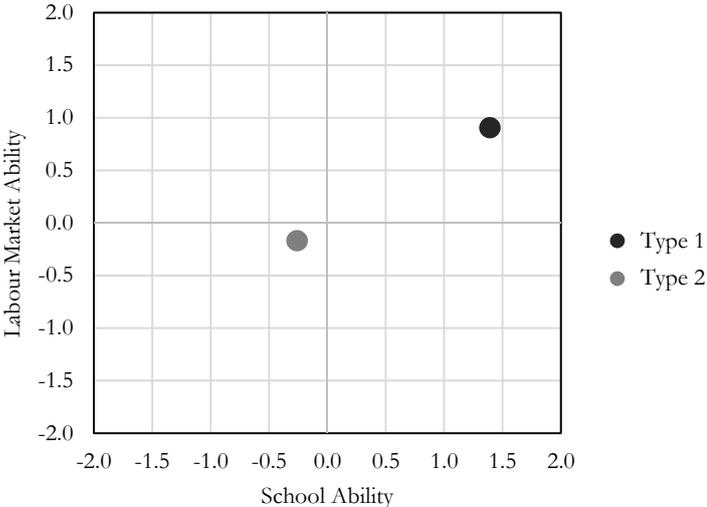


Figure 3.A.3: Distribution of unobserved ability.

Table 3.A.1: Probability of Passing Conditional on Repetition History

	Passing Gr 9-11	Passing Gr 9-11	Passing Gr 12	Passing Gr 12
Constant	0.829 (0.010)	0.883 (0.014)	0.645 (0.027)	0.745 (0.040)
Repeat = 1		-0.061 (0.023)		-0.168 (0.064)
Repeat = 2		-0.128 (0.030)		-0.220 (0.085)
Repeat \geq 3		-0.283 (0.043)		-0.226 (0.099)
N	1435	1405	301	296
R²	0.000	0.037	0.000	0.041

Table 3.A.2: Expected Wages and Probability of Employment Conditional on Repetition History

	Empl	Empl	Logged Wage	Logged Wage
Constant	0.551 (0.027)	0.615 (0.039)	2.956 (0.023)	2.660 (0.063)
Repeat = 1		-0.084 (0.061)		-0.247 (0.138)
Repeat = 2		-0.150 (0.085)		-0.458 (0.197)
Repeat \geq 3		-0.299 (0.120)		-0.693 (0.381)
N	336	336	211	211
R²	0.000	0.025	0.000	0.034

Note: Sample was restricted to only include black males with matric who were between the ages of 25 and 35.

Table 3.A.3: Model Results

	OLS Model (Model 1)	Homogenous Model (Model 2)	Heterogeneous Model (Model 3)
Wage/Employment			
Constant	1.120 (0.830)	-1.5883 (0.067)	-1.592 (0.058)
Education	-0.083 (0.037)	-0.122 (0.005)	-0.125 (0.003)
Education Squared x 100	0.013 (0.002)	0.013 (0.001)	0.008 (0.001)
Age	0.036 (0.065)	0.276 (0.001)	0.275 (0.001)
Age Squared x 100	-0.007 (0.001)	-0.005 (0.000)	-0.004 (0.000)
S.D. of Wage Shock		0.902 (0.018)	0.833 (0.019)
S.D. of Employment Shock		4.593 (0.916)	4.981 (0.838)
Employment Threshold		2.269 (0.215)	1.422 (0.323)
Non-Working Utility			
Utility at home		-0.580 (0.326)	-1.378 (0.238)
Utility at school		1.172 (0.227)	0.145 (0.148)
Utility at tertiary institution		1.566 (0.596)	-1.684 (0.473)
S.D. of School Shock		9.739 (1.010)	4.551 (0.628)
Passing			
Difficulty parameter for Gr1-9		-1.411 (0.009)	-3.794 (0.013)
Difficulty parameter for Gr10-11		-1.006 (0.029)	-2.039 (0.320)
Difficulty parameter for Gr12+		0.085 (0.028)	-0.041 (0.050)
Relative Noise Ratio ($\sigma_{\epsilon S}^S/\sigma_{\epsilon S}^m$)			2.642 (0.404)
Ability			
Proportion Type 1			0.157 (0.017)
Schooling Ability Type 1			1.392 (0.155)
Market Ability Type 1			0.906 (0.056)
N	1617	22408	22408
R²	0.166		
Log Likelihood		-22300.4	-22420.5

Appendix 3.B: Choosing whether to sit Matric

The choice to stay in school and attempt to sit the final school leaver's examination is a vital question within the South African schooling system. Unlike previous grades, the final exam is nationally monitored. Our model finds, like Lamb et al.'s research (2011), that the noise component is smaller in matric than for previous grades. Passing grade 12 therefore provides both students and future employers with a stronger signal of actual ability than earlier grades do.

To illustrate how our model works, we consider the enrolment choice for matric from the perspective of a student who has just finished grade 11.

Type 1 students have a 93.5% chance of passing matric, while type 2 students have only a 43.6% probability of passing matric. Unfortunately, students only have partial knowledge with regard to which one of these ability types they belong. Rational students are able to use their knowledge about their expected ability type to calculate their expected probability of passing.

$$E[P(\textit{pass})] = \sum_{k=1}^2 P(\textit{type} = k)P(\textit{pass}|\textit{type} = k)$$

Without any additional information, students will assume probability weights, in which case the expected probability of passing matric would be 47.6%.

In our model, however, students are able to use their academic history to update their expectations about their type. The equation can be rewritten in terms of one's educational history.

$$E[P(\textit{pass})|\textit{educ}, \textit{fails}] = \sum_{k=1}^2 P(\textit{type} = k|\textit{educ}, \textit{fails})P(\textit{pass}|\textit{type} = k)$$

A student who has completed grade 11 without ever failing a grade will have a 60% chance of passing matric the first time. The probability of passing will be lower if a student had failed a previous grade and even lower if a student had failed more than once¹⁰⁰. If a student had only failed once and that was in an early (easier) grade, the likelihood of passing matric will be 51%, while if such a student had failed only once, but that failure was in either grade 10 or grade 11, the probability of passing matric will be 54%.

In our model we assume that because the matric exam is standardized, the exam will be less noisy. Students are therefore likely to learn more about their own ability from sitting the matric exam than they would from sitting a normal non-standardized exam. In our simulated model, only 30% of those students who failed the matric exam decided to come back to attempt it again. To the author's knowledge this model is the only one that is capable of explaining how rational individuals who found it worthwhile to write the matric exam the year before will opt out in the current year following failure the year before.

¹⁰⁰ Having failed an easier grade provides more information on a student's ability than having failed a tough grade, as it signals that one is more likely to be of a weaker draw.

Conclusion

A reading of the current literature may lead one to deduce that the labour market is mobile and that reservation wages are exuberant in South Africa, which suggests that most of the unemployed would be able to find work if they were willing to work for lower wage offers. Multivariate regressions also show that the returns to education are high and convex. It is however important to remember that these methods are descriptive in nature, they merely convey the observed relationship between covariates. Before one is able to make any causal inferences from these reduced form estimates one needs to make sure that the measured effect is not biased and that the estimate has an economic interpretation. This, in turn, requires one to address the endogeneity concerns that are inherent to most of these estimates and to derive a theoretical model that is capable of giving meaning to such estimates.

In this thesis, I contribute to the literature, by choosing three such descriptive results and exposing the inconsistencies that undermine their interpretation. In all three chapters, I look at the conventional estimates, discuss why I think they may be biased and then proceed to find more reliable estimates that are more equipped to explain the behavioural outcome that are observed in the data.

In chapter one, I show that the naïve 1-period employment transition estimates that are commonly used to describe employment mobility are inconsistent with the observed age-employment profile and with other multiple-periods transition estimates. In chapter two, I show that self-reported reservation wages are prone to over-optimism – individuals report reservation wages that are higher than their expected earnings and are seen accepting jobs that pay less than their reported reservation wages. In chapter three, I investigate whether the returns to education are truly as high as and convex as ordinary least squares model report them to be or whether the returns are biased by ability bias.

In my analysis I make use of both structural and reduced form models that are better suited to the South African labour market landscape – allowing for unemployment and unobserved individual-specific heterogeneity. Theoretical models rarely simplify to linear relationships that are recoverable through conventional ordinary least squares regression. The models that I employ are no exception in this regard. In all three chapters I make use of maximum likelihood, with customised likelihood functions, to recover the set of parameters that were most likely to have produced the data, given the behavioural assumption about how optimizing individuals should behave. The economic theory that is imposed on our models allows one to interpret the results in a causal manner.

In the first chapter, I show that the transition rates that I derive from classic transition matrices are too high. If the probability to transition into employment is truly as high as the reported estimates claim, then young individuals would transition into employment at a much quicker rate than what is currently observed. It is unclear whether the divergence is driven by misclassification error or unobserved individual-specific

heterogeneity, but what one can say with confidence is that conventional transition rates do not provide an accurate estimate of the employment transition probabilities that are faced by all the individuals in the labour market. Most black males experience far lower job entry and exit rates than are generally reported. The results show that the probability of transitioning into employment among young black males within a six month period may be as low as 4%.

In chapter two, I show that current reservation wages are not consistent with the transition and wage data. I look at the transition between period and the observed wage distribution to derive a plausible wage offer and job offer distribution. The results suggest that unemployment is driven by low job arrival rates rather than high reservation wages. Only 10% of the job offers that the non-employed are offered get declined. The job refusal rate that I recover is much lower than that of Levinsohn and Pugatch (2010). According to their results 71% of job offers get declined among young inhabitants in Cape Town.

Descriptive results show that those individuals that partake in active search are more likely to transition into employment. The job-search model in chapter two shows that most of the difference in job offer rates among the searching and non-searching non-employed is due to the positive selection into searching by the more able and more educated. The results that I obtain show that individuals with more desired attributes are more likely to select into employment since the gains to search are higher for them – the incremental change in their job offer rate is higher and the expected wage offers that they will receive is higher.

In chapter three, I show that the returns to education estimates are not aligned with the educational attainment figures. Descriptive results suggest that the returns to education in South Africa are high and convex. Yet, most people appear to drop out of school right before the largest increases are about to occur. This behavior would be difficult to rationalize from the perspective of rational, well-informed agents that are not liquidity constrained. I find that a large part of the difference in wages is attributed to ability differences. Education increases the likelihood of finding employment and the expected returns, but not as drastically as descriptive regressions would have us believe. The estimates that I obtain when I control for ability bias are much lower than the reduced form estimates.

Lam et al. (2001) showed that the feedback mechanism is blunt within previously black schools and that the link between hard work (or ability) and achievement in those schools are often distorted. In my structural model I show how able students who operate within these school are unlikely to be aware of their true potential and that these students will drop out earlier than they would have if they were aware of their true ability.

One of the advantages of the structural approach is that it provides us with deep structural parameters. Unlike the reduced form estimates, these estimates are not subject to the Lucas critique. These estimates thus allows one to predict how rational optimizing individuals will respond to never before seen policy

changes or economic shocks. I show that if the noise within earlier grades was reduced then students would be better informed about their true ability. The simulated results also suggests that more capable students would be more likely to stay in school for longer and attain more education.

Bibliography

- AGUIRREGABIRIA, V., MIRA, P., 2010. Dynamic Discrete Choice Structural Models: A Survey. *Journal of Econometric*, 156(1), 38-67.
- ANDERSON, K., CASE, A., LAM, D., 2001. Social Dynamics. Causes and Consequences of Schooling Outcomes in South Africa: Evidence from Survey Data. *Social Dynamics*, 27(1), 1-23.
- ARROW, K., 1973. Higher Education as a Filter. *Journal of Public Economics*, 2, 193-216.
- ATTANASIO, O., MEGHIER, C., SANTIAGO, C., 2012. Structural Model and a Randomized Experiment to Evaluate PROGRESA. *Review of Economic Studies*, 79(1), 37-66.
- BALLEN, J., FREEMAN, R., 1986. Transitions between Employment and Nonemployment. In R. Freeman and H. Holzer, eds., *The Black Youth Employment Crisis*, Chicago: University of Chicago Press.
- BANERJEE, A., GALIANI, S., LEVINSOHN, J., MCLAREN, Z., WOOLARD, I., 2006. Why has Unemployment Risen in the New South Africa? *The Economics of Transition*, 16(4), 715-40.
- BECKER, G.S., 1964. Human Capital. *New York: Columbia University Press*.
- BECKER, G.S., 1967. Human Capital and the Personal Distribution of Income: An Analytical Approach: Ann Arbor, MI: University of Michigan, Institute of Public Administration.
- BELZIL, C., HANSEN, J., 2002. Unobserved Ability and the Return to Schooling. *Econometrica*, 70(5), 2075-2091.
- BELZIL, C., 1996. Efficiencies and Comparative Advantage in Job Search. *Journal of Labor Economics*, 14(1), 154-173.
- BHORAT, H., OOSTHUIZEN, M., 2007. Young People and the Labour Market. *Africa Insight*, 37(3), 388-403.
- BIEMER, P.P., BUSHERY, J.M., 2000. On the Validity of Markov Latent Class Analysis for Estimating Classification Error in Labor Force Data. *Survey Methodology*, 26(2), 139-152.
- BLAU, D.M., 1991. Search for Nonwage Job Characteristics: A Test of the Reservation Wage Hypothesis. *Journal of Labor Economics*, 9(2), 186-205.
- BLAU, D.M., ROBIN, P.K., 1990. Job Search Outcomes for the Employed and Unemployed. *Journal of Political Economy*, 98(3), 637-655.

- BOUND, J., BROWN C., MATHIOWETZ, N., 1999. Measurement Error in Survey Data. In *Handbook of Econometrics, vol. 5.*, Heckman, J., Leamer, E., (ed.). Amsterdam: North Holland
- BRENT, R.P., 1973. Algorithms for Minimization without Derivatives, Prentice-Hall, Englewood, New Jersey.
- BROWN, S., TAYLOR, K., 2013. Reservation wage, market wages and Unemployment: Analysis of Individual Level Panel Data. *Economic Modelling*, 28(3), 1317-27.
- BURGER, R.P., TEAL, F., 2013. Measuring the Option Value of Education. *Stellenbosch Economic Working Papers*, 15/13.
- BURGER, R.P., VON FINTEL, D., 2006. Rising Unemployment in South Africa Labour Market: A Dynamic Analysis Using Birth Cohort Panels. *Stellenbosch Economic Working Papers*, 24/09.
- CAHUC, P., ZYLBERG, A., 2004. *Labour Economics*, Cambridge, MA: MIT Press
- CARD, D., 1999. The Causal Effect of Schooling on Earnings. In *Handbook of Labor Economics* , Ashenfelter, O., Card, D., (ed.). Amsterdam: North Holland
- CARDOSO, M.F., SALCEDO, R.L., DE AZEVEDO, S.F., 1996. The Simplex-Simulated Annealing Approach to Continuous Non-Linear Optimization. *Computers and Chemical Engineering*, 20(9), 1065-1080.
- CICHELO, P.L., FIEDS, G.S., LEIBBRANDT, M. 2005. Earnings and Employment Dynamics for Africans in Post-Apartheid South Africa: A Panel Study of KwaZulu-Natal. *Journal of African Economics*, 14(2), 143-190.
- CICHELO, P.L., LEIBBRANDT, M., WOOLARD, I., 2014. Winners and Losers: South African Labour-Market Dynamics between 2008 and 2010. *Development Southern Africa*, 31(1), 65-84.
- CHUA, T.C., FULLER, W.A., 1987. A Model for Multinomial Response Error Applied to Labor Flows. *Journal of American Statistical Association*, 82(397), 46-51.
- CLARK, K.B., SUMMERS, L., 1979. Labour Market Dynamics and Unemployment: A Reconsideration. *Brookings Papers on Economic Activity*, 1, 13-60.
- DINKELMAN, T., 2004. How Household Context Affects Search Outcomes of the Unemployed in KwaZulu-Natal, South Africa: A Panel Data Analysis. *South African Journal of Economics*, 72(3), 484-521.
- ECKSTEIN, Z., WOLPIN, K.I., 1999. Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities. *Econometrica*, 67(6), 1295-1339.

- FARHI, E., 2011. The iFit Data Analysis Library: ILL, *Computing for Science*.
 [on-line: <http://ifit.mccode.org>]
- FARHI, E., DEBAB, Y., 2013. iFit: A New Data Analysis Framework. Applications for Data Reduction and Optimization of Neutron Scattering Instrument Simulations with McStas. *Journal of Neutron Research*, 17(1), 5-18.
- FLINN, C., HECKMAN, J., 1982. New Methods for Analyzing Structural Models of Labor Force Dynamics. *Journal of Econometrics*, 18(1), 115-168.
- GRONAU, R., 1971. Information and Frictional Unemployment. *The American Economic Review*. 61(3), 290-301.
- HECKMAN, J.J., 1979. Sample Selection Bias as a Specification Error. *Econometrica*. 46, 1: 153-161
- HECKMAN, J.J., 2010. Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature*, 48(2), 356-398.
- HECKMAN, J.J., SINGER, B., 1984. Econometric Duration Analysis. *Journal of Econometrics*, 24(1984), 63-132.
- HERTZ, T., 2003. Upward Bias in the Estimated Returns to Education: Evidence from South Africa. *American Economic Review*, 93(4), 1354-1368.
- IMAI, S., KEANE, M.P., 2004. Intertemporal Labor Supply and Human Capital Accumulation. *International Economic Review*. 45(2), 601-641.
- JENSEN, R., 2010. The (Perceived) Returns to Education and the Demand for Schooling. *The Quarterly Journal of Economics*. 152(2), 515-548.
- JENSEN, P., WESTERGAARD-NIELSEN, N., 1987. A Search Model Applied to the Transition from Education to Work. *Review of Economic Studies*. 54(3), 461-472.
- KOCH, S.F, NTEGE, S.S., 1987. Potential Screening in South Africa's Labour Market. *SARB Conference 2006*, 143-150.
- KEANE, M.P., 2010. Structural vs. Atheoretic Approaches to Econometrics. *Journal of Econometrics*, 156(1), 3-20.
- KEANE, M.P., WOLPIN, K.I., 1997. The Career Decisions of Young Men. *Journal of Political Economy*, 105(3), 473-522.

- KESWELL, M., POSSWELL, L., 2004. Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. *The South African Journal of Economics*. 72(4), 834-860.
- KIEFER, N.M, NEUMANN, G.R., 1979. An Empirical Job-Search Model, with a Test of the Constant Reservation-Wage Hypothesis. *Journal of Political Economy* 87(1), 89-107.
- KINGDON, G.G., KNIGHT, J., 2001. What Have We Learned About Unemployment from Microdatasets in South Africa? *Social Dynamics*, 27(1), 79-95.
- KRUEGER, A.B., MUELLER, A.I, 2014. A Contribution to the Empirics of Reservation Wages. *National Bureau of Economic Research. working paper 19870*.
- LAM, D., ARDINGTON, C., LEIBBRANDT, M., 2011. Schooling as a Lottery: Racial Differences in School Advancement in Urban South Africa. *Journal of Development Economics*. 95(2), 121-136.
- LAM, D., LEIBBRANDT, M., MLATSHENI, C., 2008. Education and Youth Unemployment in South Africa. A Southern Africa Labour and Development Research Unit Working Paper Number 22. Cape Town: SALDRU, University of Cape Town.
- LANGHEINE, R., 1988. Manifest and Latent Markov Chain Models for Categorical Panel Data. *Journal of Educational Statistics*, 13(4), 299-312.
- LEVINSOHN, J., PUGATCH, T. 2010. The Role of Reservation Wages in Youth Unemployment in Cape Town, South Africa: A Structural Approach. Unpublished, University of Michigan.
- LEWBEL, A., 2000. Identification of the Binary Choice Model with Misclassification. *Econometric Theory*, 16(4), 603-609.
- MAGNAC, T., ROBIN, J.M., 1994. An Econometric Analysis of Labour Market Transitions Using Discrete and Tenure Data. *Labour Economics*, 1(3), 327-346.
- MAGNAC, T., VISSER, M., 1999. Transition Models with Measurement Errors. *The Review of Economics and Statistics*, 81(3), 466-474.
- MAGNAC, T., THESMAR, D., 2002. Analyse économique des politiques éducatives: l'augmentation de la scolarisation en France de 1982 à 1993. *Annales d'Économie et de Statistique*, 66, 1-35.
- MCCALL, J.J., 1970. Economics of Information and Job Search. *The Quarterly Journal of Economics*. 84(1), 113-126.

- MEYER, B.D., 1988. Classification-Error Models and Labor-Market Dynamics. *Journal of Business and Economic Statistics*, 6(3), 385-390.
- MINCER, J., 1974. *Schooling, Experience and Earnings*. New York: National Bureau of Economic Research.
- MLATSHENI, C., ROSPABE, S., 2002. Why is Youth Unemployment so High and Unequally Spread in South Africa? *DPRU Working Paper No. 02/65*.
- MOORE, J.M., VISCUSI, W.K., 1990. Models for Estimating Discount Rates for Long-Term Health Risks Using Labor Market Data. *Journal of Risk and Uncertainty*, 3, 381-401.
- MORTENSEN, D.T., 1970. Job Search, the Duration of Unemployment, and the Phillips Curve. *American Economic Review*, 60(5), 505-517.
- MOLL, P., 1996. The Collapse of Primary Schooling Returns in South Africa 1960-90. *Oxford Bulletin of Economics and Statistics*, 58(1), 185-209.
- MOLL, P., 1998. Primary Schooling, Cognitive Skills and Wage in South Africa, *Economica*, 65(?), 263-84.
- MURPHY, K.M., WELCH, F., 1990. Empirical Age-Earnings Profiles. *Journal of Labor Economics*, 8(2), 202-228.
- MWABU, G., SCHULTZ, T.P., 1996. Education Returns across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa. *American Economic Review*, May 1996 (*Papers and Proceedings*), 86(2), 335-39.
- NATRASS, N., WALKER, R., 2005. Unemployment and Reservation Wages in Working-Class Cape Town. *South African Journal of Economics*, 73(3), 498-509.
- PASERMAN, M.D., 2008. Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation. *Economic Journal*, 118(531), 1418-52.
- PETTERSON, S.M., 1997. Are Young Black Men Really Less Willing to Work? *American Economic Review*, 62(4), 605-613.
- PFEFFEMANN, D., SKINNER, C., HUMPHREYS, K., 1988. The Estimation of Gross Flows in the Presence of Measurement Error Using Auxiliary Variables. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 161(1), 13-32.

- PORTEBA, J.M., SUMMERS, L.H., 1984. Response Variation in the CPS: Caveats for the Unemployment Analyst. *Monthly Labor Review*, 107, 37-42.
- PORTEBA, J.M., SUMMERS, L.H., 1986. Reporting Errors and Labor Market Dynamics. *Econometrica*, 54(6), 1319-1338.
- PSACHAROPOULOS, G., 1979. On the Weak versus the Strong Version of Screening Hypothesis. *Economic Letters*. 4, 181-145.
- RANKIN, N.A., ROBERTS, G., 2011. Youth Unemployment, Firm Size and Reservation Wages. *South African Journal of Economics*. 79(2), 128-145.
- RANCHOD, V., DINKELMAN, T., 2008. Labor Market Transitions in South Africa: What can We Learn from Matched Labor Force Survey Data? *Southern Africa Labor and Development Research Unit Working Paper*, No. 14.
- RASCH, G., 1980. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, IL: The University of Chicago Press. (Original work published in 1960)
- REPUBLIC OF SOUTH AFRICA., 1996. The South African Schools Act. Act No 84 of 1996. Pretoria: Government Printers.
- ROSEN, S., 1977. Human Capital: A Survey of Empirical Research. In *Research in Labour Economics: An annual Compilation of Research*, Ehrenberg, R.G., (ed.). Greenwich: Jai Press, vol 1, 3- 40.
- RUST, J., 1992. Do People Behave According to Bellman's Principles of Optimality? Stanford, CA: Hoover Institution Working Paper no. E-92-10.
- SALDRU (SOUTH AFRICA LABOUR AND DEVELOPMENT RESEARCH UNIT), 2015a. National Income Dynamics Study 2008, Wave 1 (dataset). Version 5.3. Cape Town: Southern Africa Labour and Development Research Unit (producer), 2015. Cape Town: DataFirst (distributor)
- SALDRU (SOUTH AFRICA LABOUR AND DEVELOPMENT RESEARCH UNIT), 2015b. National Income Dynamics Study 2010-2011, Wave 2 (dataset). Version 2.3. Cape Town: Southern Africa Labour and Development Research Unit (producer), 2015. Cape Town: DataFirst (distributor)
- SALDRU (SOUTH AFRICA LABOUR AND DEVELOPMENT RESEARCH UNIT), 2015c. National Income Dynamics Study 2012, Wave 3 (dataset). Version 1.3. Cape Town: Southern Africa Labour and Development Research Unit (producer), 2015. Cape Town: DataFirst (distributor)

- SAIRR (SOUTH AFRICA INSTITUTE OF RACE RELATIONS), 2012. South African Institute of Race Relations survey 2010-2011.
[on-line: <http://irr.org.za/reports-and-publications/south-africa-survey/south-africa-survey-online-2010-2011/education>]
- SASA (SOUTH AFRICAN SCHOOLS ACT), 1996. Act No. 84 of 1996. *Government Gazette. No 17579*. Pretoria: Government Printers.
[on-line: <http://www.gov.za/sites/www.gov.za/files/Act84of1996.pdf>]
- SASSA (SOUTH AFRICAN SOCIAL SECURITY AGENCY), 2008. Annual Statistical Report on Social Grants.
[on-line: <http://www.sassa.gov.za/index.php/knowledge-centre/annual-reports>]
- SINGH, A.C., RAO., J.N.K., 1995. On the Adjustment of Gross Flow Estimates for Classification Error with Application to Data from the Canadian Labour Force Survey. *Journal of the American Statistical Association*, 90, 478-488.
- STATS SA (STATISTICS SOUTH AFRICA), 2006. The South African Labour Force Panel Study – Methodology Document. National Statistics System Division (NSSD), Statistics South Africa.
- STATS SA (STATISTICS SOUTH AFRICA), 2015a. Quarterly Labour Force Survey: Quarter 2: 2015, Statistical Release, P0211
[on-line: <http://www.statssa.gov.za/publications/P0211/P02112ndQuarter2015.pdf>]
- STATS SA (STATISTICS SOUTH AFRICA), 2015b. National and Provincial Labour Market: Youth: Q1: 2008 - Q1: 2015, Statistical Release, P0211.4.2
[on-line: <http://www.statssa.gov.za/publications/P02114.2/P02114.22015.pdf>]
- STIGLITZ, J.E., 1975. The Theory of “Screening” Education, and the Distribution of Income. *The American Economic Review*, 65(3), 283-300.
- TOPEL, R.H., WARD, M.P. 1992. Job Mobility and the Careers of Young Men. *The Quarterly Journal of Economics*, 107(2), 439-479.
- VAN DER BERG, S., 2007. Apartheid's Enduring Legacy: Inequalities in Education. *Apartheid's Enduring Legacy: Inequalities in Education*, 16(5), 849-880.
- WEST, M., 1993. Mixtures Models, Monte Carlo, Bayesian Updating and Dynamic Models. *Computer Science and Statistics*, 24, 325-333.
- WOLFE, R., CARLIN, J.B., PATTON, G.C., 2003. Transitions in an Imperfectly Observed Binary Variable: Depressive Symptomatology in Adolescents. *Statistics in Medicine*, 22(3), 427-440.

WOLPIN, K.I., 1977. Education and Screening. *American Economic Review*, 67(5), 949-958.

WOLPIN, K.I., 1987. Estimating a Structural Search Model: The Transition from School to Work. *Econometrica*, 1987(4), 801-817.

ZOCH, A., 2014. How Reliable are Responses to Survey Questions About Reservation Wages? CSAE Conference Paper, Oxford, 2014.

[on-line: https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CSAE2014&paper_id=896]