# Estimating the shape of the South African schooling-earnings profile

Rulof Burger

Oriel College, Oxford University






Thesis submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy University of Oxford, Trinity Term 2011

# Estimating the shape of the South African schooling-earnings profile

**Rulof Burger, Oriel College, Oxford University**

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, University of Oxford, Trinity Term 2011

## Abstract

This study uses a variety of econometric techniques to estimate the shape of the South African schooling-earnings profile. The two main approaches to measuring the effect of schooling on worker productivity are to either estimate a production function that compares how output responds to the hiring of better educated workers, or to estimate an earnings regression that reveals the effect of an additional year of schooling on an individual's expected earnings. Both techniques are vulnerable to econometric issues such as unobserved productivity differences, sample selection, non-linearities in the schooling-productivity profile, heterogeneous parameters and measurement error. Within the class of earnings functions, two very different methods have been developed to deal with endogenous schooling: instrumental variable techniques and discrete choice dynamic programming models. This study estimates the South African schooling-earnings profile using a wide range of identification strategies and estimators, and investigates whether accounting for certain econometric issues can produce consistent estimates of this profile across different approaches.

Chapter 2 uses instrumental variable techniques to estimate the causal effect of schooling on the earnings of black South African men in the period directly following political transition. It explores different ways to identify the shape of the schooling-earnings profile in the presence of endogenous schooling, heterogeneity in the schooling returns parameter and sample-selection bias due to the highly selective employment process. Our analysis shows that the presence of sample selection changes both the conditions under which the parameters of interest can be estimated, as well as the interpretation of these estimates. The empirical results show that the schooling returns in South Africa are high, on average, and an increasing function of the level

of schooling. However, the degree of convexity estimated by the control function approach is considerably less than that suggested by either ordinary least squares or two-stage least squares, especially when also correcting for sample selection bias or allowing for a non-linear relationship between schooling and market unobservables.

Many recent descriptive studies find convex schooling-earnings profiles in African countries. Apart from the potential endogeneity issues that may confound such estimates, it also is not clear whether a convex earnings profile can be reconciled with the fact that most individuals choose to obtain intermediate schooling outcomes. Chapter 3 uses dynamic programming techniques to investigate the schooling decisions of rational, dynamically optimising agents in a model that seeks to incorporate key aspects of the South African labour market. Our results produce a schooling-earnings profile which is convex, although the degree of convexity is less than suggested by previous studies that use OLS to estimate this relationship. Furthermore, the marginal cost of schooling is required to be a steeply increasing function of schooling years if schooling decisions are to be reconciled with convex earnings profile without resorting to assumptions of irrational expectations, imperfect information or schooling restrictions. Our analysis indicates that the main benefit of completing early schooling years is that it allows individuals access to more advanced schooling years for which the benefits are much higher.

Chapter 4 estimates production functions for South African industries using a novel panel dataset that combines education and employment data from a series of household surveys with output and physical capital data from the South African Reserve Bank Quarterly Bulletins. The analysis starts by exploring the estimates produced by the pooled ordinary least squares (POLS) estimator, and then proceeds to alternative estimators that exploit less restrictive identifying assumptions that allow for industry fixed effects, endogeneity in the factors of production, measurement error, cross-sectional dependence and parameter heterogeneity. The POLS results indicate that the returns to education are substantial and concave. Although the effect of education disappears when industry fixed effects are included in the regression, once an allowance is made for measurement error, parameter heterogeneity and cross-sectional dependence, the results are not significantly different from POLS.

# Contents


1. Introduction ..... 1
2. Estimating the shape of the schooling-earnings profile using IV techniques ..... 8
   2.1 Introduction ..... 8
   2.2 Identifying the causal effects of schooling ..... 10
   2.3 Endogeneity and the returns to education ..... 16
      2.3.1 Econometric problems with the Mincerian earnings regression ..... 17
      2.3.2 The local average treatment effect interpretation ..... 22
      2.3.3 Average schooling-earnings profile interpretation ..... 27
   2.4 Data and instrumental variables ..... 34
      2.4.1 Policy shifts and the identification of schooling returns ..... 34
      2.4.2 Data ..... 40
      2.4.3 Instrumental variables ..... 45
   2.5 Econometric analysis ..... 52
      2.5.1 Schooling and employment regressions ..... 52
      2.5.2 Earnings regressions ..... 61
      2.5.3 Sensitivity analysis ..... 71
   2.6 Conclusions ..... 73
3. A discrete choice dynamic programming model of the South African schooling-earnings nexus ..... 75
   3.1 Introduction ..... 76
   3.2 Discrete choice dynamic programming studies of schooling returns ..... 77
      3.2.1 The strengths and weaknesses of structural modeling ..... 79
      3.2.2 The Belzil and Hansen model ..... 82
   3.3 Schooling, earnings and employment in the South African labour market ..... 87
   3.4 Estimating discrete choice dynamic programming models ..... 91
      3.4.1 Structure of discrete choice dynamic programming models ..... 91
      3.4.2 Solving the agent's dynamic optimisation problem ..... 93
      3.4.3 Estimating the structural parameters ..... 94
   3.5 A discrete choice dynamic programming model of the South African schooling-earnings nexus ..... 97
      3.5.1 Model assumptions ..... 97
      3.5.2 Solving the agent's dynamic optimisation problem ..... 100
      3.5.3 Estimating the structural parameters ..... 102
   3.6 Empirical analysis ..... 103
      3.6.1 Data ..... 103
      3.6.2 Estimation results ..... 104
   3.7 Conclusion ..... 114

4. Estimating the schooling-productivity profile using the production function approach ..... 115
4.1 Introduction ..... 115
4.2 The effect of education on labour productivity ..... 116
4.3 Data ..... 121
4.4 Production function and model identification ..... 123
4.4.1 Pooled OLS and random effects estimators ..... 126
4.4.2 Fixed effects, time effects or two-way fixed effects estimators ..... 127
4.4.3 Parameter heterogeneity ..... 129
4.4.4 Cross-sectional dependence ..... 131
4.4.5 Measurement error ..... 134
4.5 Empirical results ..... 137
4.5.1 Measurement error ..... 141
4.5.2 Parameter heterogeneity ..... 145
4.5.3 Cross-sectional dependence ..... 147
4.6 Conclusion ..... 149
5. Conclusion ..... 151
6. Bibliography ..... 155
7. Appendix ..... 170
7.1 First-stage schooling regression ..... 170
7.2 Earnings regressions ..... 172
7.2.1 The local average treatment effect interpretation ..... 172
7.2.2 Average schooling-earnings profile interpretation ..... 173
7.3 Sensitivity analysis ..... 175
7.4 Production function estimates ..... 177
7.5 List of frequently used abbreviations ..... 181

## List of Figures

Figure 1: Real value of grants (2000 prices), by sample year ..... 40
Figure 2: Share of grant receivers, by sample year ..... 40
Figure 3: Schooling outcomes for black males, by birth year ..... 45
Figure 4: Share of black males with completed secondary education, by birth year and province ..... 46
Figure 5: Share of over-aged learners, by birth year and province ..... 51
Figure 6: Estimated LATEs of schooling for black males, by schooling year ..... 65
Figure 7: Estimated marginal returns to schooling for black males, by schooling year ..... 71
Figure 8: Mean log earnings and formal employment share for black males, by schooling year ..... 88
Figure 9: Schooling histogram for black males, 2001-2004 ..... 89
Figure 10: Schooling frequency density of black males, by birth decade, 2001-2004 ..... 91
Figure 11: Predicted schooling frequencies of black males, by birth decade, 2001-2004 ..... 107
Figure 12: Decomposition of expected value of additional schooling year ..... 111
Figure 13: Scatterplot of heterogeneous schooling, wage and employment intercepts ..... 113

## List of Tables

Table 1: First-stage OLS regressions on years of education (sample of black males), 1995-2007 ..... 55
Table 2: Probit regressions on crossing various education thresholds (sample of black males), 1995-2007 ..... 58
Table 3: Formal sector employment probit regressions (sample of black males), 1995-2007 ..... 59
Table 4: OLS and 2SLS earnings regressions using different IVs (sample of black males), 1995-2007 ..... 62
Table 5: 2SLS education coefficients using different instrumental and conditioning variables (sample of black males), 1995-2007 ..... 64
Table 6: Earnings regressions comparison of OLS, 2SLS and CF estimators (sample of black males), 1995-2007 ..... 67
Table 7: CF earnings regressions with different selection controls (sample of black males), 1995-2007 ..... 70

Table 8: DCDP coefficient estimates ..... 105
Table 9: Heterogeneous intercept estimates from DCDP model ..... 112
Table 10: Estimates of production function coefficients, using various panel data estimators ..... 138
Table 11: Variance decomposition for factors of production ..... 141
Table 12: Panel data estimates of production function coefficients ..... 142
Table 13: FE and FD production function regressions with fixed employment coefficient values ..... 143
Table 14: Various heterogeneous parameter panel data estimates of production function coefficients ..... 148
Table 15: First-stage OLS regression on years of education (sample of black males), 1995-2007 ..... 171
Table 16: OLS and 2SLS earnings regressions using different IVs (sample of black males), 1995-2007 ..... 173
Table 17: CF earnings regressions with different education residual controls (sample of black males), 1995-2007 ..... 174
Table 18: Earnings regressions comparison of OLS, 2SLS and CF estimators (sample of black males), 1995-2007 ..... 175
Table 19: CF earnings regressions with different birth year controls (sample of black males), 1995-2007 ..... 175
Table 20: CF earnings regressions with different control and instrumental variables (sample of black males), 1995-2007 ..... 176
Table 21: POLS estimates of production function coefficients, with various human capital link functions ..... 177
Table 22: OLS time series regression estimates of production function, by industry ..... 177
Table 23: IV-TE estimates of production function coefficients ..... 178
Table 24: IV-FE estimates of production function coefficients ..... 179
Table 25: IV-FD estimates of production function coefficients ..... 180

**To my wife Ansa**

**For all your loving support and encouragement.**

# Acknowledgements

This thesis would not have been possible without the support of so many people. Firstly, I am indebted to members of the Oxford University Economics department and the Centre for the Study of African Economies, particularly Marcus Eberhardt, Victoria Prowse, Andrew Kerr, Simon Quinn, John Knight, Andrew Zeitlin, Alex Sienaert, Alberto Behar, Bob Rijkers, Matthew Pollison, Howard Smith and Geeta Kingdon.

I also wish to thank my family: my wife Ansa, for her loving support and encouragement; my siblings and colleagues Ronelle and Cobus, for their advice during many discussions of my research, assistance in proof-reading and emotional support; and my parents, for their love and inspiration.

My colleagues at the University of Stellenbosch also deserve acknowledgement: Servaas van der Berg, who planted the seed of my interest in labour economics as an undergraduate students; Stan du Plessis, who convinced me to continue my graduate studies in the UK; and Andrie Schoombee for his willingness to accommodate my studies as head of department.

I would also like to acknowledge the financial support I received from the Clarendon scholarship, the H.B. and M.J. Thom Scholarship, the Herbert and Ilse Frankel Scholarship, as well as Stellenbosch University's Economics Department.

I reserve my most grateful thanks to my supervisor, Francis Teal, whose patient and insightful guidance has been invaluable in this process.

Thank you.

# 1. Introduction

Economists have long been interested in the causal effect of schooling on earnings. The positive correlation between schooling and earnings has been called the "second ... most well-established fact in economics" (Pritchett, 2001, p. 368)[1], and has now been verified in hundreds of studies spanning many decades and countries[2]. On the basis of such empirical research George Psacharopoulos argues that the "returns"[3] to investment in schooling is high on average, and higher still when undertaken in less developed countries or for lower levels of education.

Such research has prompted many policy documents – including the Millennium Development Goals (United Nations, 2000), the Washington Consensus (Williamson, 1990), and the New Economic Plan for African Development (NEPAD, 2001) – to prescribe investment in education as a means for poor countries to put themselves on the road to economic development, while also emphasising the importance of primary schooling over secondary or tertiary education. African countries have followed these recommendations in terms of setting government expenditure priorities (Collier & Gunning, 1999, p. 71) and this has had the desired effect of improving schooling access for its citizens (UNESCO, 2011). However, such improvements have often failed to translate into the expected acceleration in economic growth. In a similar vein, cross-country economic growth regression typically struggle to find a direct human capital effect on worker productivity that corresponds to the magnitude of effects observed for individual earnings equations (Benhabib & Spiegel, 1994; Krueger & Lindahl, 2001; Temple, 1999). This has been called the *micro-macro paradox* (Pritchett, 2001).

There are many reasons why the schooling coefficient from a Mincerian earnings regression may not be interpretable as an unbiased estimate of the causal effect of additional schooling on labour productivity. The concern that the robust statistical correlation between schooling and earnings may reflect inherent

1 Pritchett (2001, p. 368) believes that the distinction of the most well-established fact is held by Engel's law.
2 See Cohn and Addison (1998) and Psacharopoulos and Patrinos (2004) for surveys of recent studies.
3 This terminology is not entirely appropriate when referring to the causal effect of schooling on wages in a context that does not take the explicit cost of schooling investment into consideration. However, its use is so well-established that we will go along with the convention of using "returns to schooling" to denote the proportional increase in earnings associated with an additional year of schooling.

pre-schooling productivity differences rather than the causal effect of schooling is almost as old as the empirical human capital literature itself. If the reason for the positively sloped schooling-earnings profile is that individuals with higher market ability also choose to obtain more schooling years, then Africa's attempts to jump start growth by improving access to education is ill-fated. In this case schooling will be endogenous in an earnings equation and an ordinary least squares regression will provide misleading estimates of the true causal effect of schooling on worker productivity.

Two different approaches have emerged to address concerns regarding the endogeneity of schooling in the earnings equation. The first and more popular approach uses instrumental variables to isolate the schooling variation that is orthogonal to market abilities. Any correlation between earnings and this exogenous schooling variation will then reflect the causal effect of schooling without being biased by unobserved abilities. The last two decades have seen numerous studies use natural experiments which induced exogenous variation in schooling outcomes, and the estimated returns are generally slightly higher than those obtained from Mincerian regressions (Ashenfelter, Harmon, & Oosterbeek, 1999). The second method estimates discrete choice dynamic programming models of the schooling-earnings relationship. Such studies are much less common and have not been used to estimate the schooling returns for an African country before. The evidence from estimating such models on US data suggests that the schooling returns may be considerably lower than found by Mincerian or instrumental variable studies (Belzil, 2007).

Apart from providing an interesting comparison of the schooling return estimates obtained from different identifying assumptions, these two approaches are also complementary. The fact that instrumental variable techniques can provide estimates of the causal effect of schooling with very few behavioural assumptions means that it is relatively robust compared to other methods. On the other hand, since these techniques do not usually specify a full behavioural model they are often uninformative about the channels through which schooling will affect labour market outcomes. Estimating discrete choice dynamic programming models requires a completely specified, internally consistent model which can answer some of the questions that instrumental variable methods cannot.

The schooling variable may also be endogenous if individual educational outcomes are measured with error, if the sample of wage earners is not randomly selected from the population of working-age adults or if the effect of education varies across individuals. Addressing measurement error, the omission of unobserved market ability or firm productivity, the heterogeneity of schooling returns or selection into employment each requires different techniques, and very few studies have investigated how to estimate the causal effect of schooling on earnings if the data is affected by more than one of these problems at once.

Furthermore, there is a growing body of empirical evidence to suggest that the schooling-earnings profiles in African countries are convex. Instrumental variable methods have often ignored the issue of non-linearities in the schooling returns, although Card's (1999; 2001) re-interpretation of this literature asserts that such non-linearities are crucial in understanding their results. However, estimating the shape of the schooling-earnings profile with instrumental variable methods – without resorting to a homogenous coefficient assumption – is not a straightforward task. Heckman, Lochner & Todd (2008) also survey the instrumental variable returns to schooling literature, and conclude that the link between the estimation and the theory is generally weak. They recommend using a structural approach that explicitly allows for non-linearities in the schooling-earnings profile as well as uncertainty and the sequential revelation of information. The option value of schooling that can appear under these conditions is potentially important in understanding the schooling decisions of optimising agents.

The instrumental variable and discrete choice dynamic programming techniques are able to facilitate these statistical issues to varying degrees. The local average treatment effect interpretation of two-stage least squares estimates was specifically developed to incorporate heterogeneity, unobserved abilities and measurement error, but usually ignores the issues of sample selection and non-linearities in the schooling-earnings profile. In contrast, dynamic programming techniques are able to model unobserved abilities, non-linearities and selection into employment in a very flexible manner, but find it more difficult to incorporate measurement error and heterogeneous returns.

Apart from the earnings function method described above, the effect of schooling on worker productivity can also be estimated via a production functions approach. This method compares the productivity of similar firms with different average levels of schooling amongst their workers, and so offers a more direct approach to measuring of the causal effect schooling on labour productivity. Empirical production function studies on African firm- or sector-level data have generally been unable to replicate the large education effects estimated in African earnings regressions, which presents a puzzle very similar to the micro-macro paradox. One explanation is the existence of labour market institutions that drive a wedge between the marginal productivity of workers and their remuneration. Alternatively, the production function approach may be similarly vulnerable to different types of omitted variables, measurement errors and aggregation bias, and could therefore produce inaccurate measures of the returns to schooling investment.

This dissertation will use three different approaches and a myriad of estimators to estimate the schooling-earnings profile for black[4] South African men in the period directly following political transition. These estimates can be used to calculate the average returns to investing in schooling, and whether the returns are usually lower or higher at more advanced schooling years. We are particularly interested to see whether the divergent results reported in the literature for different approaches can be explained by addressing certain econometric shortcomings. The different techniques used also provide information about the relationship between the different types of unobserved heterogeneity which, if left unaccounted for, could bias our schooling returns estimates. Perhaps the most distinct feature of the South African labour market is its highly selective employment process, which presents another potential source of bias. We will investigate how schooling affects an individual's probability of finding work, and also how selection into employment can be incorporated into our estimates so as not to bias the schooling returns estimates.

[4] The South African government defines four main South African population groups (as in the Employment Equity Act (Act 55 of 1998)): African, Coloured (mixed race), Indian and White. In government documents, "Blacks" is often used to include Africans, Coloureds and Indians, although in this thesis we will use black to refer to the African group only.

Chapter 2 investigates the schooling returns using individual-level data and instrumental variable techniques. It starts by discussing the literature on instrumental variable techniques and some of the shortcomings of the current local average treatment effect interpretation of two-stage least squares estimates. We propose two approaches to estimate the shape of the schooling-earnings profile for black South African men in the presence of heterogeneous returns and omitted market ability: comparing the local average treatment effects of different instrumental variables and estimating the parameters from a more restrictive Becker-Card model. Our analysis also explores how the presence of a highly selective employment process could bias these estimates, and the conditions under which this issue can be addressed.

Chapter 3 estimates a discrete choice dynamic programming model of the schooling choices and labour market outcomes of black South African men. The model predictions are shown to fit the observed decisions of actual learners relatively well. We also develop a decomposition technique that demonstrates why rational forward-looking agents would make the observed schooling decisions given their expected employment and earnings outcomes.

Chapter 4 estimates production function coefficients using a novel industry-level panel dataset constructed from different data sources. We report results from a variety of estimators that will provide unbiased estimates of the production parameters in the presence of industry productivity differentials, parameter heterogeneity, cross-sectional dependence and measurement error. Finally, chapter 5 concludes.

# 2. Estimating the shape of the schooling-earnings profile using IV techniques


## Abstract

This chapter uses various instrumental variable techniques to estimate the causal effect of schooling on the earnings of black South African men in the period directly following political transition. It explores the identifying assumptions and most appropriate interpretations for different estimators that could identify the shape of the schooling-earnings profile in the presence of endogenous schooling, heterogeneity in the schooling returns parameter and sample-selection bias due to the highly selective employment process. Our analysis shows that the presence of sample selection changes both the conditions under which the parameters of interest can be estimated, as well as the interpretation of these estimates. The estimators all confirm that schooling returns in South Africa are high on average. Most estimators also suggest that returns are higher at more advanced schooling levels, although the degree of convexity estimated by the control function approach is considerably less than that suggested by either ordinary least squares or two-stage least squares, especially when also correcting for sample selection bias or non-linear interaction between the schooling and market ability terms. Our results also show that mistakenly assuming that schooling is exogenous in the employment process tends to exaggerate the convexity in the schooling-employment profile.


## 2.1 Introduction

This chapter estimates the shape of the schooling-earnings profile for black South African men during the period directly following political transition. It seeks to identify the causal effect of the different schooling years on expected earnings using three policy shifts – the rollback of apartheid era restrictions on black access to schooling, the stricter enforcement of restrictions on over-aged learners and an increase in the

value and coverage of social grants – that produced exogenous variation in individuals' schooling and participation decisions. Recent advances in microeconometrics have shown how to estimate different weighted averages of schooling returns in the presence of endogenous schooling and heterogeneous returns. However, we estimate the slope of the schooling-earnings profile by comparing the local average treatment effects for instruments that affected schooling decisions at different points of the schooling ladder, as well as estimating the parameters from the Becker-Card model which allows us to plot the expected schooling-earnings profile for a randomly chosen individual. The latter approach requires estimating the average treatment effect of the endogenous schooling variable while simultaneously controlling for a quadratic schooling term, and is estimated using two-stage least squares and a control function regression. Our analysis considers the identifying assumptions required for all of the different estimators as well as the most sensible interpretation of their estimates. We argue that our instruments changed the cost of education at various schooling levels, thus providing us with a unique opportunity to study the shape of the schooling-earnings profile over the whole range of schooling outcomes. We also explore how the highly selective employment process that is an important feature of the South African labour market can affect the identifying assumptions and interpretation of the schooling returns estimates.

The chapter is organised as follows: Section 2.2 reviews the literature and highlights some of the econometric problems associated with estimating the returns to schooling. A theoretically tractable estimable earnings equation that allows for endogenous schooling, heterogeneous returns and sample selection is derived in section 2.3, before discussing estimators that could conceivably estimate the shape of the schooling-earnings profile under various identifying assumptions. Section 2.4 introduces three policy shifts and the data that are used to construct instrumental variables and exclusion restrictions and section 2.5 reports and discusses the results of the various estimators. Section 2.6 concludes.

## 2.2 Identifying the causal effects of schooling

In order to consider the causal effect of an additional year of schooling $s$ on the earnings of agent $i$ in period $t$, we write the structural equation for log hourly earnings as

$$w_{it}(s) = \boldsymbol{z}_{it1}\boldsymbol{\delta} + f_i(s) + a_{it} \quad [1]$$

where $\boldsymbol{z}_{it1}$ is a vector of observable characteristics that affect individual earnings and $a_{it}$ represents the composite effect of unobservable wage determinants. $f_i(.)$ is an individual-specific function that maps the different schooling years onto expected earnings, and therefore represents the schooling-earnings profile which is the primary concern of this dissertation.

Most early studies of the schooling returns used some variant of Mincer's (1974) human capital earnings function to estimate the returns to education using OLS while controlling for characteristics such as experience, demographic factors and geographical controls. This specification replaces the schooling function in equation [1] with $f_i(s) = bs$, where the estimate of $b$ is then usually interpreted as the causal effect of an additional year of education on labour market earnings. Ashenfelter et al. (1999) find that the cross-country average for such OLS estimates is between 6 and 7%, whereas the return in South Africa (as surveyed in Keswell and Poswell (2004)) is usually found to be between 15% and 30%.

The econometric shortcomings of using Mincerian earnings regressions to estimate the returns to schooling are well documented (Denison (1964), Gronau (1974), Lemieux (2006), Heckman, Lochner, & Todd (2006), to mention a few), albeit still frequently disregarded. In our own analysis in section 2.5 we specifically focus on four of these shortcomings, each of which may bias the estimate of the average schooling returns: i) the selective process through which individuals enter the sample of wage earners, ii) the possible correlation of schooling and unobserved ability, iii) the variation of returns to schooling across schooling years, and iv) the variation of returns to schooling across individuals. The Mincerian approach has also been criticised for restricting its attention to the effect of schooling attainment (Hanushek & Woessmann, 2008), whereas much of the recent developing country policy debate has focussed on

improving the quality of schooling. Although our own analysis will discuss the potential biases introduced by differential school quality, data limitations preclude us from explicitly modelling both schooling quantity and quality.

Gronau (1974) first noted the problems posed by sample selection bias in the context of a female earnings regression where wage data are only available for those women who received acceptable wage offers. He recognised that female labour force participants differed systematically from those who chose not to participate in the labour market, and that these differences would yield biased estimates of the effects of the various wage offer determinants. In response to this problem James Heckman produced first a maximum likelihood estimator (Heckman, 1974) and then a two-step procedure (Heckman, 1979) that are able to recover unbiased estimates of the equation of interest in the presence of certain sample selection processes. Specifically, proper identification of the parameters of interest require observing variables that determine selection into employment without affecting earnings, exogeneity of all the regressors in both the selection and earnings equations, a distributional assumption regarding the unobservable determinants of selection and knowledge about the relationship between the unobservables in the selection and earnings equations up to some estimable parameter. Concerns that certain distributional assumptions are too strong in many applications have led to the development of semi-parametric variants of this technique (as surveyed in Vella (1998)) that do not require distributional assumptions for identification. However, these methods still require that all employment and wage regressors are exogenous, which is perhaps a more serious concern than the non-normality of the unobservables in the context of an earnings regression.

The last two decades have seen a dwindling of interest in sample selection techniques (Moffitt (1999, p. 1372)). Heckman (1990) acknowledges that in many applications simpler econometric techniques may provide equally reliable estimates. Imbens (2010) ascribes the emergence of the experimentalist approach to the "failure of specific structural models, such as the Heckman selection model ... to satisfactorily address endogeneity issues in the context of estimation of causal effects of labor market programs". As a result, almost all studies that use instrumental variable techniques to estimate the returns to schooling ignore the issue of selection into employment. Although this is perhaps a justifiable omission when

studying high participation, low unemployment groups – such as men in developed countries – applying the same approach to developing countries where a relatively small share of workers are employed in the formal sector runs the risk of producing substantially biased coefficient estimates.

The concern that the robust statistical correlation between schooling and earnings may reflect inherent pre-schooling productivity differences rather than the causal effect of schooling (Denison, 1964; Spence, 1973) is almost as old and as well-established as the empirical human capital literature itself (Mincer, 1958; Becker G. S., 1964). This criticism has prompted a burgeoning literature that uses instrumental variable (IV) techniques to address the bias that arises from not being able to control for the wage effect of unobservable abilities. These studies usually exploit supply-side determinants of the individual's education investment decision, such as institutional characteristics of the education system, to obtain variation in schooling that is orthogonal to unobserved abilities. It is argued that any correlation between such exogenous schooling shocks and earnings must reveal the true causal effect of additional schooling investment. Some of the instrumental variables that have been used to identify the causal effect of schooling includes quarter of birth (Angrist & Krueger, 1991), distance to the nearest school (Kane & Rouse, 1995), changes in the minimum school leaving age (Harmon & Walker, 1995), and region and time variation in school construction (Duflo, 2001). Ashenfelter et al. (1999) find that IV estimates are on average almost 2% points higher than OLS estimates of the same equation and with the same data, which appears to contradict concerns that OLS estimates are upwardly biased owing to the omission of unobserved ability from such regression. This result is often explained by invoking Grilliches' (1977) argument that the misreporting of true schooling levels induces a downward attenuation bias in the Mincerian schooling returns[5]. If this downward bias exceeds in magnitude the positive ability bias, then IV techniques should find estimates that are higher than the corresponding OLS estimates.

Most of the earlier returns to schooling studies attempt to estimate *the* returns to education, implicitly based on the notion that this rate is identical across different individuals and levels of schooling. The

[5] Card (1999, p. 1816) casts some doubt over the validity of this interpretation by showing that the size and mean-regressive nature of schooling measurement error implies that it is unlikely to bias the schooling coefficient by more than 10%.

recent literature has relaxed this assumption in two ways: allowing for a non-linear but deterministic schooling-earnings profile or completely unrestricted heterogeneity across individuals. The first line of research departs from Mincer's (1974) standard[6] human capital earnings function by allowing schooling to enter the earnings function in a more flexible manner, usually as a quadratic or spline function. On the basis of empirical research that spans three decades George Psacharopoulos (Psacharopoulos, 1985; 1973; 1994; Psacharopoulos & Patrinos, 2004) argues that, within countries, returns are higher at lower levels of education: the schooling-earnings profile is concave. However, this characterisation of the earnings profile has recently come under intense scrutiny. Mincer (1996), Heckman, Lochner, & Todd (2008) and Lemieux (2006) all find that wages in the US has been increasingly convex since 1980. Banerjee and Duflo (2005) question the quality of data from which Psacharopolous draws his inference, and after omitting those countries considered to have "poor" or "very poor" quality data by Bennell (1996) the results are found to differ from the conventional pattern suggested by Psacharopoulos. The evidence for developing countries has been particularly conclusive in its rejection of concavity: convexity has been found in Mexico (Binelli, 2008), Columbia, Hong Kong and Kenya (Carnoy, 1995), Botswana (Siphambe, 2000), Zambia (Nielsen & Westergard-Nielsen, 2001), Ghana, Egypt, Côte d'Ivoire, Kenya and Tanzania (Appleton, Hoddinott, & MacKinnon, 1996; Teal, 2001; Whaba, 2000) and South Africa (Keswell & Poswell, 2004). However, most of these studies use estimators that require schooling to be exogenous in the earnings equation, and little work has been done to investigate how the shape of the schooling-earnings function is biased by schooling endogeneity.

The second branch of the heterogeneous returns literature entails a more dramatic departure from the conventional Mincerian approach by allowing completely unrestricted heterogeneity in the education returns across individuals and years of schooling. This approach follows a trend in microeconometrics that uses instrumental variables and very few behavioural assumptions to identify the local average treatment effect (LATE) of an endogenous regressor. In the context of our research question, the LATE can be

[6] Mincer (1974) experimented with alternative functional forms, including one that included the years of schooling quadratically. However, it is the earnings function with a linear schooling variable that would come to be the workhorse of empirical labour economics and bear his name.

interpreted as the weighted average of the marginal schooling returns for those who alter their schooling investment decisions due to a change in the IV values. Card (1999), who re-interprets much of the IV returns to education literature through the lens of LATEs, concludes that the IV estimates are often higher than OLS estimates because the instruments have a larger effect on those with lower levels and hence higher returns to schooling. However, this interpretation clearly requires that the schooling-earnings profile is concave.

Although this IV procedure has dominated the returns to education literature since the 1990s, some prominent detractors have expressed concern about the use of an approach that is largely driven by the availability of interesting natural experiments rather than important policy questions. Heckman and Urzúa (2010, p. 28) state that economists should attempt to "solve well-posed economic questions with minimal assumptions", whereas the LATE is "often very difficult to interpret as an answer to an interesting economic question". Deaton (2009) adds that "[t]he LATE may or may not be a parameter of interest … and in general there is no reason to suppose that it will be." The experimentalist research agenda has also been criticised for only being loosely based on economic theory (Keane, 2010). Although the reduced form equations are often discussed with references to economic models, the estimable model is rarely derived directly from behavioural equations. Furthermore, the pre-eminence it affords to the internal validity of the estimates often comes at the cost of external validity. Since the estimated treatment effect is expressed in terms of the individual effects of a specific (pseudo-) experiment rather than structural parameters, the results are usually more difficult to translate into predictions of the effects of other policies or even the same policy implemented under different circumstances.

On the other hand, the experimentalists refer to the strong and often untested behavioural assumptions that have in the past prevented structural econometricians from unearthing the true effects of policy interventions – a prominent study by Lalonde (1986), which demonstrated that structural models could not replicate the results obtained from the experimental evaluation of job training programs, is often cited in this regard. Furthermore, although the LATE is criticised for only identifying the average treatment effect

for a subsample of observations that the econometrician may have no specific interest in, the proponents of this technique point out that identifying the effects for any other individuals would necessarily require assumptions that effectively extrapolate the LATE to observations that were unaffected by the instrument and for which the treatment effect is therefore unobserved. In that sense, the LATE interpretation is an honest reflection of the limitations of the data.

A series of papers by Garen (1984), Wooldridge (1997; 2003) and Heckman & Vytlacil (1998) have investigated the conditions under which IV techniques can be used to estimate the average treatment effect (ATE), the causal effect averaged over the population as a whole. These conditions are discussed in more detail in section 2.3.3.1, but are considerably more heroic than the identifying assumptions required to estimate the LATE. The ATE has the benefit of being more readily interpretable than the LATE, although it is still not as informative as a full set of structural parameter estimates. Another benefit of the ATE is that is can be used to estimate the slope of the schooling-earnings profile, something which a regular LATE is completely uninformative about.

Although the number of studies that use IV methods to estimate the returns to schooling have grown exponentially since the first such studies in the early 1990s, and are now available for a large number of countries and periods, there are clearly still a number of shortcomings that beset this literature. In the international literature very few studies have attempted to simultaneously control for endogeneity and sample selection bias. The ones that do (García, Hernández, & López-Nicolás (2001), Arabsheibani & Mussurov (2007), Chen & Hamori (2009)) assume parameter homogeneity and usually omit schooling from the employment equation. In terms of the four shortcomings listed above, we can identify these two strands as addressing either endogeneity and returns that vary between schooling years and individuals (but neglecting sample selection), or endogeneity and sample selection (but ignoring heterogeneity). To our knowledge, no study has yet explored how to estimate the schooling-earnings profile in the presence of endogeneity, heterogeneity and sample selection. Although most of the South African empirical literature

uses estimators that allow for sample selection and non-linear returns it has almost uniformly ignored the issues of endogenous schooling and returns that vary across individuals.

## 2.3 Endogeneity and the returns to education

The aim of this chapter is to estimate the average causal effect that the different schooling years have on an individual's earnings capacity. Our literature survey highlighted four econometric problems that we should consider when estimating the coefficients of the earnings function: the potential endogeneity of schooling, returns that are non-linear in schooling, returns that vary across individuals, and sample selection. In order to accurately interpret the estimation results it is useful to derive the estimable equations from an explicit theoretical model (Heckman, 2010). Towards this end we briefly review two models that demonstrate the effect that endogeneity, heterogeneity and sample selection can have on the coefficient estimates of a Mincerian earnings regression. The first is the model presented in Card (1999; 2001), which allows for endogenous schooling decisions and a non-linear schooling-earnings profile that varies across individuals. The second is the model of selection into employment first proposed by Gronau (1974) and operationalised by Heckman (1979).

After investigating the econometric issues that can plague estimates of the schooling returns, we proceed to explore the different sets of identifying assumptions that would allow us to identify the average schooling-earnings profile in their presence. The primary interest of this paper is in estimating the shape of the schooling-earnings profile, and particularly in determining whether the marginal returns generally increase, decrease or are unchanged at higher schooling years. We explore two ways to answer this question when returns are heterogeneous and schooling is endogenous. The first is to estimate the LATE with instruments that have different effects on the various schooling years. If the schooling-earnings profile is indeed convex, then we would expect the instruments that have a stronger effect on higher schooling years to produce larger LATEs than those that mainly affect lower schooling years. The second is to estimate the parameters of the Becker-Card model, in which the marginal return is a linear function of

schooling with a constant slope but different intercepts across individuals. These two strategies are discussed in sections 2.3.2 and 2.3.3.

### 2.3.1 Econometric problems with the Mincerian earnings regression

The literature survey in section 2.2 refers to the large number of studies that use the Mincerian earnings function to estimate the return to investment in schooling. This model is based on the estimable equation

$$w_{it} = \boldsymbol{z}_{it1}\boldsymbol{\delta} + bs_i + a_{it} \quad [2]$$

where $\boldsymbol{z}_{it1}$ includes a constant, years of experience and experience squared, demographic characteristics and geographical controls. Ordinary least squares (OLS) can provide unbiased estimates of the vector of coefficients $(\boldsymbol{\delta}, b)$ if $E(a_{it}|\boldsymbol{z}_{it1}, s_i) = 0$, which requires that schooling is exogenous in the earnings regression. Furthermore, this model implicitly assumes that the schooling returns, $b$, are constant across individuals and different schooling years. Some studies choose to circumvent the restrictiveness of this assumption by including schooling as a quadratic or a spline function, but this still does not allow the returns to vary across different individuals. Finally, it is also assumed that the wage earners on which the earnings regression is estimated are randomly sampled from the population as a whole. As we will see below, these assumptions may lead us to draw invalid inferences regarding the causal effect of an additional schooling year on an individual's expected earnings.

#### *2.3.1.1 The Becker-Card model of endogenous schooling and heterogeneous returns*

In a pair of seminal articles that aim to explain the high schooling return estimates obtained using IV methods (compared to returns estimated with OLS), Card (1999; 2001) derives an economically tractable model of the relationship between earnings and schooling by building on ideas first introduced by Becker (1967) in his Woytinsky lecture. This model consists of individuals who are heterogeneous in both

opportunity and ability and who make decisions in order to maximise their utility functions subject to an intertemporal budget constraint. Under the maintained assumption that the conditions for an interior solution are met a representative agent (denoted $i$) will continue investing in schooling, $s_i$, until the marginal cost and marginal benefit are equated. Log earnings is additively separable in schooling and experience – which means that the earnings equation can be expressed as in equation [1] – and both the marginal returns and marginal costs are linear functions of schooling. Card (2001, p. 1811) further assumes that "either because of credit market considerations or taste factors" the marginal cost of schooling is an increasing function of the years of schooling, $r_i + \kappa_2 s_i$ where $\kappa_2 > 0$, and that proportional wage growth is a linearly decreasing function of schooling, $b_i - \kappa_1 s_i$, where $\kappa_1 > 0$. Individuals are heterogeneous in both the cost of schooling that they face and the marginal benefit to schooling, which is what Becker (1967) refers to as "inequality in opportunity" and "inequality in ability", respectively. These assumptions provide us with the following structural earnings function[7]

$$w_{it} = \boldsymbol{z}_{it1}\boldsymbol{\delta} + b_i s_i - \frac{1}{2}\kappa_1 s_i^2 + a_{it} \quad [3]$$

where $E(a_{it}) = 0$. Utility maximising individuals choose to invest in $s_i^* = \frac{b_i - r_i}{\kappa}$ years of education, where $\kappa = \kappa_1 + \kappa_2$. If we allow the relative cost term $r_i$ to depend on the observable wage determinants as well as a set of (instrumental) variables that do not directly affect earnings, $\boldsymbol{z}_{it2}$, then the schooling equation can be rewritten as

$$s_i^* = \boldsymbol{z}_{it1}\boldsymbol{\pi}_1 + \boldsymbol{z}_{it2}\boldsymbol{\pi}_2 + u_{it} \quad [4]$$

where $u_{it}$ is now the demeaned sum of the unobservable relative returns and cost terms.

Conditional on schooling, individual heterogeneity affects earnings via general labour market productivity, $a_{it}$, and the ability to translate education into earnings, $b_i$. Deschenes (2004) refers to these components as

[7] The notation used here differs slightly from that used in the original Card (2001; 1999) papers, as our objective is to derive an estimable equation with time-varying error terms and conditioning variables.

"relative ability" and a "comparative advantage in the use of education". We will refer to the latter term as the "relative returns" coefficient, while being mindful of the fact that the actual marginal returns to education are $b_i - \kappa_1 s_i$. Card (2001) maintains the assumption that the schooling-earnings profile is concave, but as long as the conditions for an interior solution are met[8] there is no reason why this model is inconsistent with convex schooling-earnings profiles.

Rewriting the heterogeneous relative returns coefficient as $b_i = \bar{b} + v_i$, where $\bar{b} = E(b_i)$ gives us an estimable earnings equation

$$w_{it} = \boldsymbol{z}_{it1}\boldsymbol{\delta} + \bar{b}s_i - \frac{1}{2}\kappa_1 s_i^2 + v_i s_i + a_{it} \quad [5]$$

Card (2001, p. 1134) demonstrates that the linear schooling coefficient from an OLS regression of equation [2] is likely to be an upwardly biased estimate of the average marginal returns $\bar{b} - \kappa_1 \bar{s}$ since the optimal level of schooling is correlated with the OLS regression residual (which now consists of the last two terms on the right-hand side of equation [5]). Furthermore, his derivations (Card, 2001, p. 1133) also suggest that an OLS earnings regression with schooling included as a quadratic function will tend to over-estimate the degree of convexity in the expected schooling-earnings profile. Clearly then, estimating the Mincerian earnings function with OLS – and including schooling either linearly or as a quadratic function – will yield biased estimates of the average schooling returns parameters if schooling is endogenous in the earnings function or individuals are heterogeneous in schooling returns.

It is worth noting that although the Becker-Card model allows the schooling returns to vary by schooling year and individual, it still implies a substantive restriction on how additional schooling affects expected earnings compared to the more general equation [1]. Specifically, the slope of the marginal returns and cost curves are both assumed to be linear and identical for all individuals, neither of which is implied in equation [1]. This more general specification is increasingly common in the programme evaluation literature and can be used if the econometrician is satisfied with estimating a local average of the marginal

[8] The second-order sufficient condition requires that the slope of the marginal cost curve is larger than that of the marginal benefit curve.

returns. This approach to estimating the shape of the schooling-earnings profile is discussed in section 2.3.2.

#### 2.3.1.2 The Heckman-Gronau-Roy model of selection into employment

The Becker-Card model ignores any issues related to labour force participation, job search or unemployment, which makes it inappropriate to model the workings of a high unemployment labour market such as South Africa's. The sample selection that arises from a highly restrictive employment process is most commonly addressed with techniques based on the simple model first proposed by Gronau (1974) and operationalised by Heckman (1979). Suppose that each person has log reservation wage $w^r$ such that

$$w_{it}^r = \boldsymbol{z}_{it1}\boldsymbol{\mu}_1 + \boldsymbol{z}_{it3}\boldsymbol{\mu}_3 + \mu_4 s_i - \tfrac{1}{2}\mu_5 s_i^2 + p_i s_i + q_{it}$$

and will choose to participate in the labour market only if the offered wage exceeds this reservation wage. Assuming that the wage offer can be expressed as in [5], the employment outcome is then

$$m_{it}(s_i) = \mathbf{1}\left(\boldsymbol{z}_{it1}\boldsymbol{\psi}_1 + \boldsymbol{z}_{it3}\boldsymbol{\psi}_3 + \psi_4 s_i + \frac{1}{2}\psi_5 s_i^2 + c_i s_i + e_{it} > 0\right) \qquad [6]$$

where $\boldsymbol{\psi}_1 = \boldsymbol{\delta} - \boldsymbol{\mu}_1, \boldsymbol{\psi}_3 = -\boldsymbol{\mu}_3, \psi_4 = \bar{b} - \mu_4, \psi_5 = \kappa_1 - \mu_5$ , $c_i = v_i - p_i$ and $e_{it} = a_{it} - q_{it}$.

The addition of an employment process has important implications for how we wish to estimate and interpret our model. Firstly, if $E(e_{it}a_{it}) \neq 0$ then all of the regressors in the wage equation will generally be endogenous – even in the homogeneous parameter case where $v_i = p_i = 0$ – and instrumenting for schooling will not necessarily produce unbiased coefficients of the parameters of interest. Intuitively, if the effect of schooling on the wage offer exceeds its effect on the reservation wage, then individuals with lower schooling levels will be less likely to accept wage offers and are less likely to be in the sample of wage earners unless they also have high levels of relative ability $a_{it}$. This induces an artificial negative

correlation between schooling and relative ability amongst the employed which will cause a downward bias in our estimated schooling returns parameter. Heckman (1974; 1979) showed that this type of endogeneity can be addressed by using estimators that explicitly take sample selection into consideration.

Secondly, the fact that our sample of wage earners is a non-random draw from the population has implications for the interpretation of our parameters of interest. The Becker-Card model equates the marginal benefit of education with the wage increase associated with having more schooling, which ignores one of the most important benefits that the highly educated enjoy in high unemployment economies: an increased likelihood of being employed. Also, in a heterogeneous parameter framework where our object of interest is the marginal returns averaged over some subsample of individuals, our definition of this parameter should be explicitly altered to recognise that our sample only includes individuals in wage employment. How exactly this parameter relates to the population average of the marginal returns is not clear.

Finally, because the effect of education on the employment outcome changes the marginal benefit of schooling, this has implications for the optimal schooling level as derived in the Becker-Card model. This in turn affects the behavioural assumptions required to consistently estimate the causal effect of schooling on earnings. For example, where consistent estimation of the wage returns to schooling in a full-employment environment requires that the instruments are uncorrelated to these returns, the presence of sample selection will add the condition that instrumental variables must also be uncorrelated to the employment returns of schooling. This chapter will not attempt to recast the whole Becker-Card model in a high unemployment environment, but we do consider identifying conditions that allow us to consistently estimate some average of these returns in the presence of sample selection in sections 2.3.2.2 and 2.3.3.2.

### 2.3.2 The local average treatment effect interpretation

Studies that attempt to estimate the effect of schooling on expected earnings are increasingly choosing to use instrumental variable methods and interpret their results as an instrument-specific local average treatment effect. The benefit of this approach is that it does not rely on strong behavioural assumptions in order to identify the parameter of interest. Suppose that the earnings function is specified as in equation [1]. It is convenient to rewrite this model in terms of marginal returns

$$w_{it}(s) = \boldsymbol{z}_{it1}\boldsymbol{\delta} + \sum_{j=1}^{s_{max}} 1(s_i \geq j).\, b_{ij} + a_{it} \qquad [7]$$

where the marginal returns[9] $b_{ij}$ are allowed to vary without restrictions across individuals and schooling years. Clearly it would be impossible to estimate the values of the heterogeneous returns $b_{ij}$ for each individual and schooling year, but we may be able to estimate the average value of this coefficient for some subsample of the population. However, in choosing a specific summary statistic it is sensible to consider not only what we *want* to identify, but also what the data *will allow us* to identify. This consideration, combined with the very unrestrictive conditions under which the LATE can be consistently estimated, explains the recent popularity of interpreting the two-stage least squares (2SLS) estimates as a LATE: the average treatment effect for those individuals whose receipt of the treatment was the result of variation in the exogenous instrumental variable.

#### *2.3.2.1 Full employment case*

Suppose wages are generated according to equation [7] and that the schooling equation is written as

$$s_i = \boldsymbol{z}_{it1}\boldsymbol{\pi}_1 + \boldsymbol{z}_{it2}\boldsymbol{\pi}_2 + u_{it} = \boldsymbol{z}_{it}\boldsymbol{\pi} + u_{it} \qquad [8]$$

[9] The unrestricted marginal returns $b_{ij}$ is distinct from the heterogenous component of the marginal schooling returns in equation [3], $b_i$.

where $\boldsymbol{z}_{it1}: 1 \times K_1$ is the exogenous conditioning variables from the earnings equation and contains a constant, $\boldsymbol{z}_{it2}: 1 \times K_2$ is a set of instrumental variables, and $K = K_1 + K_2$. These equations are easily related to equations [3] and [4] of the Becker-Card model. Equation [7] represents a structural equation that characterises the process through which $w_{it}$ was generated and from which the parameters $(\boldsymbol{\delta}, b_{ij})$ derive their causal interpretation. On the other hand, equation [8] is a reduced form equation and results from a linear projection of $s_i$ on $\boldsymbol{z}_{it}$. The $\boldsymbol{\pi}$ parameters may contain information supporting or contradicting the hypothesised channels through which the instrumental variables are supposed to affect schooling outcomes, but interpreting these coefficients as the structural parameters from a behavioural equation would require additional assumptions about how individuals make their schooling decisions. For now we are only interested in estimating some local average of the $b_{ij}$ coefficients while allowing for the possibility that $s_i$ is endogenously determined. Imbens and Angrist (1994) derived the conditions under which two-stage least squares (2SLS) can be used to estimate the LATE:

**Assumptions LATE:** If we assume that i) $E(\boldsymbol{z}_{it} a_{it}) = \boldsymbol{0}$, ii) $\operatorname{rank} E\left(\boldsymbol{z}_{it}'(\boldsymbol{z}_{it1}, s_i)\right) = K_1 + 1$, and iii) the effect of $\boldsymbol{z}_{it2}$ on $s_i$ is monotonic, then 2SLS will provide consistent estimates of local average treatment effect of $s$.

Assumptions LATE.i and LATE.ii are the regular instrument exogeneity and relevance assumptions required for 2SLS consistency in the constant parameter case. Conceptually, these conditions state that the instrumental variables are correlated to the schooling outcomes ($\boldsymbol{\pi}_2 \neq 0$) and uncorrelated to unobserved earnings ability $a_{it}$. The exogeneity condition LATE.i implicitly rules out sample selection of the type that arises in the Heckman-Gronau-Roy model. Monotonicity (LATE.iii) requires that a change in the instrumental variables will never affect the expected value of the endogenous variable in different directions for different individuals, i.e. increase the expected schooling years for some and decrease it for others. However, consistency of the 2SLS estimates requires no further assumptions about how schooling decisions were made. If we were also interested in predicting the causal effect of a certain policy on

schooling outcomes – and it isn't difficult to think of cases where we would be – then the relationship between $u_{it}$ and $\boldsymbol{z}_{it}$ would have to be considered. For now, however, the linear projection used to obtain equation [8] is purely mechanical. This is particularly fortunate given the discrete characteristics that $u_{it}$ inherits from a count variable like schooling, which could make an assumption of independence between $u_{it}$ and $\boldsymbol{z}_{it}$ difficult to justify.

Much of the programme evaluation literature is written with regards to a binary treatment that is instrumented with a binary instrumental variable. In this case Imbens and Angrist (1994) demonstrated that the LATE is the average of the treatment effects for all "compliers": individuals who would have received treatment under one IV value but not under the other. In the case of multiple treatments (such as schooling) the LATE can be interpreted as the weighted average of the individual treatment effects, where individuals and treatments that are more affected by the instrumental variable are weighted more heavily. Stated more formally and in the notation used for our theoretical model: if the variation in the instrumental variable $z_{it2}$ induced a change in agent $i$'s schooling outcomes of $\Delta_i s$, then the LATE can be expressed as $\frac{E(b_{ij}\Delta_i s)}{E(\Delta_i s)}$. Clearly the LATE of any instrument will weigh more heavily the marginal returns of the type of individual who responds to a change in the value of the instrument and the additional schooling years that these individuals decide to obtain. Having multiple IVs (or similarly, IVs with multiple values) further changes the estimate to the weighted average of each of the IV-specific LATEs with weights proportional to the strength of the instrument. Finally, when conditioning variables are included in the outcome equation the interpretation of the estimate is further extended to the weighted average of each of the covariate-specific LATEs.

In section 2.4.1 we introduce two policies that may help us capture exogenous variation in schooling outcomes and hence to estimate the LATE of schooling for these two specific policies. As discussed above the LATE of the IVs associated with each policy can be used to investigate the causal effect of these policies on earnings under a very general specification of earnings and schooling decisions. Unfortunately such estimates do not tell us anything about the shape of the schooling-earnings profile, which is the

primary concern of this chapter and dissertation. However, if the two policies – and the different IVs constructed to reflect different aspects of each policy – affected schooling decisions at different points of the schooling distribution, then this could provide important information about the shape of the schooling-earnings profile. Our identification strategy is therefore to compare the LATE for IVs that affect mainly early schooling year to the LATE of those that affect later schooling years. If the latter exceeds the former, then this is *prima facie* evidence in favour of a convex schooling-earnings profile in the presence of heterogeneous returns and endogenous schooling. Of course, the validity and applicability of this approach is limited by two caveats. Firstly, the weights of the LATE vary not only across schooling years but also across individuals. If one IV affects mainly low return students at earlier schooling years and another impacts high return students at advanced schooling years, then a comparison of the LATEs will overstate the convexity of the actual schooling-earnings profiles. Secondly, the LATE approach does not allow us to recover the expected effect of the IV on each schooling outcome. This means that even if a comparison of low-schooling and high-schooling IVs is informative about whether the marginal returns usually rise or fall at higher schooling years, the exact shape of this profile cannot be recovered without more information or stronger assumptions.

#### 2.3.2.2 Sample selection case

The estimation and interpretation of the LATE is complicated by the presence of sample selection. In the case of a single binary instrumental variable, the 2SLS estimator can be written as

$$b_{2SLS} = \frac{E(w_{it}|\boldsymbol{z}_{it1}, z_{it2} = 1) - E(w_{it}|\boldsymbol{z}_{it1}, z_{it2} = 0)}{E(s_i|\boldsymbol{z}_{it1}, z_{it2} = 1) - E(s_i|\boldsymbol{z}_{it1}, z_{it2} = 0)} = \frac{E(b_{ij}\Delta_i s)}{E(\Delta_i s)}$$

because $E(a_{it}|\boldsymbol{z}_{it1}, z_{it2} = 1) - E(a_{it}|\boldsymbol{z}_{it1}, z_{it2} = 0) = 0$. However, if we estimate 2SLS on the sample of employed individuals only, then the estimator should be rewritten as

$$b_{2SLS} = \frac{E(w_{it}|\boldsymbol{z}_{it1}, z_{it2}=1, m_{it}=1) - E(w_{it}|\boldsymbol{z}_{it1}, z_{it2}=0, m_{it}=1)}{E(s_i|\boldsymbol{z}_{it1}, z_{it2}=1, m_{it}=1) - E(s_i|\boldsymbol{z}_{it1}, z_{it2}=0, m_{it}=1)}$$

$$= \frac{E(b_{ij}\Delta_i s)}{E(\Delta_i s)} + \frac{E(a_{it}|\boldsymbol{z}_{it1}, z_{it2}=1, m_{it}=1) - E(a_{it}|\boldsymbol{z}_{it1}, z_{it2}=0, m_{it}=1)}{E(s_i|\boldsymbol{z}_{it1}, z_{it2}=1, m_{it}=1) - E(s_i|\boldsymbol{z}_{it1}, z_{it2}=0, m_{it}=1)}$$

If $E(a_{it}|\boldsymbol{z}_{it1}, z_{it2}=1, m_{it}=1) = E(a_{it}|\boldsymbol{z}_{it1}, z_{it2}=0, m_{it}=1)$ then the 2SLS coefficient remains a consistent estimator of the LATE, although our interpretation should make clear that the group of compliers over which the marginal returns are averaged now only includes employed individuals.

Suppose that selection into employment was determined by a slightly different version of equation [6], in which the schooling instruments can affect the employment probability

$$m_{it}(s_i) = \mathbf{1}\left(\boldsymbol{z}_{it1}\boldsymbol{\psi}_1 + \boldsymbol{z}_{it2}\boldsymbol{\psi}_2 + \boldsymbol{z}_{it3}\boldsymbol{\psi}_3 + \psi_4 s_i + \frac{1}{2}\psi_5 s_i^2 + c_i s_i + e_{it} > 0\right) \quad [9]$$

In this case the 2SLS estimator will be appropriate if either $E(e_{it}a_{it}) = 0$ or $\boldsymbol{\psi}_2 = 0$. The probit form of the employment function means that the expected value of $e_{it}$ (conditional on being a wage earner) depends on the observable determinants of employment. If $E(e_{it}a_{it}) \neq 0$ then the expected value of the unobserved market ability of the employed will also be correlated to observable determinants of employment. This means that if $\boldsymbol{\psi}_2 \neq 0$, so that a change in the value of the instrumental variable affects the employment likelihood, then the expected value of $a_{it}$ will also vary with the IV values and the 2SLS estimates will be biased. In the presence of sample selection the conditions under which instrumenting strategies are able to identify a LATE are more stringent than is otherwise the case. Even if the instruments are uncorrelated to unobservable market ability in the population, this may not be true for the sample of wage earners if the instruments cannot also be excluded from the employment regression.

It is not generally possible to test whether $\boldsymbol{\psi}_2 = 0$ or to control for this sort of sample selection bias, even when making the strong assumptions that would allow us to combine elements of the Rivers-Vuong (1988) two-step regression and 2SLS. Such an approach would customarily require us to find some estimate of the conditional expectation of the employment error term $E(e_{it}|\boldsymbol{z}_{it1}, \boldsymbol{z}_{it2}, \boldsymbol{z}_{it3}, m)$ and then

control for its correlation with the earnings error term by including it as an additional explanatory variable in a 2SLS earnings regression[10]. The Rivers-Vuong approach would proceed by including the residual from a first step schooling regression of equation [8] as an additional explanatory variable in the probit regression of equation [9]. However, since this schooling residual is a deterministic function of $\boldsymbol{z}_{it1}$, $\boldsymbol{z}_{it2}$ and $s_i$, perfect multicolinearity in the probit regression can only be avoided if we assume that $\boldsymbol{\psi}_2 = 0$. However, in that case the LATE interpretation of the 2SLS – which now recognises that the compliers only include those in wage employment – is already valid. The best way to ensure that the LATE estimates are not biased may be to include as controls in the earnings regression any correlates of the IVs that are also employment shifters, $\boldsymbol{z}_{it3}$, although this remedy will exacerbate any weak instrument problems in the data.

### 2.3.3 Average schooling-earnings profile interpretation

Another approach to estimating the shape of the schooling-earnings profile for the "average" individual is to start from the assumptions of the Becker-Card model and to estimate the $\bar{b}$ and $\kappa_1$ parameters from equation [5]. These coefficient estimates will allow us to construct the expected schooling-earnings profile for a randomly selected individual, which corresponds to the object of interest of this chapter.

#### *2.3.3.1 Full employment case*

Suppose that the earnings function is given by equation [5] and that individuals' schooling decisions are represented by equation [8]. Estimating the parameters of interest now involves estimating the ATE in what Heckman & Vytlacil (1998) termed the correlated random coefficient (CRC) model. The conditions under which this parameter can be estimated have been a point of some contention. Garen (1984) was the first to consider the estimation of the ATE in a CRC model – he too considered a wage function in which

[10] This is similar to the approach taken by Arabsheibani & Mussurov (2007) and Chen & Hamori (2009), although their analysis assumes parameter homogeneity.

the returns to schooling may be correlated to endogenous schooling, although his earnings equation omitted the quadratic schooling term from equation [5] – and he argued that 2SLS cannot recover consistent estimates of $(\boldsymbol{\delta}, \bar{b})$ because $E(v_i s_i) \neq 0$. Instead he proposed using the control function estimator to estimate these parameters.

The control function (CF) approach entails first estimating the schooling equation before using the regression residual $\hat{u}_{it} = s_i - \boldsymbol{z}_{it}\hat{\boldsymbol{\pi}}$ as an additional control in an OLS regression of the structural earnings equation. In the case of a linear endogenous variable with a constant parameter this approach can be shown to be numerically identical to 2SLS. However, this equivalence disappears if schooling enters the earnings function non-linearly or heterogeneously, in which case the two estimators use different identifying assumptions to estimate the parameters of interest. Although $u_{it}$ is not observed, the ability to consistently estimate it allows us to treat it as if it were observed asymptotically when using it as a control variable in the earnings regression. The fact that the CF estimator uses a generated regressor means that the OLS estimate of the second-stage regression coefficient covariance matrix must be adjusted to reflect this additional source of variation[11]. In the context of a CRC model, Garen's (1984) approach is to estimate the schooling residuals, $\hat{u}_{it}$, and to then estimate the earnings regression with both $\hat{u}_{it}$ and $\hat{u}_{it}.s_i$ included as additional regressors. He demonstrated that this estimator will consistently estimate the structural model parameters under fairly strong conditions that included the normality of $u_{it}$ and that $(u_{it}, v_i)$ are independent of $\boldsymbol{z}_{it}$.

Wooldridge (1997) showed that although the presence of $v_i s_i$ in the model error term will necessarily bias the 2SLS estimate of the intercept coefficient, the (generally more interesting) slope coefficients can be consistently estimated under conditions that are in many respects weaker than Garen's (1984) assumptions for CF consistency. Specifically, using 2SLS to estimate the ATE of schooling requires that $E(u_{it}|\boldsymbol{z}_{it}) = 0$, $E(u_{it}^2|\boldsymbol{z}_{it})$ is constant and $E(v_i|u_{it}) = \rho_2 u_{it}$. He also argues that the additional computations required to adjust the standard errors of the CF estimator in order to account for the

[11] In our empirical analysis in section 2.4 we do this using bootstrapping techniques.

generated regressors makes Garen's (1984) proposed methodology less attractive than 2SLS. Heckman & Vytlacil (1998) developed a two-step plug-in estimator for the same model, and demonstrated that this will provide consistent estimates of the ATE under weaker assumptions than those in Wooldridge (1997) for 2SLS consistency. Wooldridge (2003) responded by deriving consistency conditions for the 2SLS estimator that are weaker than those in Wooldridge (1997) and also weaker than required for the Heckman & Vytlacil (1998) estimator. He argues that the 2SLS estimates of the ATE in a CRC model will therefore be more robust than those of the Heckman & Vytlacil (1998) approach, and simpler to calculate than Garen's (1984) CF estimator. The conditions required for 2SLS consistency (Wooldridge, 2007) are stated below:

**Assumptions 2SLS:** If we assume that i) $E(a_{it}|\boldsymbol{z}_{it}) = 0$, ii) $\text{rank}\, E\left(\boldsymbol{z}_{it}'\left(\boldsymbol{z}_{it1}, s_i, s_i^2\right)\right) = K_1 + 2$, iii) $E(v_i|\boldsymbol{z}_{it}) = 0$ and iv) $E(v_i s_i|\boldsymbol{z}_{it}) = E(v_i s_i)$, then 2SLS will provide consistent estimates of the equation [5] parameters $\left(\boldsymbol{\delta}, \bar{b}, \kappa_1\right)$.

The mean independence exogeneity condition in assumption 2SLS.i is clearly stronger than the uncorrelatedness condition required in LATE.i, although one may be concerned about the exogeneity of an instrumental variable that is uncorrelated with but not mean-independent of unobserved ability. Furthermore, the instrument relevance condition 2SLS.ii now requires that we have at least two IVs that do not perfectly overlap in their ability to explain schooling outcomes. A more restrictive assumption is 2SLS.iii, which requires the relative returns to be mean independent of $\boldsymbol{z}_{it}$. Since schooling is correlated to the instruments by assumption 2SLS.ii, this would rule out any cases in which the returns coefficient $v_i$ varied systematically with the level of schooling. However, the inclusion of the quadratic schooling term should go some way in addressing this concern.

Card (2001, p. 1138) warns that assumption 2SLS.iv – which requires that the average mapping from the relative schooling returns to schooling are unaffected by the IV – may also be violated in the wage-schooling model. His analysis of the U.S. NLS data shows that even though ability appears to be independent of college proximity (a popular IV for schooling), the correlation between schooling and

ability is stronger for those who grew up close to colleges. In this case the expected value of $v_i s_i$ will shift with the values of $\boldsymbol{z}_{it}$, so that 2SLS will not provide consistent estimates of the ATE of schooling.

The CF estimator allows a different set of assumptions under which the ATE can potentially be recovered. This approach requires first estimating the schooling equation [8], and then using the residuals $\hat{u}_{it}$ and the residuals interacted with schooling $\hat{u}_{it}.s_i$ as additional controls in the OLS regression of equation [5]. Wooldridge (2007) presents sufficient conditions for the CF to consistently estimate the ATE as follows:

**Assumptions CF:** If we assume that i) $(a_{it}, v_i, u_{it})$ is independent of $\boldsymbol{z}_{it}$, ii) $\text{rank}\, E\big(\boldsymbol{z}_{it}'(\boldsymbol{z}_{it1}, s_i)\big) = K_1 + 1$, iii) $E(a_{it}|u_{it}) = \rho_1 u_{it}$ and iv) $E(v_i|u_{it}) = \rho_2 u_{it}$ then the CF estimator will provide consistent estimates of the equation [5] parameters $\big(\boldsymbol{\delta}, \bar{b}, \kappa_1\big)$.

CF.i imposes a "substantive restriction on the structural and reduced form error terms" (Wooldridge, 2007). Firstly, requiring $\boldsymbol{z}_{it}$ to be independent of $a_{it}$ implies a stronger exogeneity condition for the instruments than uncorrelatedness or mean independence: we now also preclude the instrument from affecting non-linear functions of market ability, such as its variance. Perhaps more restrictively, assuming that the schooling error $u_{it}$ is independent of $\boldsymbol{z}_{it}$ seems to suggest having more knowledge about the process that generated schooling outcomes $s_i$ than was the case when using a 2SLS estimator that merely performed a linear projection on the schooling outcome. Although we will continue to refer to equation [8] as the reduced form equation, the assumption that $u_{it}$ is independent of $\boldsymbol{z}_{it}$ comes close to assuming that we have correctly specified the structural equation for $s_i$. This assumption also rules out models in which the endogenous variable has discrete characteristics, which could be a problem in an analysis such as ours where schooling is a count variable. Finally, assuming that $v_i$ is independent of $\boldsymbol{z}_{it}$ is an even stricter version of assumption 2SLS.iii, which precludes any systematic variation between schooling and the returns that are not already captured by the quadratic schooling term.

Assumption CF.iii specifies the precise functional form of the relationship between the structural and reduced form errors. When combined with assumptions CF.i and the form of equation [8] it implies that $E(a_{it}|\boldsymbol{z}_{it}, s_i) = \rho_1 u_{it}$, which is enough to ensure that we can remove any endogeneity arising from correlation between $a_{it}$ and $u_{it}$ by the inclusion of $\hat{u}_{it}$ as a regressor. Similarly, assumption CF.iv will allow us to control for the relationship between $v_i$ and $u_{it}$ by including $\hat{u}_{it}.s_i$ as an additional control variable. These functional form assumptions appear very restrictive, but it is the fact that $E(a_{it}|\boldsymbol{z}_{it}, s_i)$ and $E(v_i|\boldsymbol{z}_{it}, s_i)$ are known functions of a consistently estimable variable that is required for consistency rather than the linearity of these functions. Equally valid identifying assumptions would be obtained by replacing CF.iii and CF.iv with alternative specifications $E(a_{it}|u_{it}) = \vartheta_a(u_{it})$ and $E(v_i|u_{it}) = \vartheta_v(u_{it})$ where $\vartheta_a(u_{it})$ and $\vartheta_v(u_{it})$ could be something like low-order polynomials that are believed to accurately approximate a broader class of relationships between the structural and reduced form errors. In this case the CF approach is applied by including the polynomial terms of $\hat{u}_{it}$ (on its own and interacted with $s_i$) as second-stage regressors. This is exactly the approach taken in Söderbom et al. (2006).

The assumed independence between $u_{it}$ and $\boldsymbol{z}_{it}$ technically rules out using the CF approach outlined above unless the endogenous variable is continuous. Wooldridge (2007) remarks that all data are discrete in practice, but that certain endogenous count variables may be "too discrete" for this method to be appropriate. Some returns to schooling studies (Söderbom M. M., Teal, Wambugu, & Kahyarara, 2006) evidently consider the schooling variable to be sufficiently continuous to exploit this independence assumption. Furthermore, this is not a problem that only arises in the context of the CF approach. Assumption 2SLS.iv is also untenable if schooling is discrete, yet a number of studies (including Heckman and Vytlacil (1998)) have used estimators that require this assumption in the context of an earnings model with endogenous schooling. The fact that observed schooling outcomes for our sample of black South Africa males span the full range of schooling outcomes from 0 to 16 means that discreteness will be less of a problem than in many other countries.

If the schooling variable is judged to be too discrete to use with assumption CF.i, then one alternative is to explicitly model the process that produced this discreteness. However, such techniques typically require strong distributional assumptions. Suppose, as in our case, that $s_i$ is ordered but discrete, and is known to have been generated according to the following process: $s_i = \sum_{j=0}^{J} 1(v_j < s_i^\dagger < v_{j+1}).j$, where $v_0 = -\infty$, $v_{J+1} = \infty$, $s_i^\dagger = \boldsymbol{z}_{it}\boldsymbol{\pi}^\dagger + u_{it}^\dagger$ and the econometrician knows the distribution of $u_{it}^\dagger$. This information can be combined with assumptions CF.ii, CF.iii and the independence of $u_{it}^\dagger$ and $\boldsymbol{z}_{it}$ to consistently estimate $E(u_{it}^\dagger|\boldsymbol{z}_{it}, s_i)$. This term can then be included as an additional control in the wage regression (on its own and interacted with schooling) to yield consistent estimates of the equation [5] parameters. If $u_{it}^\dagger$ is distributed normally then schooling can be modeled using an ordered probit estimator, after which it is simple to calculate the values of $E(u_{it}^\dagger|\boldsymbol{z}_{it}, s_i)$. Alternatively, data-intensive non-parametric techniques can be used in order to obtain a flexible fit of this term without any assumptions about the distribution of $u_{it}^\dagger$, although this approach is not pursued in this chapter.

#### *2.3.3.2 Sample selection case*

Only a handful of studies have attempted to address the issues of endogeneity and sample selection simultaneously, but all those that we are aware of do so in a constant parameter setting. García, Hernández & López-Nicolás (2001) control for selection into employment in an IV regression, but treat schooling as exogenous in the selection equation. However, if unobservable schooling abilities are important in determining wages, then it seems natural that the same abilities would affect employability. Arabsheibani & Mussurov (2007) and Chen & Hamori (2009) both drop schooling from the selection equation but use all of the exogenous variables (including the IVs and employment exclusion restrictions) in estimating the schooling, selection and earnings equations. They justify this by claiming to "[f]ollow the suggestion made by Wooldridge (2002)" (Arabsheibani & Mussurov, 2007, p. 352; Chen & Hamori, 2009, p. 145). Although

Wooldridge (2002, p. 567) discusses this example, in his model schooling is assumed not to determine employment, rather than it being omitted in order to address its endogeneity[12].

Although sample selection is clearly an important aspect of developing country wage-schooling models, correcting for this becomes very difficult once we allow for a heterogeneous schooling return parameter. If we construct the selection equation as in section 2.3.1.2 by subtracting the reservation wage from the wage offer, then the employment equation will also contain an endogenous schooling term with a heterogeneous coefficient. In the Heckman selection model the effect of sample selection can be controlled for because of our ability to consistently estimate the conditional expectation of the schooling error $e_{it}$. However, if the schooling effect on employment is heterogeneous, then equation [6] can be rewritten as $m_i = 1\left(c_i s_i + e_{it} > -\left(\boldsymbol{z}_{it1}\boldsymbol{\psi}_1 + \boldsymbol{z}_{it3}\boldsymbol{\psi}_3 + \psi_4 s_i + \frac{1}{2}\psi_5 s_i^2\right)\right)$. Even after estimating $E(c_i s_i + e_{it}|\boldsymbol{z}_{it}, s_i, m_i)$, this still does not help us control for the unobservable wage component because we cannot separately identify $c_i s_i$ and $e_{it}$. Intuitively, there are two types of individuals who enter into employment: those who have – conditional on their schooling level and other observable characteristics – sufficiently high levels of the employment error term, $e_{it}$, or those that have high levels of the ability to transform schooling into an increased employment probability, $c_i$. Unfortunately the data does not allow us to distinguish between these two types of individuals, and hence we cannot separately control for these two traits in our earnings regression. The only way to make any progress is to assume that the schooling coefficient in the employment equation is homogeneous, which would be the case if $v_i = p_i$ in equation [6]. In this scenario we can proceed using a technique that combines elements of the sample-selection approach due to Rivers & Vuong (1988) and 2SLS.

**RV-2SLS:** If we assume that i) $(a_{it}, u_{it}, v_i, e_{it})$ is independent of $(\boldsymbol{z}_{it1}, \boldsymbol{z}_{it2}, \boldsymbol{z}_{it3})$, ii) $E(a_{it}|u_{it}, e_{it}) = \rho_{11}u_{it} + \rho_{12}e_{it}$, iii) $E(v_i|u_{it}, e_{it}) = \rho_2 u_{it}$ , iv) $E(e_{it}|u_{it}) = \rho_3 u_{it}$ and v) $e_{it}$ and

[12] In the homogenous parameter case considered in these papers, omitting schooling from the employment probit regression but including the schooling residual as a regressor means that the inverse Mills ratio from this regression will be a consistent estimate of $E(e_{it} + \psi_4 s_i|\boldsymbol{z}_{it1}, \boldsymbol{z}_{it2}, \boldsymbol{z}_{it3}, s_i, m_{it} = 1)$. Controlling for this term in a 2SLS earnings regression will provide consistent estimates of the earnings coefficients (since the instrumental variables are exogenous with respect to the composite error term), but this may simply be a fortuitous coincidence, since none of these papers justify the omission of schooling on these grounds.

$u_{it}$ are normally distributed, then the method described below will provide consistent estimates of $(\boldsymbol{\delta}, \bar{b}, \kappa_1)$.

Estimating the schooling error term, $\hat{u}_{it}$, from the 2SLS first-stage regression (now estimated on the whole population rather than just the sample of wage earners) and including it as an additional regressor in an employment probit regression will produce consistent estimates of the coefficients in equation [9], from which a consistent estimate of the conditional mean of the employment error $E(e_{it}|\boldsymbol{z}_{it1}, \boldsymbol{z}_{it2}, \boldsymbol{z}_{it3}, s_i, m_{it} = 1)$ can be calculated. If the inverse Mills ratio of this term[13] is included as an additional regressor in the second-stage of the 2SLS regression together with $\hat{u}_{it}$ and $\hat{u}_{it}.s_i$ then this will correct for sample selection and endogeneity in the earnings equation.

## 2.4 Data and instrumental variables

### 2.4.1 Policy shifts and the identification of schooling returns

We now discuss three recent policy shifts that affected the schooling and employment decisions of South Africans and that are central in our strategy of identifying the schooling returns in the presence of endogenous schooling and selection into employment. In order to instrument for endogenous schooling, we need some variation in schooling that is uncorrelated to its potential benefit. Towards this goal, we explore two sets of government policy shifts that affected the schooling outcomes of South African learners. The first is the rollback of apartheid era restrictions on black access to schooling and the second is the stricter enforcement of restrictions on enrolling over-aged learners. The ability to correct for selection into employment requires finding variables that determine an individual's willingness to accept wage offers without affecting these offers themselves. This leads us to the recent expansion in the value and coverage of government social grants.

[13] Note that when calculating this term it is necessary to subtract the contribution of the schooling residual to the predicted value of the employment index.

#### 2.4.1.1 A brief history of restrictions on black schooling

Prior to forming the Union of South Africa in 1910 the Cape, Natal and Transvaal colonies and the Orange Free State each had separate schools for their white and black students, and the quality of instruction in the black schools was generally poor. This was particularly true in rural areas, where both schools and qualified teachers were in short supply (Baloyi, 2004, p. 54). Most primary schools were run by missions, and the location of these schools were mainly dictated by "denominational concerns and rivalries, rather than by any sense of planning to meet real needs" (Hartshorne, 1992, p. 24). At the same time there was almost no system of secondary schools for black learners. After union formation black enrolments rates were gradually improved (Malherbe, 1977, p. 257) due to increased subsidies paid for black learners – which led to an expansion of secondary mission schools (Hartshorne, 1992, p. 62), especially in black townships – and the Union government's efforts to build more state schools. However, black enrolment rates remained low because of the lack of schools in most areas (Baloyi, 2004, p. 60). There were also stark geographical differences in black education in terms of the distribution of school buildings, the prescribed curriculum and policies with respect to language of instruction.

In 1948 the National Party won the general election on a platform of "separate development", and started enacting the collection of laws that would come to be known as *apartheid*. Some of these policies were explicitly aimed at restricting black access to education. *The Bantu Education Act of 1953* called into existence a Black Education Department whose responsibilities included compiling a curriculum that would prepare black workers for the menial occupations that the apartheid government envisioned for them. It also restricted the training of black teachers (Christie, 1985, p. 79), led to the closure of most mission schools and further limited the amount of funding for black education. In 1959 the *Extension of University Education Act* precluded all black students from attending "white" universities. Between 1953 and 1970, expenditure on black students decreased from one-eighth of what was spent on each white student, to one-sixteenth (Christie, 1985, p. 100).

Although many of these laws were met with resistance from black students and other organisations, it was the ministerial decision to implement Afrikaans as the medium of instruction in black secondary schools that sparked the 1976 Soweto uprisings and led to an era of widespread student unrest and civil revolt against the apartheid system. The National Party government initially responded by reversing their language policy and implementing piecemeal reforms, but this was followed by more substantial reforms with the promulgation of the *National Policy for General Affairs Act (No. 76) of 1984*. Per-pupil spending on black education increased to one-fifth of white levels by the late 1980s (Baloyi, 2004, p. 129) and improvements were made in the remuneration and training of black teachers. The rate at which new classrooms were built also accelerated (Goodey, 1988), which led to further increases in black enrolment rates. *The Private Schools Act of 1986* allowed racially mixed private schools and in 1990 conditions were published under which white state schools could admit black students. Although historically white universities were not technically allowed to enrol black students during the 1980s, the existence of certain loopholes meant that by 1990 28% of places at white English medium universities were filled by black students (Bunting, 2002).

In 1990 the National Party government announced the end of apartheid and the various political parties started a process of negotiations regarding ways to deracialise South Africa's education systems. As a result the racial gap in schooling expenditure was further narrowed, although these changes were only gradually phased in over time due to budget constraints. More school buildings were also built in predominantly black areas and existing ones were upgraded, although the backlog of school facilities – particularly in rural areas – meant that even after these improvements the conditions of many schools were still very poor. Following political transition in 1994, a new set of policies were implemented that aimed to improve educational access for black learners, including open admission to all public schools and free and compulsory education until the age of 15.

#### 2.4.1.2 Restricting access for over-aged learners

A high share of South African youths have historically chosen to stay in school for long beyond the normal school-going age due to late school entry, high repetition, drop-out and drop-in rates and the high unemployment rate amongst young workers with less than completed secondary education. In 1998 60% of Grade 12 learners were older than the correct grade-age (Guluza & Hoadley, 1998, p. 1). This contributes to the already high pupil-teacher ratios in disadvantaged schools, which imposes a negative externality on pupils of the correct grade-age while tying up scarce resources in the education system.

Shortly after political transition the Department of Education (DoE) discussed – albeit in very vague terms – a strategy to reduce the large numbers of over-aged learners in schools (Republic of South Africa, 1995). This document refers to eliminating over-aged enrolment "in time" (1995, p. 36) at which point such a policy would be "enforced by the provincial Ministries of Education for the designated age group, on the basis of designated magisterial districts, until by stages the whole country is covered". This seems to suggest a plan to phase in such a policy by geographical regions. Given that South Africa's constitution guarantees all learners a basic level of education, the implementation of this policy was fraught with complications from the start. It was initially envisioned that the affected group of over-aged learners would be absorbed into alternative Further Education and Training (FET) colleges outside of the schooling system, where they would gain the appropriate skills required to successfully transition into the labour market (Republic of South Africa, 1995). However, Guluza & Hoadley (1998) believe that by the time this policy was suggested the "FET alternative" was already inaccessible to many potential FET candidates due to the scarcity of such institutions in rural areas and the often high fees. Better-located and cheaper schools therefore remained a more suitable alternative for most of these individuals.

Nevertheless, a shift in focus towards normalising the age profile was clearly a policy prerogative. In 1998 the Minister of Education published the *Age Requirements for Admission to an Ordinary Public School*, which defined the appropriate age to for admission to a certain grade as "the grade number plus 6" (Department of Education, 1998). However, the notice itself seems to be more concerned with the enrolment of under-

age children in Grade 1, and – perhaps due to the sensitive nature of this policy – official DoE documents describing the implementation of restrictions on over-aged learner admissions are hard to come by. We are therefore left to infer much about the implementation details from school data, the policy documents of provincial education departments and other research papers.

DoE administrative data show that in the middle of a long-term trend of gradually growing enrolment numbers there occurred a sharp and sudden drop in enrolment between 1998 and 2002 (Kraak, 2008, p. 12), which Perry and Arends (2003) ascribe to the policy of age requirements for school admissions. Similarly, the number of matric candidates decreased by nearly 20% over the same 4 year period (Kraak, 2008, p. 14). Provincial policy documents indicate that by 2003 these directives had already been adopted, along with the promotion of the FET system (Western Cape Education Department, 2003). Schools in the Western Cape were to deny admission to learners who were more than two years older than the appropriate grade-age. Given that schooling is compulsory for all learners until the age of 15, the affected individuals would be those who were i) older than 15, ii) more than two years older than the correct grade-age, and iii) wanted to enrol in school after 1998.

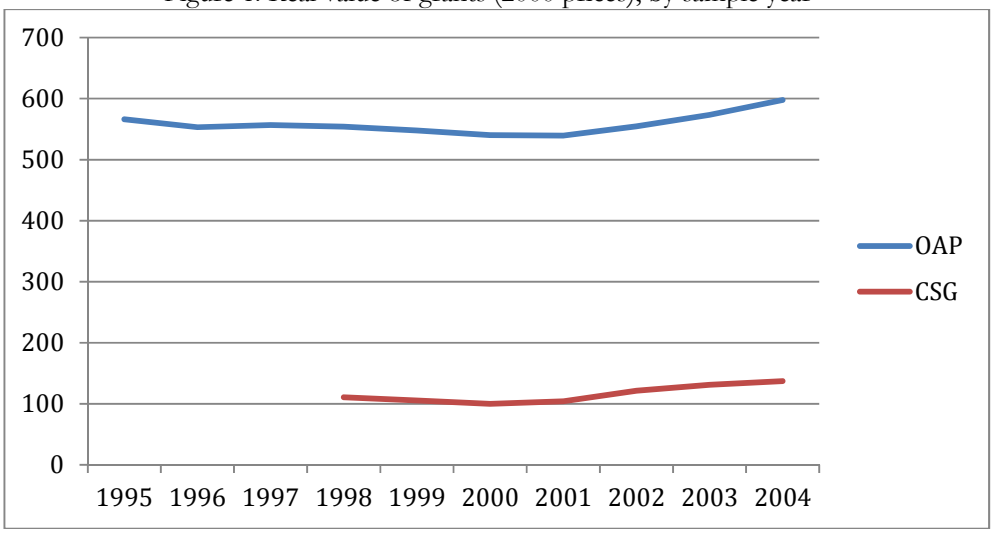
It is not clear to what extent this policy was phased in across provinces in the way that was initially envisioned, but a study by Guluza & Hoadley (1998) found that – at least with respect to restricting access to under-aged learners – schools were "caught in the middle between policies which require them to exclude [these] learners and pressures within the school and from the community which push them towards accepting these learners". It therefore seems unlikely that this policy would have been uniformly enforced in all schools. Given that both the initial distribution of over-aged learners and the implementation of the policy may have varied across provinces, we would want our instrumental variable to allow geographical heterogeneity in the implementation of this policy.

#### 2.4.1.3 Expansion of social grants

Social grants have long played an important role in supporting the South African poor in an economic environment marked by low employment rates and a relatively small informal sector. These grants are paid to a number of groups who are deemed to be vulnerable and deserving of state assistance, including veterans, foster children and the disabled. However, in this study we are only interested in the effect of the two largest grants, the child support grant (CSG) and the social old-age pension (OAP). The OAP is paid to individuals over the age of 60 and the CSG to households with a child below a certain age. Eligibility for both grants is subject to a means test.

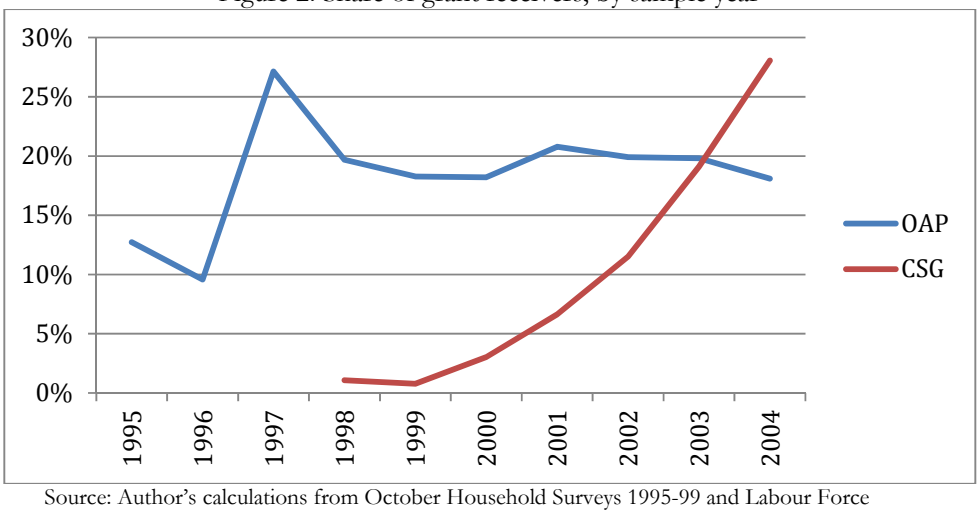
The CSG was rolled out in 1998 to replace the racially biased state maintenance grant. The OAP initially suffered from similar race-based eligibility criteria and benefit levels, but by 1993 these differences were removed (Van der Berg, 2002). Although the post-transition South African government initially embarked on a policy of fiscal austerity, by 2002 accelerated economic growth, improved efficiency in tax collection and a decrease in debt interest payments meant that more resources were available for poverty alleviation programmes. This allowed the National Treasury the scope to increase the real value of grant payments, as well as relaxing the eligibility criteria for the CSG. Figures 1 and 2 plot the time trends in eligibility and grant value between 1995 and 2004. Between 2000 and 2004 the value of an OAP payment increased by 11% and that of the CSG by 37%, which represented a large fluctuation in non-labour market income for South African households. Over a similar period the age threshold for the CSG was gradually increased from 7 in 1999 to 9 in 2003, 11 in 2004 and finally to 14 in 2007. The effect can be observed from the share of individuals who live in eligible households (Figure 2). A number of studies (Posel, Fairburn, & Lund (2006), for example) have found that being a member of households that receives a social grant affects participation decisions.

Figure 1: Real value of grants (2000 prices), by sample year




Source: Calculated from Eyal & Woolard (2011).

Figure 2: Share of grant receivers, by sample year




Source: Author's calculations from October Household Surveys 1995-99 and Labour Force Surveys 2000-2004 (Statistics South Africa, various years)

### 2.4.2 Data

In order to identify the returns to schooling using the exogenous variation generated by the policy shifts discussed above, we need surveys that span the entire relevant period. Towards this end we use a series of nationally representative household surveys that were conducted by Statistics South Africa (StatsSA) in the post-transition period and which included comprehensive labour market modules: the five October Household surveys collected annually between 1995 and 1999 and the Labour Force Surveys collected bi-

annually between February 2000 and March 2007. The frequency and large sample sizes of these surveys means that we can construct a dataset that is large by African standards. However, a number of papers have investigated the changes in sampling methodology and questionnaire design that occurred between surveys and found that they complicate comparison of the data over time (Casale, Muller, & Posel, 2004; Kingdon & Knight, 2005; Burger & Yu, 2006). We will briefly consider some of these issues before describing how our instrumental variables were generated.

In October of 1993 StatsSA conducted the first of a series of nationally representative household surveys known as the October Household Surveys (OHS), which was replaced by the Labour Force Surveys (LFS) in February of 2000. Until 1996 these surveys used sampling frames based on the 1991 Census and which excluded the nominally independent homelands of Transkei, Bophuthatswana, Venda and Ciskei. However, since the 1995 OHS coincided with an Income and Expenditure Survey, StatsSA went to great lengths to provide a more representative sampling frame for these areas. The improved coverage in 1995 has prompted StatsSA (1996, p. 2) to warn that the 1994 and 1995 surveys were "not directly comparable in all respects", and we therefore follow convention in using 1995 as the baseline year in our study.

The 1996 OHS was the first to use a sampling frame no longer based on the 1991 census. The new frame, which was derived from the preliminary database of enumerator areas (EA) created during the demarcation phase of the 1996 census, had the effect of enumerating more non-urban EAs (Statistics South Africa, 1999, p. 13) where wages are usually lower and unemployment is higher. This shift has led some studies to draw overly pessimistic inferences regarding wage and employment trends by comparing 1995 to subsequent years (Kingdon & Knight, 2005; Casale & Posel, 2002). In 1998 the sampling frame was again adjusted, this time to achieve a more adequate coverage of those residing in mining hostels (Statistics South Africa, 2000, p. 6). In 1999 the first StatsSA master sample was created from the 1996 census, and this was used to draw a sample for the 1999 OHS and the first Labour Force Surveys, until a new master sample based on the 2001 Census was introduced for the September 2004 LFS.

The samples for all the OHSs and LFSs were drawn using a multistage sample design, usually stratifying the sample explicitly by province and area type and implicitly (within magisterial districts) by population group or income levels. Starting with the September 2004 LFS the samples were explicitly stratified by the 53 district councils alone, although they continued to be post-stratified by province. Before this point individuals were easily identified as residing in either urban or rural areas based on the area type strata, but this distinction is no longer possible after 2004. Whereas all of the OHSs were independent cross-sections, the LFSs were originally designed as a 20% rotating panel, although the rotation scheme had to be adjusted in order to address "response fatigue" (Statistics South Africa, 2002, p. i) and possibly some other survey shortcomings as well. Most of the surveys we use sampled around 30,000 households, with the exception of the 1996 OHS (which coincided with the enumeration for the 1996 Census), the 1998 OHS (budget issues) and the February 2000 LFS (pilot study). In this chapter (and the thesis as a whole) we are primarily interested in black males between the ages of 15 and 65 – StatsSA's definition of the working age population – which leaves us with slightly more than 1.3 million observations when using all the available sample years.

The survey questionnaires have also undergone various adjustments since 1995, including changes in the formulation of the questions, differences in how individuals are allowed to respond, and variations in the order in which questions are asked. Such changes have affected the measurement of some of the variables we would like to use in our analysis (including years of education, area of residence, labour market earnings and employment), so that some discussion of these changes is warranted. The questions used in order to determine the highest level of education completed changed slightly between the surveys. In most years respondents were allowed to specify either exactly how many years they completed in the school system or the highest post-secondary certificate, diploma or degree they obtained. However, in certain years the surveys allowed less detailed responses: some surveys grouped grades 1 and 3 into a single category (1995), banded together national technical certificates (NTC) I, II and III with grades 10 to 12, respectively (1995, 1997, 1998), did not distinguish between undergraduate and post-graduate degrees (1995, 1996), between different types of post-graduate degrees or between diplomas and certificates (1995, 1996, 1999 and all the

LFSs until September 2005). Clearly this limits the amount of information that our "educational attainment" measure can reflect, unless we are willing to restrict our analysis to the few years for which the more detailed answers are available. Unfortunately, restricting our sample to these years substantially reduces the variation in our instruments so that we rather construct a slightly less detailed "years of education completed" measure. This variable contains values of 0 to 12 that correspond to the number of years of completed in the school system. NTC I, II and III are counted as 10, 11 or 12 years of education, and those with grades 1-3 in 1995 were assigned 2 years of education. Anyone who obtained a diploma or certificate was given 13 years of education if they were also known to have completed secondary school and 12 years otherwise. All those with degrees were assigned 16 years of education.

Furthermore, detailed questions regarding household characteristics, such as asset ownership or alternative sources of income, were only included sporadically. As discussed in section 2.4.1.3, social grant receipts provide a promising source of employment variation that could be excluded from the earnings regression. Any survey inconsistencies in how this data was collected would therefore restrict the sample years we can use when estimating a sample selection corrected model of the schooling returns. Specifically, some of the surveys asked questions regarding whether the household received different types of government social grants: the child-support and disability grants (all of the OHSs, and the September rounds of the 2000-2004 LFSs), and the social old-age pension (the same, except not captured in 1995).

Although these changes in survey design restrict the survey years we can use or the level of detail with which we can construct certain variables, the more serious comparability issues arise from changes in the way in which labour market earnings and employment were captured. The questions asked to determine the employment status of individuals changed frequently during the early survey years. In 1995 respondents were asked "[w]hat [they] did most during the last 7 days" and were allowed to specify working either part-time or full-time. In 1996 respondents were asked whether "during the past 7 days [they] actually [did] work for pay, profit or family gain?"[14] The 1997 and 1998 surveys added "casual work"

[14] Note that "most" was dropped from the question between 1995 and 1996, so that individuals who did some work, but did not feel that this is what they did most in the last week may have answered in the negative in 1995 but not in 1996.

as a third potential response to the question, and in 1999 this option was relabelled "casual / seasonal work". From 2000 onwards, respondents were asked whether they had engaged in any one of numerous specific (and mostly low-income) activities such as "guarding cars" or "making things for sale" in the preceding week. It is clear that there was an increased effort to capture informal and low-income employment, and that this could lead to an artificial increase in reported employment growth over this period (as well as an artificial decrease in average earnings). A closer analysis (Yu, 2007) shows that the changes in the employment questions and sampling frames are usually mirrored in employment trends. However, most of this artificial employment volatility occurs in the reporting of self-employment and in the informal and agricultural sectors.

Another change in the questionnaire design that is of direct relevance to the current analysis is the responses that respondents were allowed to specify when asked about their labour market earnings. All the surveys provided respondents with the choice of reporting their income either as an actual value or to specify an income interval (except for the 1996 OHS which only allowed intervals). Those who chose to report in intervals were more likely to work in high-paying jobs (Von Fintel, 2007). Furthermore, before the standardisation of the income intervals in 1999, all workers in 1995 and the self-employed in 1996 to 1998 were allowed to specify higher income brackets than were available to respondents in the subsequent years, which contributed to an uneven distribution of outliers across the years. A careful investigation of the raw data also suggests differences in how survey responses were coded or cleaned by StatsSA (Yu, 2007), although no mention of such a shift is made in any of the documentation accompanying the survey data releases. Point estimates were inferred for the earnings of those who reported using earnings bands by running an interval regression on the log of the thresholds for the interval reporters only, and using the mean and variance estimates to assign to each person their conditional expectation of earnings. This assumes a log-normal distribution of earnings, but allows the distribution to differ from those who reported point estimates. Burger and Yu (2006) investigate different ways of addressing these inconsistencies and conclude that a comparable sample of wage earners can be obtained by omitting the .01% of outliers who claimed to earn more than R200 000 per month (many of whom may well represent

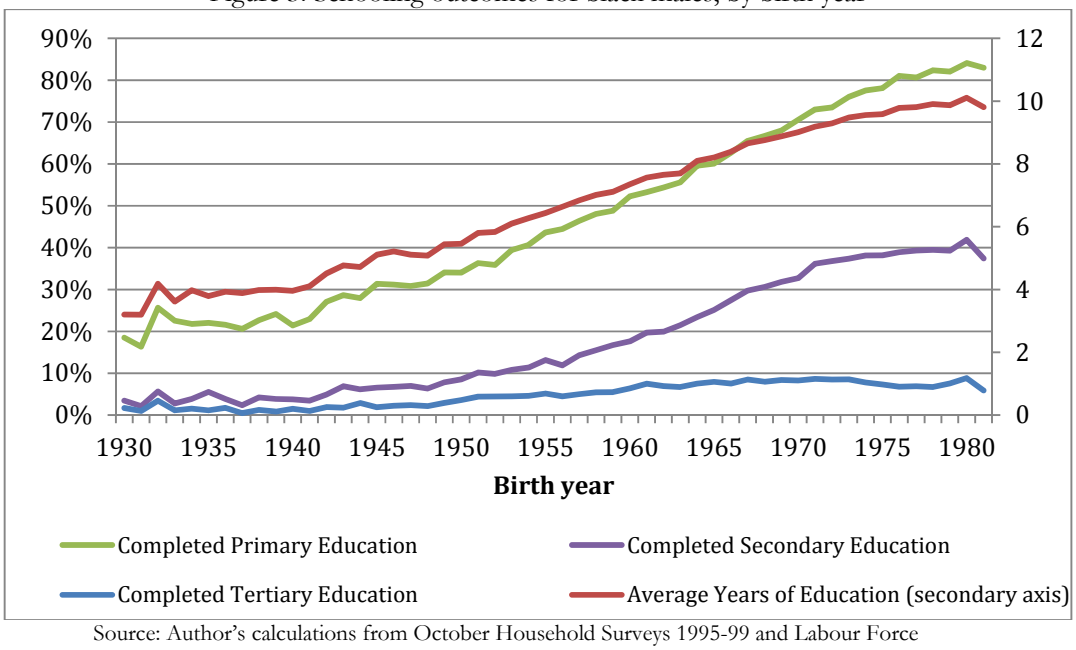
coding errors), zero income reporters and those working in self-employment or the informal sector. For the remainder of this study we will therefore restrict our attention to formal sector employees, not because they are the only workers of interest, but because they are the only workers for whom we have comparable data spanning the whole post-transition period.

### 2.4.3 Instrumental variables

Our econometric models in section 2.3 indicates that we need three sets of exogenous variables for our empirical analysis: conditioning variables, $\boldsymbol{z}_{it1}$, instrumental variables, $\boldsymbol{z}_{it2}$, and employment exclusion restrictions, $\boldsymbol{z}_{it3}$. When estimating the earnings function we control for either experience or age (linearly and quadratically), province of residence and sample year. Our sample is restricted to black males, which serves to further control for the effects of ethnicity and gender, two dimensions that still interact with many other features of the South African labour market. Some of our specifications also include a birth year variable to control for unobserved differences across generations, like education quality.

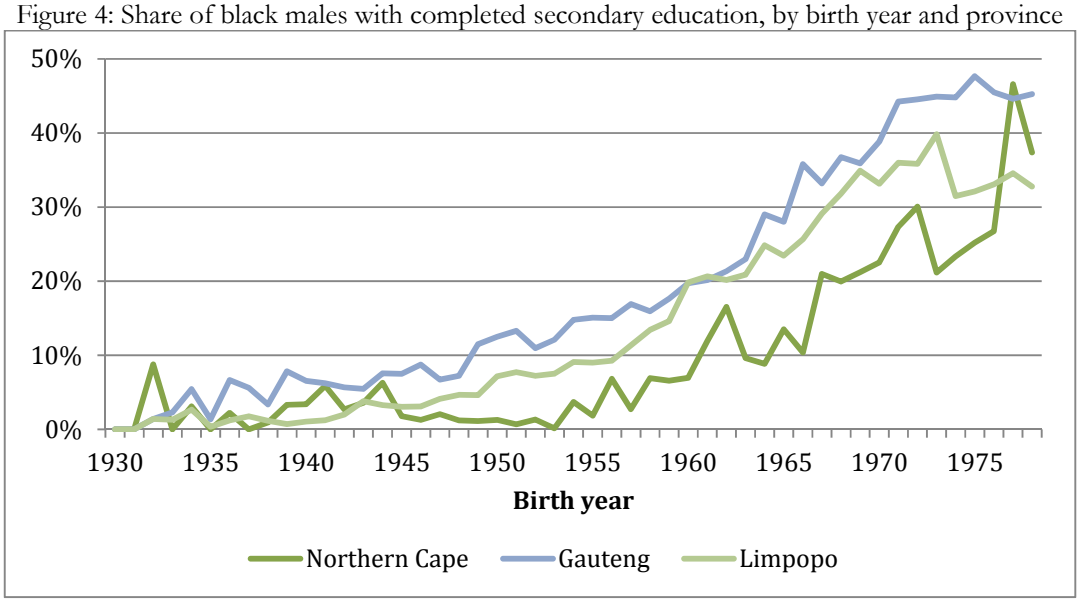
The empirical analysis below uses two sets of instrumental variables derived from policy changes that affected individual schooling decision. The first is the expansion of access to education as apartheid restrictions on black education were gradually relaxed, and the second is the policy of normalising the ages of learners in the education system. We now discuss how these policy shifts are operationalised into instrumental variables for our econometric analysis in section 2.5.

Figure 3: Schooling outcomes for black males, by birth year




Source: Author's calculations from October Household Surveys 1995-99 and Labour Force Surveys 2000-2007 (Statistics South Africa, various years)

Figure 4: Share of black males with completed secondary education, by birth year and province




Source: Author's calculations from October Household Surveys 1995-99 and Labour Force Surveys 2000-2007 (Statistics South Africa, various years)

As demonstrated in Figure 3, the rollback of apartheid restrictions had the expected effect on the schooling outcomes of black males, with those born more recently being much more likely to obtain

higher levels of education[15]. Black males from younger generations were clearly more likely to have had access to educational institutions, and if schooling indeed has a causal effect on earnings we would expect to observe a commensurate effect on the earnings of more recently born cohorts. Furthermore, many factors that affect the schooling decision – such as distance to the nearest school, school fees or the aggressiveness with which new policies were implemented – also possess a geographical dimension. This is illustrated in

Figure 4, which depicts the share of black males older than 25 with completed secondary education, by province of residence and birth year. Clearly black males currently residing in the highly urbanised Gauteng province had a better chance of completing high school than their counterparts in the more rural provinces of the Northern Cape or Limpopo. Improvements in access to secondary school also occurred at an earlier date (and hence affected older cohorts) for those residing in Gauteng than for those from Limpopo, where progress in turn preceded that in the Northern Cape.

Assuming that these increases in years of completed schooling were – after partialling out the time-invariant provincial effect, sample year effects, and perhaps a universal birth year time trend – the result of changing policies with respect to the education system, we can calculate the cohort average of various schooling outcomes and interpret these variables as a policy-induced variations in the cost of schooling. Belonging to a cohort with a higher probability of schooling outcomes implies a decreased likelihood of having faced apartheid restrictions on access to education oneself. The differential access to education across black cohorts is best conceptualised in the Becker-Card model as different values of the marginal cost of education $r_i$. Since political transition was associated with an increase in educational attainment across the whole range of schooling years (as shown in Figure 3) rather than at just one specific point, this offers a set of instrumental variables with a rare ability to identify the shape of the schooling-earnings

[15] This figure, and indeed all of the regressions estimated below, uses a series of cross-sectional surveys between 1995 and 2007 and therefore ignores any mortality trends which – given the potential correlation between schooling outcomes and life expectancy – may distort the actual pattern of education attainment across birth cohorts. However, we believe that this bias would tend to under- rather than overstate the increase in educational attainment over successive birth cohorts.

profile. In order to gauge the effect of apartheid legislation on schooling decisions, the share of black men that completed the various schooling grades were calculated for each province-birth year combination. The education level of the observation itself was not used in this calculation. In this way we obtain fifteen[16] instrumental variables that each reflects the impediments faced by individuals of a certain cohort in gaining access to and completing a certain grade.

The validity of this instrument depends on the assumption that the expected unobserved abilities of individuals from different cohorts are uncorrelated to the schooling restrictions imposed on them by apartheid legislation. This exogeneity condition can be conceptualised as consisting of two parts: firstly, that the likelihood and severity of restrictions were uncorrelated to the cohort's potential schooling returns and secondly, that the instrumental variable affects wages only through its effect on schooling outcomes. As discussed in section 2.4.1.1, South African schooling laws were mainly driven by political ideology, internal and external pressures and macro-economic conditions rather than their potential effect on students. Similarly, the geographic variation in schooling restrictions were probably driven by domestic political and cost considerations more than anything else. In this sense, the "treatment" can be considered to be as-good-as-randomly distributed across individuals in our sample, at least after controlling for province and birth year effects.

However, there are at least three reasons why the variation in our instrumental variables may be correlated to earnings potential apart from its effect on schooling outcomes: migration patterns, the macro-economic effects of political liberalisation, and changes in educational quality. Much of migration is driven by geographical differences in expected labour market outcomes and if high-ability (and education) individuals all migrate to high wage provinces such as Gauteng, then this may induce a correlation between province of residence and earnings which is not caused purely by variations in education. If this migration pattern was identical across birth years, then including province of residence amongst the conditioning variables

[16] As discussed in section 2.4.2, the pooled OHS-LFS dataset allows us to distinguish between fifteen categories of educational attainment, and the cohort averages of these schooling years give us fifteen instrumental variables. However, these variables are restricted to add up to unity, so that the variable representing the share of each cohort that completed grade 0 is omitted from our regressions in section 2.5.

should address this concern, but since the process of political transition also entailed the relaxing of restrictions on black migration, this pattern may be stronger for more recent birth years. In order to investigate whether our instruments actually affected earnings through this channel, section 2.5.3 re-estimates our earnings regressions using IVs with only birth year and no geographical variation.

The second objection is that political liberalisation coincided with a changing macroeconomic environment, such as relaxing restrictions on capital flows, changing perceptions amongst foreign investors or the removal of trading sanctions. One would expect such changes to affect labour market outcomes, and if there is a high degree of persistence in such outcomes (Beaudry & DiNardo, 1991) then higher wages may be associated with those who went to school in more liberalised period for reasons other than their higher levels of education. These objections can be addressed by also controlling for macro-economic variables (such as the interest or economic growth rate at the time of entering the labour market or the current cohort unemployment rate) in the regression. The effect of adding these additional controls is also explored in section 2.5.3.

Finally, the sudden increase in access to education may have attributed to a deterioration in the quality of schooling. This implies that the same individuals who were more likely to obtain access to education (by province or birth year) are also more likely to have received a poorer quality of education. All other things being equal, this would induce a negative correlation between our first set of schooling instrumental variables and wages, abstracting from the causal effect that works through greater educational attainment. In section 2.5.3 we also re-estimate our earnings regression while controlling for an exhaustive set of birth year dummies to control for unobserved birth year productivity differentials that are not attributable to differences in educational attainment.

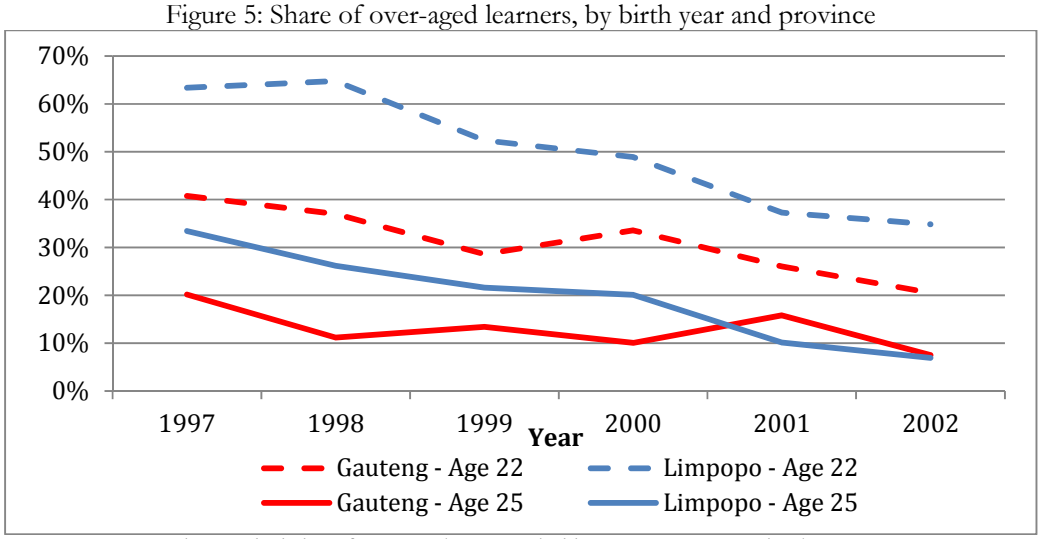
The second schooling policy shift we consider is the measures taken by the DoE to remove the large number of over-aged learners from the school system. Figure 5 compares the share of black males of various birth years and in different provinces who would have been classified as over-aged learners at different ages, as calculated from household surveys between 1995 and 2007. The effect of the age

normalisation policy is immediately evident. The same trend can be observed from administrative data, which shows a 21% decline in matric candidates between 1997 and 2002. Figure 5 also demonstrates that this policy was never uniformly enforced, and that there is time and geographic variation in its implementation. This provides another set of useful instrumental variables for schooling. There is little reason to suspect that members in the cohorts (defined according to province and birth year) that were subject to the stricter enforcement of the age restrictions would have different unobservable characteristics than those from unaffected cohorts. On the other hand, since the policy was aimed at school repeaters older than 15, the policy shift probably had a larger effect on those with lower than average abilities, at older ages and higher than average schooling years.

In order to operationalise this policy into instrumental variables we calculate the share of "would-be over-aged learners" – those whose age exceeds their schooling level by more than 8 years – who are still in the school system, for individuals of each age-year-province combination. Since the implementation of this policy occurred between 1997 and 2002, we group together the samples before 1998 and after 2001 and extrapolate their averages to all those who attended school before or after the sample period. The resulting variables give us an indication of how likely a person would be (based on their province and birth year) to have still been allowed to enrol in secondary school as an over-aged learner at different ages. In our analysis we use the variables for ages 16 to 26, which give us another eleven IVs.

One objection to the exogeneity of this set of IVs may be that the principals of schools in regions (or periods) of relatively high unemployment or low wages may have faced more pressure from the community not to expose their students to the labour market, and if the policy was less likely to be enforced in such schools this would induce a correlation between wages and the second set of schooling instruments that does not work purely through variation in educational attainment. This concern can be investigated by including the cohort unemployment rate and the economic growth rate while in school as a conditioning variable in the schooling and earnings regression, which is done in section 2.5.3.

Figure 5: Share of over-aged learners, by birth year and province




Source: Author's calculations from October Household Surveys 1995-99 and Labour Force Surveys 2000-2007 (StatsSA: various years).

In section 2.3.3.1 we briefly mentioned Card's (2001, p. 1138) concerns regarding IVs that affect the mapping from the relative returns parameter to schooling outcomes. This point is worth considering in the context of our own schooling model too. Even if the instruments may have no effect on the expected value of the relative schooling returns, $E(v_i|\boldsymbol{z}_{it}) = 0$, it is still possible that the correlation between the returns and schooling depend on the value of the instruments. For example, it is conceivable that during the height of apartheid only students with very strong academic abilities were able to proceed beyond secondary school, whereas the mapping from unobservable abilities to schooling has since grown weaker. Similarly, for the second policy it is possible that by weeding out poorer students the stricter enforcement of over-aged enrolment restrictions may have resulted in a stronger mapping from high returns to high education outcomes. In any of these cases the assumptions required for 2SLS to consistently estimate the ATE will be violated. Although the CF estimator does not require the same assumption in order to identify the shape of the schooling-earnings profile, its own identifying assumptions imply something similarly strong about how instruments are allowed to interact with the unobservable components. The literature in

this regard is still embryonic, and there does not appear to be any solid guidelines in how to evaluate the validity of these stronger IV assumptions[17].

## 2.5 Econometric analysis

We now proceed to estimate the different econometric models discussed in section 2.3, using the instrumental variables and exclusion restrictions that were argued to provide exogenous variation in schooling and employment in section 2.4.3 and the data discussed in section 2.4.2. Section 2.5.1 starts by analysing the first-stage regressions on years of schooling and employment. The importance of the schooling regression in our particular identification strategies were discussed in section 2.3: our LATE approach requires that different IVs shift schooling outcomes at different points of the schooling distribution, whereas the validity of the CF estimator depends on our ability to estimate a correctly specified schooling equation. This is followed by the earnings regressions interpreted according to the LATE approach in section 2.5.2.1 and the average schooling-earnings profile in section 2.5.2.2. Finally, section 2.5.3 performs some additional regressions to investigate alternative channels through which our IVs could be correlated to expected earnings and which could potentially invalidate our instrument exogeneity assumption.

### 2.5.1 Schooling and employment regressions

All of our IV estimators require instrumental variables that are informative about individuals' schooling decisions. A natural point of departure is therefore to investigate the estimates from a first-stage regression of years of schooling completed on the instrumental and conditioning variables as per equation [8]. Given

[17] One possibility may be to use an overidentifying restriction test. Since the stronger assumptions allow us to estimate the ATE, which is the unconditional expectation of the heterogeneous returns component, this estimate should be invariant to the choice of (valid) instrumental variables. Comparing whether different IVs produce different estimates would thus offer one way to test whether the IVs fulfil the stronger exogeneity conditions stated in assumptions 2SLS or CF, assuming that the weaker conditions in *Assumption LATE* are met.

the relatively strong exogeneity condition required when using the CF estimator – a condition that essentially requires that the first-stage regression is a correctly specified function of the schooling decision – this equation should be scrutinised more carefully to make sure that it also makes sense as a structural equation, rather than it just being used to generate weak instrument test statistics. In particular, we need to consider how expected schooling years completed depends on an individual's age or experience, birth year, province of residence, the sample year and the two sets of policy-related instrumental variables.

Most individuals choose to complete their education before entering the labour market, so that the age or experience regression coefficients in a schooling equation should imply a minimal effect for those approaching their late twenties. Given South Africa's history of racial restrictions on access to education, one may expect the period directly following the end of apartheid to have included a larger group of adults who returned to the education system to obtain their desired level of schooling. Section 2.4.1.2 also alluded to the tendency of unsuccessful learners to remain in the school system for longer than is the case in most other countries. These facts notwithstanding, one would not expect the *average* individual's years of completed education to increase by much once they reach their late-twenties.

The secular rise in educational attainment observed in most countries, particularly those in Africa, would suggest that education is a positive function of birth year for reasons other than apartheid era schooling restrictions coming to an end. We would therefore expect to find a positive birth year coefficient in a structural schooling equation. Once the birth year and age effects are completely controlled for, the additional sample year effect should be small, especially when already controlling for policy changes that may have shifted the schooling cost over time via our instrumental variables. Geographical differences in the concentration of schools or average school fees may induce a correlation between province and education outcomes, and one would expect those residing in urban areas to have had easier access to schools and hence to have higher expected schooling outcomes. Finally, in section 2.4.3 we argued that our instrumental variables – representing policies that expanded the access to education in line with political reform and the stricter restrictions on the enrolment of over-aged learners – should also be correlated to

education outcomes. We would expect schooling outcomes to be positively correlated to both the first and second sets of instruments, as defined in section 2.4.3.

Two multicolinearity issues arise in estimating the schooling regression. Firstly, age, birth year and sample year are linearly dependent and so cannot be included together without imposing additional restrictions on how these variables enter the schooling function. Secondly, potential experience is defined as the difference between age and schooling years[18] and will therefore be correlated with completed schooling years for reasons that have nothing to do with the causal effect of age/experience on schooling, particularly if the other control variables are able to closely replicate the sample variation in age. In order to avoid this problem we use age rather than experience when controlling for the experience-earnings profile in the earnings regression.

Table 1 reports the first-stage regression estimates for different specifications. The regression in the first column includes the instrumental variables associated with the first policy (the roll-back of apartheid era schooling restrictions), the second includes the instrumental variables associated with the second policy (the stricter enforcement of restrictions on access to over-aged learners) and the third column includes all the instrumental variables associated with either policy. All three regressions also control for age, sample year and province of residence.

[18] Experience is actually defined as age minus schooling years completed minus six (the typical school-going age) and with an adjustment so that experience only starts accumulating after the legal working age is reached. Although this adjustment rules out perfect multicolinearity, the partial correlation between experience and schooling is still very close to minus one when also controlling for birth year.

Table 1: First-stage OLS regressions on years of education (sample of black males), 1995-2007

| | | | |
|---|---|---|---|
| Grade 1 (cohort %) | -0.844 | | -1.261 |
| Grade 2 (cohort %) | 0.833 | | 0.499 |
| Grade 3 (cohort %) | 0.374 | | 0.312 |
| Grade 4 (cohort %) | 1.27*** | | 1.129** |
| Grade 5 (cohort %) | 1.422*** | | 1.188*** |
| Grade 6 (cohort %) | 1.111*** | | 0.845** |
| Grade 7 (cohort %) | 1.332*** | | 1.089*** |
| Grade 8 (cohort %) | 1.059*** | | 1.026*** |
| Grade 9 (cohort %) | 1.596*** | | 1.27*** |
| Grade 10 (cohort %) | 2.131*** | | 1.934*** |
| Grade 11 (cohort %) | 3.102*** | | 2.851*** |
| Grade 12 (cohort %) | 5.331*** | | 4.698*** |
| Post-school dipl. /cert. (cohort %) | 6.29*** | | 5.359*** |
| Degree (cohort %) | 4.796*** | | 3.851*** |
| P(allow overage), age = 16 | | 1.933** | 1.252 |
| P(allow overage), age = 17 | | 3.573*** | 2.043** |
| P(allow overage), age = 18 | | 3.84*** | 2.985*** |
| P(allow overage), age = 19 | | 2.192*** | 0.701 |
| P(allow overage), age = 20 | | 1.469*** | 0.945** |
| P(allow overage), age = 21 | | -0.474 | -0.595 |
| P(allow overage), age = 22 | | 0.859** | 0.3 |
| P(allow overage), age = 23 | | 1.799*** | 0.816* |
| P(allow overage), age = 24 | | 1.034** | 1.282*** |
| P(allow overage), age = 25 | | 1.138* | 0.28 |
| P(allow overage), age = 26 | | -0.685 | 0.264 |
| Age | 0.19*** | 0.036*** | 0.11*** |
| Age^2 | -0.003*** | -0.002*** | -0.002*** |
| Year = 1996 | 0.186** | 0.119 | 0.187** |
| Year = 1997 | 0.441*** | 0.432*** | 0.457*** |
| Year = 1998 | 0.057 | 0.026 | 0.074 |
| Year = 1999 | 0.244*** | 0.278*** | 0.285*** |
| Year = 2000 | 0.267*** | 0.349*** | 0.326*** |
| Year = 2001 | 0.424*** | 0.611*** | 0.509*** |
| Year = 2002 | 0.564*** | 0.866*** | 0.678*** |
| Year = 2003 | 0.624*** | 1.031*** | 0.767*** |
| Year = 2004 | 0.688*** | 1.196*** | 0.859*** |
| Year = 2005 | 0.913*** | 1.547*** | 1.122*** |
| Year = 2006 | 1.08*** | 1.87*** | 1.334*** |
| Year = 2007 | 1.046*** | 1.93*** | 1.34*** |
| Western Cape | -0.623*** | 1.011*** | 0.325* |
| Eastern Cape | 0.254*** | 0.146 | 0.332*** |
| Northern Cape | -1.764*** | 0.79*** | -0.079 |
| Free State | -1.025*** | -0.646*** | -0.707*** |
| Kwazulu-Natal | -0.159** | 0.604*** | 0.283** |
| North-West | -0.656*** | 0.945*** | 0.331** |
| Gauteng | 0.053 | 1.113*** | 0.401*** |
| Mpumalanga | -0.929*** | -0.932*** | -0.854*** |
| Constant | 3.671*** | -2.064 | -1.505 |
| Observations | 129659 | 129659 | 129659 |
| R-squared | 0.18 | 0.176 | 0.181 |
| F-test (IV sets 1 & 2) | 48.34 | 30.976 | 31.642 |
| F-test (IV set 1) | 48.34 | | 31.785 |
| F-test (IV set 2) | | 30.976 | 8.923 |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

The first column shows that the effect of relaxing apartheid-era restrictions on black schooling access was as expected: the coefficients on the cohort shares that completed each grade are positive, significant and generally an increasing function of the completed schooling year. The results from the second column indicate that an increased likelihood of being allowed to enrol in the school system as an over-aged student (at various ages) is generally associated with a higher level of expected schooling[19]. The coefficients are somewhat larger for being allowed to enrol at younger ages, which suggest that the effect of over-aged restrictions were stronger when applied to those who were only a few years over the prescribed grade age. The instrumental variables in columns 1 and 2 both have an overall significant effect on the schooling decisions of individuals, as per the F-test statistics, although the first policy is relatively more important in explaining schooling outcomes. The third column includes the two sets of instruments simultaneously. The coefficient patterns observed from columns 1 and 2 are generally maintained, although the magnitude of most coefficients as well as the F-tests associated with each IV set is reduced, indicating that there is some overlap in the schooling variation explained by the two sets of IVs.

Age is included as a quadratic function mainly in anticipation of the second-stage earnings regression which is required to use the same conditioning variables, and in which we want to control for the typical concave experience-earnings profile observed in Mincerian earnings equation. However, the implied schooling-age pattern – according to which schooling increases rapidly in the teens and early twenties and then reaches a maximum in the mid-twenties – also seems plausible from a causal perspective. The year effects reveal an increase in educational attainment over the period under consideration, but is a relatively minor contributor to the overall variation in schooling. The provincial controls indicate significant geographical variation in schooling patterns. Living in Gauteng – the most urbanised province with, presumably, the highest concentration of schools – is associated with the highest expected years of completed schooling. Table 15 in the appendix re-estimates this regression using a more flexible age

[19] Interestingly many of these IV coefficients are negative if we do not also control for either the effect of the first policy or the observed wage determinants. The probability of being allowed to enrol in school as an over-aged learner is lower for more recent generations, who are the same individuals that also benefited from the shifts in education policies that accompanied political transformation. If we do not control for age, birth year or the first policy, the coefficients on the second policy IVs will adjust in order to replicate the more prominent effect of the first policy.

specification and controlling for birth year, and shows little evidence to suggest that our coefficients are biased by the more restrictive specification in Table 1. In general the schooling regression coefficients are consistent with an interpretation that views this regression as a correctly specified schooling equation, the residual of which provides an estimate of the unobserved schooling cost and ability terms as per the Becker-Card model. This should give us somewhat more confidence in making the stronger assumptions about the schooling process that are required for the CF estimator.

The ability to use the LATEs of different instruments to plot out a schooling-earnings profile rather than just a single point on this curve depends on the assumption that the different IVs affect the choices of individuals at different schooling years. Intuitively, one would expect the restrictions at lower schooling levels (as measured by the cohort-grade completion rates) to be associated with an individual's decisions at lower schooling years. Table 2 allows us to evaluate this assumption by comparing the probit estimates (using all of the IVs while controlling for province, sample year and age) of whether an individual crossed certain schooling thresholds. Columns 1 to 5 report the results from whether the individual completed at least four, seven, eleven, twelve or thirteen years of schooling. The second category corresponds to completed primary schooling and the last two categories to completing secondary school and any post-school qualifications, respectively.

There are three important patterns that can be observed from Table 2, where the cells containing coefficient estimates that are positive and significant are shaded to highlight these trends. Firstly, restrictions at lower schooling years affected mainly those outcomes lower down the schooling ladder, whereas the effects of restrictions at later secondary and tertiary education were more evenly distributed across the schooling years. This is consistent with the notion of individuals who are forward-looking when making schooling decisions and therefore consider the cost associated with the current and all future schooling years whilst ignoring the costs of completed schooling levels.

Table 2: Probit regressions on crossing various education thresholds (sample of black males), 1995-2007

| DEPENDENT VARIABLE | educ >= 4 | educ >= 7 | educ >= 11 | educ >= 12 | educ >= 13 |
|---|---|---|---|---|---|
| Grade 1 (cohort %) | -0.696* | -0.384 | -0.38 | -0.241 | -0.571 |
| Grade 2 (cohort %) | -0.295 | -0.044 | 0.158 | 0.36 | 0.366 |
| Grade 3 (cohort %) | 0.044 | 0.113 | -0.227 | -0.111 | 0.015 |
| Grade 4 (cohort %) | 0.254 | 0.04 | -0.021 | -0.073 | 0.045 |
| Grade 5 (cohort %) | 0.29* | 0.369** | 0.04 | -0.108 | -0.377 |
| Grade 6 (cohort %) | 0.292* | -0.05 | -0.339** | -0.265* | -0.296 |
| Grade 7 (cohort %) | 0.328** | 0.063 | -0.237* | -0.096 | -0.018 |
| Grade 8 (cohort %) | 0.338*** | 0.209* | -0.358*** | -0.22* | -0.36** |
| Grade 9 (cohort %) | 0.599*** | 0.197 | -0.517*** | -0.427*** | -0.422** |
| Grade 10 (cohort %) | 0.575*** | 0.46*** | -0.105 | 0.025 | -0.361* |
| Grade 11 (cohort %) | 1.018*** | 0.76*** | 0.485*** | 0.391** | 0.046 |
| Grade 12 (cohort %) | 1.285*** | 1.402*** | 1.175*** | 1.07*** | 0.573*** |
| Post-school dipl. /cert. (cohort %) | 1.038*** | 1.419*** | 1.443*** | 1.471*** | 1.394*** |
| Degree (cohort %) | 0.421 | 0.412* | 0.759*** | 0.903*** | 1.439*** |
| P(allow overage), age = 16 | 1.096 | 0.637 | 0.311 | -0.191 | 0.262 |
| P(allow overage), age = 17 | 0.552 | 1.61*** | 0.418 | 0.519 | 1.542*** |
| P(allow overage), age = 18 | 1.262** | 1.44*** | 0.242 | 0.378 | 1.782*** |
| P(allow overage), age = 19 | 0.088 | 0.336 | 0.485* | 0.683*** | 0.237 |
| P(allow overage), age = 20 | -0.287 | -0.299 | 0.18 | 0.411** | -0.184 |
| P(allow overage), age = 21 | 0.328 | 0.006 | 0.086 | -0.105 | -1.02*** |
| P(allow overage), age = 22 | -0.106 | -0.184 | 0.319** | 0.327** | 0.274 |
| P(allow overage), age = 23 | -0.116 | 0.407* | 0.329* | 0.379** | 0.535** |
| P(allow overage), age = 24 | 0.341 | -0.203 | -0.168 | 0.072 | 0.994*** |
| P(allow overage), age = 25 | -0.44 | 0.08 | 0.403 | 0.442* | 0.066 |
| P(allow overage), age = 26 | 0.371 | -0.116 | -0.2 | 0.177 | 0.584** |
| Age | 0.022*** | 0.032*** | 0.032*** | 0.019*** | 0.068*** |
| Age^2 | -0.001*** | -0.001*** | -0.001*** | -0.001*** | -0.001*** |
| Constant | -1.998** | -3.286*** | -2.395*** | -2.516*** | -5.301*** |
| Observations | 129659 | 129659 | 129659 | 129659 | 129659 |
| (Pseudo) R-squared | 0.122 | 0.113 | 0.108 | 0.083 | 0.051 |
| Chi2/F-test (all IVs) | 190.48 | 424.475 | 769.539 | 703.397 | 365.829 |
| Chi2/F-test (IV set 1) | 153.897 | 305.991 | 495.194 | 367.396 | 144.438 |
| Chi2/F-test (IV set 2) | 11.162 | 40.741 | 62.384 | 122.243 | 117.365 |
| Also control for: | Sample year, province | | | | |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Secondly, there is a tendency for enrolment restrictions at later ages to have an impact at higher schooling levels. This is somewhat puzzling given the fact that enrolment restrictions at younger ages generally had

larger coefficients in Table 1. In fact, the same pattern can be also be observed in Table 2, so that we may conclude that enrolment restrictions for older over-aged students had a weaker effect but mainly on those individuals at more advanced schooling years. Finally, the strength of the IVs (as reported by the F-tests) indicate that whereas the second policy had a relatively weak effect on early schooling years compared to its much stronger impact on secondary and tertiary schooling decisions, the effect of the first policy was more evenly distributed across the schooling years. This is entirely consistent with what one would expect from our discussion of the policies in section 2.4.1.

Section 2.2.1.2 discussed the problems that a highly selective employment process can causes in an analysis like ours and sections 2.3.2.2 and 2.3.3.2 suggested a few techniques to address this problem. The first step of such an approach usually requires modelling the employment outcome using the same conditioning variables as the wage equation as well as a set of variables that will affect the individual's employment decision without directly influencing earnings potential. The dependent variable is a dummy indicating whether each individual is in the sample of formal sector wage employees that we use to estimate our earnings regressions. We obtain exogenous variation in employment from the amount that the individual's household received in the form of either the OAP or the CSG. As explained in section 2.4.2, using these variables restricts our sample to the years in which the OHS and LFS included questions about non-labour income sources. We also control for schooling, age, sample year and province.

Table 3: Formal sector employment probit regressions (sample of black males), 1995-2007

| Generated variable | IMR1 | IMR2 | IMR3 |
|---|---|---|---|
| Education | -0.034*** | 0.163*** | 0.168*** |
| Education^2 | 0.006*** | 0.005*** | 0.005*** |
| Education residual | | -0.193*** | -1.358*** |
| Education residual*Education | | | 1.16*** |
| Age | 0.251*** | 0.24*** | 0.232*** |
| Age^2 | -0.003*** | -0.002*** | -0.002*** |
| Log child support grant | -0.03*** | -0.031*** | -0.074*** |
| Log old age pension | -0.053*** | -0.054*** | -0.084*** |
| Constant | -6.077*** | -7.764*** | -8.101*** |
| Observations | 199357 | 199357 | 199357 |
| Pseudo R-squared | 0.221 | 0.224 | 0.224 |
| Also control for: | Sample year, province | | |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Column 1 in Table 3 estimates the employment probit under the assumption that all the regressors, including schooling, are exogenous. The schooling-employment probability profile is convex, which confirms the notion that more advanced schooling years have a larger effect on an individual's labour market productivity. The age-employment profile is concave and reaches a maximum in the mid-40s. The amount of grant money received had a negative effect on the employment likelihood of individuals, and this effect was greater for the OAP than for the CSG.

If our analysis allows for endogenous schooling in the earnings process then it seems reasonable to also be concerned about the exogeneity of schooling in our employment process. In order to allow for endogenous schooling in our probit regression we use the Rivers & Vuong (1988) approach and regress employment on the same conditioning variables as in column 1, but now also control for the residual from a regression of equation [8] on the whole sample for whom grant data is available (rather than just formal sector employees). The schooling residual is highly significant in this regression, which confirms that our concerns regarding the exogeneity of schooling were justified. Treating schooling as endogenous has the effect of dramatically increasing the schooling coefficient and marginally decreasing the degree of convexity of the schooling-employment profile. The remaining coefficients are not substantially affected by this inclusion.

Finally, our analysis in section 2.3.1.2 suggests that if employment is determined by the difference between the offered and reservation wage then we should expect the schooling coefficient in the employment equation to also be heterogeneous. Unfortunately, this more flexible employment specification more or less precludes us from appropriately controlling for sample selection in the earnings regression. This problem notwithstanding, the results from a probit regression in which schooling is treated as endogenous and its effect on employment as heterogenous are reported in column 3. Although the schooling residual-schooling interaction term is significant in the employment probit, the schooling residual's coefficient is the only one to change substantially between columns 2 and 3. One way to interpret this result is that treating the schooling coefficient as homogeneous is not as important a source of bias in the employment

process as the endogeneity of schooling. When used to calculate inverse Mills ratios[20] (IMR) for our earnings regressions below, the different employment specifications produce variables IMR1, IMR2 and IMR3 respectively.

### 2.5.2 Earnings regressions

We now proceed to the earnings regressions from which the causal effect of an additional year of schooling on labour market earnings can be inferred. As discussed in section 2.3, there are at least two ways to gauge whether the schooling-earnings profile is a concave, linear or convex function of schooling: estimating and comparing LATEs using IVs that affected the schooling distribution at different points, and estimating the parameters from the Becker-Card model in order to construct the schooling-earning profile for the "average" wage earner.

#### *2.5.2.1 The local average treatment effect interpretation*

The LATE of schooling is estimated using a 2SLS regression on an earnings equation with a linear schooling term (as in equation [2]). This estimator calculates the weighted average of the marginal schooling returns for those individuals whose schooling outcome was affected by a change in the IV values. The weighting scheme attaches more weight to those schooling years where the specific IV had a larger effect and to types of individuals who responded more strongly to a change in the IV values.

Table 4 compares the OLS estimates of the earnings regression to the LATEs associated with the different sets of instrumental variables. The coefficients in column 1 are from an OLS regression of monthly earnings on education, age, sample year and province, and reveal that the OLS returns to schooling are just below 11%. As discussed in section 2.2, this is high by international standards but towards the lower end of the range of coefficient estimates from South African studies. The concave age-earnings profile

[20] These variables are calculated as described in footnote 13.

conforms to our expectations. Columns 2 to 4 use 2SLS and different combinations of the IVs to estimate the schooling returns in the presence of endogenous schooling outcomes. The estimated returns are 11.5% when using all of the instruments at our disposal (column 4) which is – as is typical in this literature – slightly higher than the OLS returns. The schooling coefficient is 1 percentage point lower when using only the IVs associated with the first policy shift (column 2) and 2 percentage points higher when only exploiting the schooling variation associated with the second policy shift (column 3).

The J-test statistic (and associated p-value) investigates whether the different instrumental variables produce similar schooling coefficients and can be interpreted as a test of instrument validity in the constant coefficient model. However, in models that allow for heterogeneous schooling coefficients, and for different individuals to respond differently to changes in the IVs, we would not expect the different IVs to produce the same LATEs. In fact, our ability to identify the slope of the schooling-earnings profile depends crucially on the assumption that the different IVs affect the choices of individuals at different schooling years. In this context the high J-test statistics merely reflect that the slope of the schooling-earnings profile is unlikely to be linear.

Table 4: OLS and 2SLS earnings regressions using different IVs (sample of black males), 1995-2007

| | **OLS** | **2SLS** | **2SLS** | **2SLS** |
|---|---|---|---|---|
| Education | 0.107*** | 0.105*** | 0.137*** | 0.115*** |
| | *(0.0008)* | *(0.0079)* | *(0.0124)* | *(0.0075)* |
| Age | 0.095*** | 0.095*** | 0.09*** | 0.093*** |
| | *(0.0018)* | *(0.0022)* | *(0.0026)* | *(0.0021)* |
| Age^2 | -0.0009*** | -0.0009*** | -0.0007*** | -0.0008*** |
| | *(0.00002)* | *(0.00004)* | *(0.00005)* | *(0.00003)* |
| Observations | 129659 | 129659 | 129659 | 129659 |
| R-squared | 0.296 | 0.296 | 0.28 | 0.295 |
| CD F-test | | 83.664 | 48.604 | 53.03 |
| J-test | | 88.181 | 211.201 | 274.689 |
| p-value | | 0.000 | 0.000 | 0.000 |
| Instrumental variables | None | IV set 1 | IV set 2 | IV set 1 & 2 |
| Also control for: | Sample year, province | | | |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Given our discussion in section 2.4.1, we would expect the effect of the second policy to be restricted to schooling decisions for those already in secondary school (which may also have a smaller knock-on effect on tertiary education decisions) whereas the effect of the first policy shift would have been much more evenly distributed across the schooling years. A convex schooling-earnings profile for the average individual would therefore imply that the IVs associated with the second policy shift should produce higher LATEs than those associated with the first policy shift, which is exactly what we observe in Table 4. Furthermore, the 2SLS estimate using all IVs (column 4) is a weighted average of the averages produced in columns 2 and 3, with more weight attached to the instruments that had a larger effect on schooling outcomes: IV set 1 in our case. Again, this interpretation is in line with what we observe in Table 4. Note that this explanation for why the 2SLS estimates exceed the OLS schooling returns is very different from the once popular interpretation that ability bias is relatively unimportant compared to measurement error.

The Cragg-Donald F-test statistics[21] are much larger than 10 in all cases, indicating that we need not worry about weak instruments. Table 16 in the appendix re-estimates these regressions while controlling for birth year effects, and finds that the general pattern observed in Table 4 is maintained. The fact that the 2SLS schooling coefficients are no lower when controlling for birth year suggests that the correlation between our instruments and earnings are not the result of omitting birth year-specific unobservables such as varying educational quality or macro-economic trends. Further sensitivity analyses are performed in section 2.5.3.

In line with the LATE interpretation of the results, we now further explore the estimated returns by using each of the different instrumental variables at our disposal separately. The schooling coefficients from these regressions, which control for age, sample year and province are shown in column 1 of Table 5. The probit schooling regressions in Table 2 showed that i) amongst the IVs associated with the first policy, restrictions at lower schooling years affected mainly earlier schooling years, ii) amongst the second policy

[21] Stock and Yogo (2005) define a weak instrument in terms of its effect on the 2SLS bias (relative to that of OLS) and its distortion of Wald tests sizes. Their test statistic is based on that of Cragg and Donald (1993), which simplifies to the regular F-test statistic in the case of a single endogenous regressor.

Table 5: 2SLS education coefficients using different instrumental and conditioning variables (sample of black males), 1995-2007

| **Controls** | **Age, year, province** | | **Age, year, birth year, province** | |
|---|---|---|---|---|
| **Instrumental variables** | **Coefficient** | **F-stat** | **Coefficient** | **F-stat** |
| Grade 1 (cohort %) | -0.014 | 29.15 | 0.023 | 17.12 |
| Grade 2 (cohort %) | -0.082 | 10.47 | -0.085 | 1.8 |
| Grade 3 (cohort %) | 0.09** | 40.34 | 0.108 | 10.01 |
| Grade 4 (cohort %) | 0.064 | 20.03 | 0.022 | 0.15 |
| Grade 5 (cohort %) | 0.087 | 15.71 | -0.596 | 0.16 |
| Grade 6 (cohort %) | 0.031 | 43.75 | 0.018 | 3.38 |
| Grade 7 (cohort %) | 0.019 | 46.08 | -1.037 | 0.01 |
| Grade 8 (cohort %) | 0.205*** | 84.3 | -13.278 | 0.01 |
| Grade 9 (cohort %) | 0.189*** | 52.98 | 0.334** | 7.48 |
| Grade 10 (cohort %) | 0.063 | 18.15 | 0.188 | 0.07 |
| Grade 11 (cohort %) | -0.164* | 39.48 | -1.817 | 0.47 |
| Grade 12 (cohort %) | 0.101*** | 746.17 | 0.155*** | 115.43 |
| Post-school dipl. /cert. (cohort %) | 0.166*** | 283.71 | 0.232*** | 77.17 |
| Degree (cohort %) | 0.167*** | 70.68 | 0.218*** | 25.36 |
| P(allow overage), age = 16 | -0.81 | 0.46 | 0.858 | 0.82 |
| P(allow overage), age = 17 | 0.081** | 43.55 | 0.284 | 2.28 |
| P(allow overage), age = 18 | 0.056** | 132.18 | -0.158 | 7.87 |
| P(allow overage), age = 19 | 0.137*** | 186.13 | 0.205** | 9.31 |
| P(allow overage), age = 20 | 0.017 | 230.54 | -1.344 | 2.48 |
| P(allow overage), age = 21 | 0.075*** | 244.6 | 3.33 | 0.22 |
| P(allow overage), age = 22 | 0.144*** | 263.57 | 0.101 | 3.33 |
| P(allow overage), age = 23 | 0.154*** | 263.52 | 0.11* | 24.93 |
| P(allow overage), age = 24 | 0.28*** | 161.39 | 0.337*** | 43.13 |
| P(allow overage), age = 25 | 0.441*** | 105.37 | 0.588*** | 34.95 |
| P(allow overage), age = 26 | 0.86*** | 17.95 | 0.846*** | 11.81 |
| IV sets 1 | 0.105*** | 83.66 | 0.125*** | 24.92 |
| IV sets 2 | 0.137*** | 48.6 | 0.228*** | 9.31 |
| IV sets 1 & 2 | 0.115*** | 53.03 | 0.146*** | 16.31 |

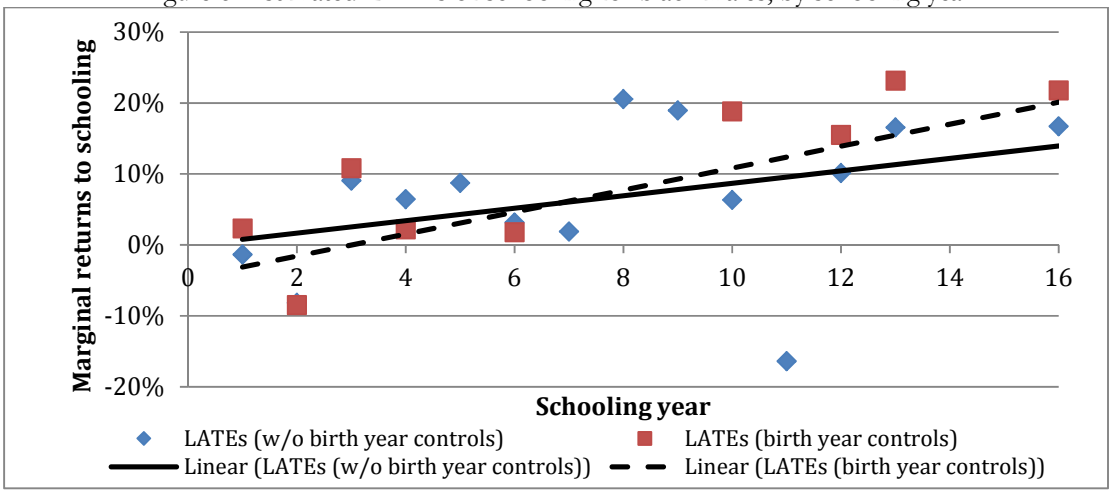
Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

IVs, enrolment restrictions at later ages had an impact at higher schooling levels, and iii) the effect of the second policy shift was mainly on later schooling years compared to the more evenly distributed effect of the first policy shift. Keeping this in mind, we can also distinguish three trends amongst the coefficients in Table 5: the estimated returns are i) an increasing function of the grade at which restrictions were relaxed, ii) an increasing function of the age of over-aged enrolment likelihood, and iii) higher on average for the

second policy than for the first. These two sets of trends indicate a positive correlation between the schooling year and the causal schooling effect of earnings along three dimensions of IV variation. The reported F-statistics in the second columns show that the estimates were generally based on a strong partial correlation between the instruments and schooling.

In a high unemployment country such as South Africa the selective process through which individuals enter formal sector employment may change what we understand as the LATE of schooling, or even bias the 2SLS estimates away from the true LATEs. Section 2.3.2.2 discussed the issues related to this problem and suggested that the most sensible way to avoid this type of bias is to control for any observable characteristics that may shift the employment probability of individuals while being correlated to the IVs. In this vein, column 3 of Table 5 re-estimates the schooling coefficients in column 1, but now uses a specification that controls for birth year (as a cubic polynomial), age (as a spline), sample year and province. The additional controls substantially reduce the explanatory power of the IVs, as demonstrated by the F-statistics reported in column 4. The LATE associated with the second policy's IVs are still higher than those associated with the first policy. Furthermore, the basic pattern in which the schooling coefficients generally increase with the cohort completion rates at higher grades and the enrolment restrictions at older ages is maintained, although these are less pronounced.

Figure 6: Estimated LATEs of schooling for black males, by schooling year




Source: Author's calculations from October Household Surveys 1995-99 and Labour Force Surveys 2000-2007 (StatsSA: various years).

Figure 6 plots the LATEs in columns 1 and 3 of Table 5 against the cohort completion rates at the different grades. The trend lines for these coefficients (which ignore those coefficients with a F-statistic below 1) clearly show that the local average of the marginal schooling returns increases as the cohort completion rates for higher grades are used. It should be noted that the LATE estimator also averages returns along dimensions other than schooling year, and may therefore fluctuate between the different IVs for reasons other than how the effect of each IV was distributed across schooling years. Still, the combined evidence from Table 2 and Table 5 provides strong *prima facie* evidence in support of a convex schooling earnings profile.

#### 2.5.2.2 Average schooling-earnings profile interpretation

The Becker-Card model assumes that the marginal return to schooling is a linear function of schooling but that the intercepts are allowed to vary across individuals. If this is indeed the case, then the schooling-earnings profile for the average individual in our sample can be constructed with the parameters from earnings equation [5]. Estimating the average of the heterogeneous schooling parameter as well as the homogeneous schooling squared parameter is analogous to the estimation of an ATE in a CRC model with exogenous conditioning variables, and section 2.3.3 discussed the conditions under which this can be achieved using the 2SLS or CF estimators. Table 6 below shows the coefficient estimates from an OLS earnings regression with various forms of the first-stage schooling residual included as additional controls. Columns 1 and 2 include no residuals, and the residuals from schooling and schooling squared respectively, and therefore correspond to the OLS and 2SLS estimates. The specification in column 3 adds the schooling residual interacted with schooling itself but omits the schooling squared residual, and hence represents the CF estimates.

Table 6: Earnings regressions comparison of OLS, 2SLS and CF estimators (sample of black males), 1995-2007

| | **OLS** | **2SLS** | **CF** |
|---|---|---|---|
| Education | -0.026*** | -0.038 | 0.073*** |
| | *(0.0024)* | *(0.0409)* | *(0.0100)* |
| Education^2 | 0.0089*** | 0.009*** | 0.003*** |
| | *(0.0002)* | *(0.0024)* | *(0.0005)* |
| Education residual | | 0.012 | -0.056*** |
| | | *(0.0406)* | *(0.0083)* |
| Education^2 residual | | -0.0001 | |
| | | *(0.0024)* | |
| Education residual*Educ | | | 0.007*** |
| | | | *(0.0005)* |
| Age | 0.095*** | 0.097*** | 0.088*** |
| | *(0.0018)* | *(0.0017)* | *(0.0016)* |
| Age^2 | -0.0009*** | -0.0009*** | -0.0008*** |
| | *(0.00002)* | *(0.00003)* | *(0.00003)* |
| Observations | 129659 | 129659 | 129659 |
| R-squared | 0.332 | 0.332 | 0.334 |
| Also control for: | Sample year, province | | |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

The OLS return estimates can be seen to imply that schooling returns are negative for the first year of education, and then increase by 1.8% for each additional year of schooling. The 2SLS estimates suggest slightly lower returns but a marginally stronger degree of convexity (and negative returns up to the second year of schooling) and would be considered more reliable in a homogenous parameter setting where it addresses the issue of endogenous schooling decisions that could bias the OLS estimates. In this case the similarity between these sets of coefficients would suggest that the exogenous schooling assumption is relatively innocuous. However, the much stronger assumptions required to justify the ATE interpretation in a heterogeneous coefficient model – including Card's concerns that instruments can affect the mapping from heterogeneous returns to schooling outcomes – should make us wary of uncritically assuming that our instruments have circumvented the more complicated type of endogeneity that arise in a CRC model. In such cases the CF estimator provides a useful alternative that will provide consistent estimates of the parameters of interest under different identifying assumptions.

This specification shows a much lower degree of convexity than suggested by either 2SLS or OLS. The wage benefit of the first schooling year is already positive and substantial, but this benefit only increases by 0.6% for each additional year of schooling. If the identifying assumptions of the CF model are more suitable to our particular schooling-earnings model, then this would suggest that previous South African earnings equation coefficient estimates discussed in section 2.2 may have over-stated the degree of convexity in the schooling-earnings profile by a factor of three. The schooling-schooling residual interaction term is significant in the earnings regression, contrary to what was found by Söderbom, et al. (2006). This suggests that the schooling error and relative return terms are positively correlated, and that heterogeneity of returns is important in understanding the South African earnings distribution. The schooling residual is significantly negative, which implies that the labour market productivity term is negatively correlated with the schooling error term.

Section 7.2.2 in the appendix re-estimates these regressions with different control variables (Table 18) and finds that the estimated schooling-earnings profiles in Table 6 are not particularly sensitive to the omission of birth year controls. However, Table 17 in the appendix estimates the CF model while allowing the schooling residual to enter the model as a low-order polynomial or after having been created through an ordered probit schooling process, and these results look markedly different from those in Table 6. In both cases the schooling-earnings profile is now concave, and in the ordered probit case the implied marginal returns are also markedly lower on average. On the one hand these results confirm the notion that the OLS/2SLS estimator may be overstating the extent of convexity in the schooling returns. On the other, it serves to emphasise how sensitive the results of the different estimators are to the specificities of our functional form assumptions, and particularly to the assumptions regarding how individuals make schooling decisions and how unobserved schooling and market ability interacts. These are issues that have not been thoroughly explored in any empirical studies that we are aware of, and are again touched upon in chapter 3 (although in a slightly different context).

In Table 7 we also explore the possibility that the presence of sample selection could be biasing our coefficient estimates by adding the IMRs calculated from the residuals in Table 3 to our CF estimates from Table 6. The social grant variables that are used in identifying exogenous variation in the employment outcomes are only available for certain sample years, so columns 1 and 2 in Table 7 start by comparing the uncorrected CF estimates for the full and restricted samples. Restricting the sample in this way clearly had little effect on our CF estimates of the schooling coefficients. Column 3 in Table 7 adds the inverse Mills ratio calculated from the regression in the first column of Table 3 (IMR1) to our CF estimator and therefore represents the sample selection corrected CF estimator under the assumption that schooling is exogenous in the employment process. The IMR1 term is significant in the earnings regression, and substantially reduces the degree of convexity in the schooling-earnings profile as well as the concavity in the age-earnings profile. Column 4 replaces the IMR1 term by the inverse Mills ratio derived from the regression in the second column of Table 3 (IMR2). The results show that treating schooling as endogenous in the employment process further reduces the degree of convexity in the schooling-earnings profile, but the additional effect is relatively small. Finally, the CF regression in column 5 uses an IMR calculated from the probit employment regression that also allowed the schooling coefficient to be heterogenous. However, as explained in section 2.3.3.2 this specification does not control for the expected value of the employment unobservables in an internally consistent manner, since the regression error term actually contains elements from two different unobservable components that cannot be separately identified. This problem notwithstanding, the results in Table 7 indicate that controlling for sample selection does not overturn the convexity in the schooling-earnings profile produced by the CF estimator. Barring the questionable results obtained from the inclusion of the IMR3 term, these regressions can also be interpreted as suggesting that not correcting for sample selection has the effect of over-estimating the rate at which the earnings benefit of schooling increases at higher schooling years.

Table 7: CF earnings regressions with different selection controls (sample of black males), 1995-2007

| | CF | CF | CF | CF | CF |
|---|---|---|---|---|---|
| Education | 0.073*** | 0.104*** | 0.106*** | 0.094*** | 0.076*** |
| | *(0.0103)* | *(0.0182)* | *(0.0131)* | *(0.0167)* | *(0.0174)* |
| Education^2 | 0.003*** | 0.0033*** | 0.0022*** | 0.0019*** | 0.0055*** |
| | *(0.0005)* | *(0.0007)* | *(0.0006)* | *(0.0006)* | *(0.0007)* |
| Education residual | -0.0563*** | -0.0802*** | -0.0746*** | -0.0881*** | -0.017 |
| | *(0.0077)* | *(0.0149)* | *(0.0113)* | *(0.0136)* | *(0.0171)* |
| Education residual*Educ | 0.0069*** | 0.005*** | 0.0052*** | 0.0069*** | -0.0002 |
| | *(0.0005)* | *(0.0008)* | *(0.0006)* | *(0.0006)* | *(0.0011)* |
| IMR1 | | | -0.296*** | | |
| | | | *(0.0234)* | | |
| IMR2 | | | | -0.182*** | |
| | | | | *(0.0230)* | |
| IMR3 | | | | | 0.043*** |
| | | | | | *(0.0065)* |
| Age | 0.088*** | 0.093*** | 0.041*** | 0.063*** | 0.105*** |
| | *(0.0022)* | *(0.0034)* | *(0.0052)* | *(0.0050)* | *(0.0033)* |
| Age^2 | -0.0008*** | -0.0008*** | -0.0002*** | -0.0005*** | -0.0009*** |
| | *(0.00004)* | *(0.00006)* | *(0.00007)* | *(0.00006)* | *(0.00005)* |
| Observations | 129659 | 54022 | 54022 | 54022 | 54022 |
| R-squared | 0.334 | 0.322 | 0.325 | 0.323 | 0.323 |
| Also control for: | Sample year, province | | | | |

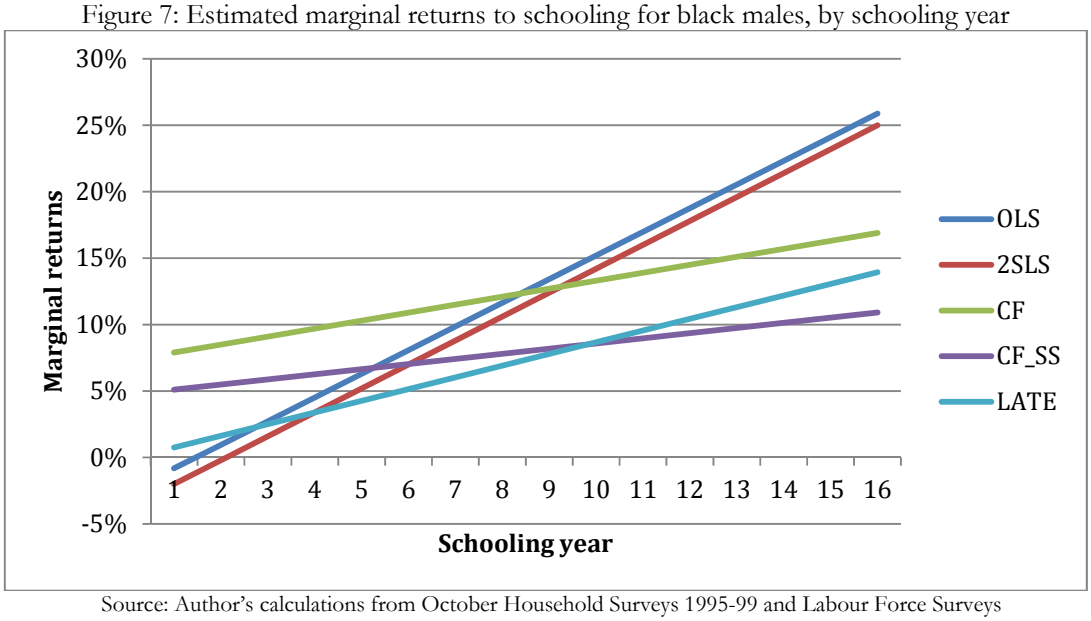
Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

The predicted marginal returns from the OLS, 2SLS and CF estimates in Table 6 and the sample selection corrected CF estimates from column 4 in Table 7 are plotted in Figure 7. We also add a linear trend line obtained from the LATE estimates associated with cohort grade completion rates as reported in column 1 of Table 5[22]. The first important result conveyed by this figure is that the marginal returns to investing in schooling for black South African men are high on average. At the mean schooling years of 8.15, the marginal returns vary between 7 and 12% depending on the estimator, which is high by international standards. Secondly, the marginal returns appear to increase with schooling years, which indicates a convex schooling-earnings profile. This results is relatively robust to the various specifications that allow for

[22] However, we should be careful in interpreting the trend line obtained from the LATE estimates as plotting a schooling-earnings profile. The LATEs are compared to the cohort share who completed the various schooling years – a measure of the restrictions on access to that schooling year – the effect of which is distributed across many different education levels.

different combinations of endogeneity, heterogeneity and sample selection[23]. Figure 7 also suggests that the OLS and 2SLS estimators may tend to overstate the convexity in the schooling-earnings profile. The various CF estimates suggest that early schooling years already had a positive effect on one's expected level of earnings, but that the benefit of more schooling did not increase quite as fast as previous studies may have led us to believe. Furthermore, not controlling for sample selection may further overstate the wage benefit of additional schooling.

Figure 7: Estimated marginal returns to schooling for black males, by schooling year




Source: Author's calculations from October Household Surveys 1995-99 and Labour Force Surveys (March & September rounds) 2000-2005 (StatsSA: various years).

### 2.5.3 Sensitivity analysis

The discussion of the instrumental variables in section 2.4.3 mentions a few potential concerns regarding their exogeneity. The first concern is that the birth cohorts differ for reasons other than their education levels, perhaps due to variations in the quality of schools over time. Although this is difficult to control for directly, we can attempt to proxy for such differences by including functions of the birth year variable (normalised to 0 at the oldest birth year of 1930). If schooling quality changes slowly over time then the inclusion of a birth year polynomial function in the earnings regression should be able to capture much of

[23] The one exception is the different specifications through which the schooling error term was allowed to enter the wage regression in Table 17, which produced concave schooling-earnings profiles.

the associated differences in worker productivity. Since the same variables are also included as conditioning variables in the first-stage regression, the exogenous variation in our instrumental variables is now also stripped of any trends in schooling outcomes that vary by birth cohort only. Table 19 in the appendix reports the results from including various birth cohort controls in a CF regression. The first column estimates the regression without any birth year controls, whereas the second, third and fourth columns include birth year as a polynomial of order one, three and five respectively. The fifth column includes a full set of birth year dummies. Controlling for birth year (linearly) has the effect of increasing the expected returns to schooling compared to no birth year controls, while decreasing the degree of convexity in the schooling-earnings profile. More flexible specifications tend to increase the degree of convexity slightly. These relatively minor differences notwithstanding, controlling for birth year does not substantially alter the results from the preceding section, which suggests that quality of schooling differences across generations did not bias our original CF estimates. Regardless of whether we control for birth year effects or not the expected schoolings-earnings profile is convex but closer to linear than the OLS or 2SLS estimators suggest.

A second concern is the possibility that our instrumental variables could be correlated to earnings through migration patterns. Since we observe only the province of residence rather than the province in which the individual attended school, any inclination amongst high ability individuals to migrate to certain provinces could induce a correlation between the provincial variation in our IVs and earnings that does not operate through schooling outcomes. Furthermore, if the same process of political liberalisation that led to the relaxing of restrictions on black education also allowed more freedom in the movement of migrant workers, then the birth year-province variation in our instrumental variables may actually serve as a proxy for inherent geographical and generational differences in productive abilities that came about through migration patterns. One way to test this hypothesis is to generate our instruments using only birth and sample year variation in our instrumental variables and to ignore any geographical differences in schooling attainment. The results from a CF estimator using these less informative instrumental variables are reported in column 2 of Table 20 in the appendix. The fact that the shape of the expected schooling-

earnings profile is very similar to that which is estimated using the IVs with provincial variation suggests that the IVs are not correlated to earnings because of migration patterns.

A third potential objection to the exogeneity of our IVs is that the relaxing of schooling restrictions coincided with a changing macro-economic environment – it is not necessarily for our purposes to speculate in which direction the causality ran – and that the macro-economic climate at the time of labour market entry has had a persistent effect on the expected earnings of individuals of different generations. We can test this hypothesis by including the average GDP growth rate and interest rate during the different points of an individual's school-going life. Since these variables also affect the direct and opportunity cost of schooling, it could be argued that these are natural conditioning variables in the schooling equation. The cohort unemployment rate is another factor that could affect wage-setting behaviour while being correlated to our instrumental variables (as discussed in section 2.4.3). Columns 3 and 4 in Table 20 add these variables to our regression and demonstrate that their inclusion has little effect on the estimated schooling-earnings profile. The argument that the IVs are correlated to earnings via its effect on schooling outcomes therefore withstands the objections raised in section 2.4.3.

## 2.6 Conclusions

This paper used three government policy shifts and various econometric techniques to estimate the causal effect of schooling on the earnings of black South African men in the period directly following political transition. Our model of earnings allows for endogenous schooling, individual heterogeneity in the returns to schooling, non-linearities in the schooling-earnings profile and a selective process through which individuals enter wage employment. This model is more general than has been estimated for South Africa (and any other country, to our knowledge) before. We take two different approaches to identifying the shape of the schooling-earnings function: comparing the LATEs of different IVs and estimating the parameters from a more restrictive Becker-Card model.

Our analysis in section 2.3 demonstrated that the usual LATE interpretation is not entirely accurate in the presence of sample selection: the IVs are required to be exogenous with respect to wage unobservables as well as excludable from the employment regression before producing valid LATEs. Even in this case, however, the subsample across which the marginal returns are averaged will only include those who were able to find work. Unfortunately, there does not seem to be any way to test for sample selection or control for its effects in a LATE framework, apart from ensuring that we control as comprehensively as possible for variables that are correlated to employment outcomes and our instrumental variables. The conditions under which the parameters of the Becker-Card model can be estimated using either a 2SLS or a CF estimator were discussed in section 2.3.3. We also briefly outline one way to address sample selection in a CF estimator, although this only works if the schooling returns coefficient is homogenous in the employment equation.

Our empirical analysis in section 2.5 demonstrated that schooling returns amongst black South African men are high on average, probably between 7% and 12% for those at the mean level of schooling attainment. Furthermore, the convexity of the schooling-earnings profile found in other studies is relatively robust to most of the estimators that we consider. There is some evidence that ordinary least squares and two-stage least squares tend to overstate the degree of convexity in the schooling-earnings profile, particularly when we allow for sample selection or more complex relationships between schooling and market unobservables. Finally, our results also show that addressing the endogeneity of schooling in an employment probit produces a schooling-employment profile which is slightly less convex than what is suggested under the exogenous schooling assumption. Our results also point towards a need for more empirical studies to help clarify the criteria under which the most appropriate IV techniques can be identified. Understanding how individuals make their schooling decisions and how unobserved abilities interact are crucial in making progress in this regard.

# 3. A discrete choice dynamic programming model of the South African schooling-earnings nexus


## Abstract

Many recent descriptive studies find convex schooling-earnings profiles in African countries. Apart from the potential endogeneity issues that may confound such estimates, it also is not clear whether a convex schooling-earnings profile can be reconciled with the fact that most individuals choose to obtain intermediate schooling outcomes. This chapter aims to contribute to the current debate regarding the shape of the schooling-earnings profile by using dynamic programming techniques to investigate the schooling decisions and labour market outcomes of South African workers. Our theoretical model builds on that of Belzil and Hansen (2002), but also allows high and persistent unemployment to play an important role in the choices made by students. We investigate the schooling decisions of the rational, dynamically optimising agents in our model using a novel decomposition technique to better understand what motivates the behaviour of South African students. Our results suggest that the schooling-earnings profile is convex, although the degree of convexity is less than suggested by previous studies that use OLS to estimate this relationship. The convexity of the schooling-employment profile is similarly reduced, and this change is much stronger than for the earnings profile. Furthermore, the marginal cost of schooling is required to be a steeply increasing function of schooling years if schooling decisions are to be reconciled with the convex earnings profile without resorting to assumptions of irrational expectations, imperfect information or schooling restrictions. Our decomposition indicates that the primary benefit of early schooling is the access that it provides to more advanced, higher yielding schooling years. The estimated pattern of unobserved abilities is more complex than the usual notion of high ability individuals who obtain more schooling, earn more, and are more likely to be employed, although this trend can be observed for the bulk of individuals who have neither very low nor very high academic abilities.

## 3.1 Introduction

Gary Becker (1964) introduced the idea of an internal rate of return to schooling that individuals use to guide their education investment decisions. This parameter is of central importance in understanding the earnings distribution, as well as the schooling investment decisions made by individuals. Becker and Chiswick (1966) demonstrated how this return can be estimated econometrically, and following the extension and popularisation provided by Mincer (1974), the schooling return soon became one of the most researched concepts in all of economics. However, the wage regression schooling coefficient that is frequently reported as an estimate of the schooling return is more accurately interpreted as the price of schooling from a hedonic market wage equation, a concept which is only loosely related to the parameter of interest. After almost half a century of empirical research Heckman, Lochner and Todd (2006, p. 311) therefore argue that the conventional econometric methods are fundamentally flawed and that the returns parameter remains "widely sought after and rarely obtained".

The problems with interpreting the earnings regression schooling coefficient as the benefit to an additional schooling year seem to be exacerbated by recent findings that the schooling-earnings profile has become increasingly convex in a number of countries. This finding holds true for numerous African countries (as discussed in Chapter 2), and also for South Africa, where the marginal returns have been found to increase by as much as 8% (Keswell & Poswell, 2004) for each additional schooling year. The finding that the marginal benefit of schooling investment is actually low for most schooling years and only starts to increase once students reach tertiary education seems contrary to the observation that most students choose to leave the school system during intermediate years rather than opting for corner solutions.

Attempts to address the empirical shortcomings of the Mincerian earnings regression have developed in two directions. Firstly, many studies argue that two-stage least squares with a valid instrument is able to estimate a weighted average of the returns for individuals who are heterogeneous along multiple dimensions without resorting to the strong behavioural assumptions that are implied by the Mincerian earnings regression. These studies only exploit the exogenous variation in schooling outcomes and can

therefore justify ignoring the nature of heterogeneity and individual schooling decisions. A second and very different approach uses dynamic programming methods to explicitly model the individuals' schooling investment decisions and labour market outcomes. These studies have often found schooling return estimates that are considerably lower than those estimated using OLS or IV techniques.

This chapter investigates the schooling-earnings nexus in South Africa, and particularly the insights afforded by using a dynamic programming model to study this relationship. It aims to contribute to the current debate on the shape of the schooling-earnings profile in African countries on three fronts. Firstly, this is (to our knowledge) the first study that uses dynamic programming techniques to study schooling decisions and labour market outcomes for an African country. Secondly, we adjust the assumptions of existing dynamic programming models to allow the unique features of the South African labour market to be factored into the decisions made by students. Finally, this chapter explicitly considers the implications of a convex earnings profile on optimal schooling outcomes, and investigates whether this can be reconciled with the observed schooling distribution within the context of a model with rational expectations and forward-looking, optimising agents. Individual schooling decisions are analysed using a novel decomposition technique which is developed in section 3.6.2.3. This method calculates the relative importance of different benefits of additional schooling, including concepts such as the "option value" and the "sequential revelation of information" that are emphasised in Heckman, Lochner, & Todd (2008).

## 3.2 Discrete choice dynamic programming studies of schooling returns

The rate of return is a vital parameter for those hoping to understand either the distribution of wages or the schooling investment decisions of individuals. Much of the early returns to schooling literature assumed that the schooling effect on log earnings was either a concave or approximately linear function of schooling year. This belief was primarily based on evidence from Mincerian earnings regressions (for early examples, see Mincer (1974) and Becker (1964)), but was also appealing from the perspective of explaining observed schooling investment decisions. In most countries only a small share of individuals choose

schooling corner solutions – obtaining no schooling at all or the maximum number of schooling years – and the prevalence of interior solutions is easily explained if the marginal cost of schooling is an increasing function of schooling year while a concave schooling-earnings profile is maintained. Assumptions of this nature form the basis for many prominent models of the schooling-earnings nexus (for example, Becker (1967), Card (1999)).

However, a number of descriptive studies have recently found that the schooling-earnings profile has turned convex, both in the US (Mincer (1996), Heckman, Lochner, & Todd (2008), Lemieux (2006)) as well as in a number of African countries (Appleton, Hoddinott, & MacKinnon, 1996; Teal, 2001; Whaba, 2000; Siphambe, 2000; Nielsen & Westergard-Nielsen, 2001; Carnoy, 1995), including South Africa (Keswell & Poswell, 2004). The reasons why such regressions cannot necessarily be interpreted as estimating the causal effect of schooling on earnings is well-documented (Denison, 1964; Spence, 1973; Griliches, 1977), and a popular strand of this literature has used supply side variation of the schooling decision as instrumental variables in order to estimate the unbiased effect of schooling on wages. IV studies generally find schooling returns that are somewhat higher than suggested by OLS (Ashenfelter et al. (1999)). Although such studies usually ignore any issues that relate to non-linearities in the schooling-earnings profile and effectively treats individual heterogeneity as a nuisance parameter, their results are often interpreted as evidence that ability bias is a relatively minor problem compared to measurement error. Chapter 2 discussed some alternative explanations for this phenomenon.

The last two decades have also seen the emergence of an alternative approach that addresses the same econometric problems in a very different way. Discrete choice dynamic programming (DCDP) models of schooling and earnings attempt to estimate policy invariant "deep parameters" (Heckman, 2010) that appear in individuals' utility and cost functions rather than some policy-dependent average of the marginal returns. Although this approach requires making strong assumptions about how the data were generated, these assumptions allow the econometrician to investigate the reasons why individuals behave in certain ways or to simulate how unobserved policies would affect schooling and earnings outcomes. Furthermore,

whereas most recent IV studies choose to interpret the 2SLS estimates as local average treatment effects that are not informative about the shape of the schooling-earnings profile, DCDP models usually estimate the exact shape of this profile. Apart from these conceptual differences, the two approaches can also produce very different results regarding the causal effect of schooling on earnings. Whereas it is not uncommon to find returns to schooling between 10% and 15% in IV studies (Card, 1999), DCDP estimates on US data are usually in the 4% to 7% range (Keane & Wolpin, 1997; Eckstein & Wolpin, 1989; Lee, 2005). Furthermore, Belzil and Hansen (2002) find that schooling returns in the US labour market are both low and convex, and that there exists a strong ability bias that would upwardly bias the OLS schooling coefficient. Given the importance of this paper, and the fact that our own theoretical model in section 3.5 uses their specification as a point of departure, it will be discussed in more depth in section 3.2.2. We start by comparing the advantages and shortcomings of DCDP models to that of IV methods in estimating the schooling returns.

### 3.2.1 The strengths and weaknesses of structural modeling

The last two decades have seen econometricians increasingly turn to flexible programme evaluation methods (as surveyed by Imbens and Wooldridge (2009)) that can identify some average of the causal treatment effects under fairly general conditions without requiring strong statistical assumptions or explicit behavioural assumptions. Imbens (2010) ascribes the emergence of the experimentalist approach to the "failure of specific structural models ... to satisfactorily address endogeneity issues in the context of estimation of causal effects of labor market programs". In the returns to schooling literature, such studies have often tried to identify supply-side determinants of the schooling decision such as institutional characteristics of the education system in order to obtain variation in schooling that is orthogonal to unobserved abilities. In this case the 2SLS estimate can most generally be interpreted as the LATE, or the weighted average of the marginal schooling returns for those who are induced to change their schooling decisions due to a change in the instrument values.

One of the most prominent differences between the structural and programme evaluation methods is the fact that fully specified structural models are more difficult to estimate. This is not a shortcoming in itself (although it may help explain the dominance of IV methods), but the simplicity of the programme evaluation methods "fosters transparency, reliability and sensitivity analyses" (Heckman, 2010, p. 17), all of which are considerably more difficult when using structural models. Belzil (2007) also points out that structural models usually assume that agents form rational expectations, the exact effect of which is difficult to quantify. Furthermore, although it is possible to allow for heterogeneous returns in a dynamic programming setting (Belzil & Hansen, 2007), most such models do not. Add to this the fact that structural methods will be sensitive to measurement error in the schooling variable (unlike IV methods) and this provides at least two biases that may taint the estimates of structural models but will not distort the results produced by programme evaluation methods. However, there is little reason to believe that measurement error bias will affect the schooling returns estimate by more than 10% (Card, 1999, p. 1834), and this would not explain why most structural estimates are lower than OLS estimates. One would also expect the correlation between heterogeneous schooling returns and schooling outcomes to produce upwardly biased returns estimates in structural models, which also does not account for actual low DCDP return estimates.

Regarding the widespread perception that structural methods require the econometrician to make stronger assumptions than do programme evaluation methods, Heckman (2010, p. 8) notes that even randomised experiments require making some assumptions about the agents' behaviour subject to randomisation. Furthermore, any experiment (either natural or randomised) takes place within a very specific economic environment and forecasting the impact of the same policy under different circumstances or a different policy, would require making additional assumptions about how people behave. More often than not economists are interested in the effect of a particular policy because it provides information about the types of interventions that may have desirable effects in the future. For example, we may believe that the effect of the construction of primary schools in Indonesia (Duflo, 2001) during the 1970s tells us something about what the effect would be of constructing secondary schools in Indonesia, or of building

primary schools in South Africa. However, unless the experiment is framed in terms of an economic model it is not clear which behavioural assumptions will allow us to extrapolate the results of one policy onto a different environment. More specifically, programme evaluation techniques generally do not require specifying the channels through which certain policies affect outcomes, but unless we know what these channels are there is no way to assess the likelihood of replicating a successful policy under different conditions. The distinction between structural and programme evaluation methods may therefore be more aptly described in terms of the types rather than the number or restrictiveness of assumptions they make, and how explicitly these assumptions are stated.

Belzil (2007) surveys the literature that uses structural models to estimate schooling returns and specifically investigates why structural and IV methods produce such divergent estimates. One of the main differences between the methods is that structural methods usually derive earnings and schooling equations with many endogenous explanatory variables and very general laws of motion. In contrast, programme evaluation methods require agents to behave in ways that will give rise to a linear schooling equation that can be neatly decomposed into exogenous and endogenous components and in which the "effect" of the instrument variable is monotonic, as well as an earnings equation in which schooling is the only endogenous variable. The fact that the econometrician can write down and estimate reduced form equations with these characteristics rather than having to derive the exact behavioural assumptions required to produce such equations is perhaps a weakness rather than a strength of programme evaluation techniques. On the other hand, structural methods are required to make strong statistical assumptions – for which there is usually little theoretical basis – about the distribution of the model unobservables whereas programme evaluation techniques typically only require assumptions about the first moments of the schooling and earnings equations. Belzil (2007) concludes that without more knowledge about the data generating process it is impossible to know whether the strong behavioural assumptions of the programme evaluation techniques or the strong statistical assumptions of structural models are preferable.

Heckman, Lochner & Todd (2006) point out that a completely specified structural model can always be extended to reflect additional features of the schooling-earnings relationship in a way that simple IV techniques cannot. For example, within a structural model it is possible to distinguish between the *ex ante* returns that individuals expected to earn when making schooling decisions and the *ex post* returns that the econometrician observes from the earnings data. Structural models have also been used to model the general equilibrium effect of certain schooling policies (Lee, 2005; Lee & Wolpin, 2006; Heckman, Lochner, & Taber, 1998). Although this flexibility is undoubtedly a useful trait of structural models in general, these extensions are not pursued in our own model, nor are they common amongst the studies surveyed in Belzil (2007).

One final distinction between the structural and programme evaluation approaches that partly motivates this study is the difference in how individual heterogeneity and schooling decisions are dealt with. The IV approach circumvents the problems posed by ability bias by only exploiting the exogenous variation in the schooling variable to estimate its effect on earnings. This allows the econometrician to ignore both the nature of individual heterogeneity – which is essentially treated as a nuisance parameter – and the way in which individuals make their schooling decisions. DCDP models of the schooling-earnings nexus explicitly model both the unobserved heterogeneity and the schooling decision. The parameter estimates of this completely specified structural model makes it possible to explore why individuals decide to continue or cease their studies, which is not possible with IV methods.

### 3.2.2 The Belzil and Hansen model

One of the most prominent examples of a DCDP model of the schooling-earnings profile, and the one on which the model in section 3.5 is based, is that of Belzil and Hansen (2002) (henceforth BH). In this model student $i$ derives instantaneous period $t$ utility from the amount of parental transfers received, $\xi_{it}$, according to the following utility function

$$u^s(.) = \ln \xi_{it} = \boldsymbol{z}_i \boldsymbol{\psi}_1 + \psi_2(s_{it}) + \eta^s_{g(i)} + \varepsilon^s_{it} \quad \text{[BH.1]}$$

where $\boldsymbol{z}_i$ is (time-invariant) family human capital, $\psi_2(.)$ is a spline function representing the marginal effect of schooling level $s_{it}$ on family transfers, $\varepsilon^s_{it} \sim nid(0, \sigma_s^2)$ is a utility shock and $\eta^s_{g(i)}$ is unobserved schooling ability that is specific to the individual's type $g$. The family human capital vector includes the father and mother's education levels, household income, the number of siblings, family composition at age 14, and regional controls. Schooling may be interrupted with exogenous probability $\zeta$[24], the outcome of which is captured by the interruption state variable $I_{it}$. Parental transfers cease once an individual decides to enter the labour market, after which utility is derived from labour market income according to

$$u^w(.) = \ln(W_{it}.m_{it})$$

where $W_{it}$ is wage earnings and $m_{it}$ is the individual's "employment rate". Log wage earnings can be represented as

$$\ln W_{it} = \chi_1(s_{it}) + \chi_3 x_{it} + \chi_4 x_{it}^2 + \eta^w_{g(i)} + \varepsilon^w_{it} \quad \text{[BH.2]}$$

where $\chi_1(.)$ is a spline function representing the causal effect of schooling on wages, $x_{it}$ is labour market experience, $\varepsilon^w_{it} \sim nid(0, \sigma_w^2)$ is a wage shock and $\eta^w_{g(i)}$ is persistent type-specific unobserved market ability. The annual employment rate is determined by schooling, experience, an employment shock $\varepsilon^m_{it} \sim nid(0, \sigma_m^2)$, but no fixed effect[25]

$$\ln\left(\ln \frac{1}{m_{it}}\right) = \kappa_0 + \kappa_1 s_{it} + \kappa_3 x_{it} + \kappa_4 x_{it}^2 + \varepsilon^m_{it} \quad \text{[BH.3]}$$

[24] BH initially express the interruption probability as a function of schooling year, $\zeta(s_{it})$, although the implementation of the model restricts this to be constant probability across individuals and schooling years, $\zeta$. This seems to remove much of the value that is added by modelling interruption in the first place.

[25] The employment equation (3) in the original article is specified as $\ln \frac{1}{m_{it}} = \kappa_0 + \kappa_1 s_{it} + \kappa_3 x_{it} + \kappa_4 x_{it}^2 + \varepsilon^m_{it}$, but this is inconsistent with their equation (7). BH also neglect to define $\mu_j$ (in their equation (7) and equation (BH.7) below), but it seems reasonable from the context to assume that $\mu_j = \kappa_0 + \kappa_1 s_{it} + \kappa_2 x_{it} + \kappa_3 x_{it}^2$. Working backwards from equation (7) gives the employment equation (BH.3) below. If this is indeed the correct employment equation, then the expected value of the employment rate varies between 92.1% and 99.3% and is an increasing function of schooling years, both of which is more sensible than what is obtained under the original equation. This error was confirmed by the authors (Belzil, personal communication, 8 Feb 2011).

The optimisation problem can be characterised in terms of a pair of functions that represent the maximised expected discounted stream of utilities from choosing to either leave school and enter the labour market (in which case $d_{it} = 0$) or from attending school for another year ($d_{it} = 1$). The value associated with entering the labour market is[26]

$$V_{it}^{w}(s_{it}, \boldsymbol{a}_{it}) = \ln(W_{it}.m_{it}) + \sum_{j=t+1}^{T} \beta^{j-(t+1)}\left[-\exp\left(\mu_j + \tfrac{1}{2}\sigma_m^2\right) + \chi_1(s_{it}) + \chi_3 x_{it} + \chi_4 x_{it}^2 + \eta_{g(i)}^{w}\right] \quad \text{[BH.7]}$$

where $\boldsymbol{a}_{it} = \left(\boldsymbol{z}_i, I_{it}, x_{it}, t, \varepsilon_{it}^{w}, \varepsilon_{it}^{m}, \varepsilon_{it}^{\xi}\right)$. The value attached to enrolling in school is[27]

$$V_{it}^{s} = \ln \xi_{it} + \beta\left\{\zeta E V_{i,t+1}^{I} + (1-\zeta) E \ \text{Max}\left[V_{i,t+1}^{s}, V_{i,t+1}^{w}\right]\right\} \quad \text{[BH.4]}$$

where $$V_{i,t+1}^{I} = \ln \xi_{i,t+1} + \beta\left\{\zeta E V_{i,t+2}^{I} + (1-\zeta) E \ \text{Max}\left[V_{i,t+2}^{s}, V_{i,t+2}^{w}\right]\right\} \quad \text{[BH.5]}$$

Using the finiteness of the optimisation timeframe and the orthogonality of random shocks, BH argue that it is now possible to calculate a closed form solution for this model. The authors then use dynamic programming methods to find the optimal schooling outcome for six types of individuals, each of which occurs with probability $p_g$ where $\sum_{g=1}^{6} p_g = 1$. Finally, the parameter values that maximise the finite mixture of likelihoods are reported as the structural estimates of the underlying model. As with any structural model, the BH model makes a number of seemingly strong behavioural and statistical assumptions in order to obtain a solvable model. Given that our model in section 3.5 will use the BH model as a point of departure, some of the paper's shortcomings will now be discussed in more depth.

Students are assumed to derive utility while in school based solely on the level of parental transfers that they receive. Given that parental transfers are unobserved, the parameters in the schooling equation (BH.1)

[26] The specification for $V_{it}^{w}(s_{it}, \boldsymbol{a}_{it})$ in the original article (equation (7)) omits $\eta_i^{w}$ as a determinant of the expected wage level. This suggests that despite observing their schooling ability (perhaps with noise as $\eta_i^{s} + \varepsilon_{it}^{s}$) and new wage offers for each year that they spend in school (the signal of which may again be hidden within $\eta_i^{w} + \varepsilon_{it}^{w}$), individuals are still unable to deduce anything about their own market ability. In fact, it assumes that all individuals will take as zero the expected value of $\eta_i^{w}$ despite the fact that their results show this term to be larger than 1 for individuals of all types. The authors (Belzil, personal communication, 8 Feb 2011) confirmed that this was an error in equation (BH.7) rather than actual modelling assumption.

[27] Since the arguments of the period $t$ value function are always $(s_{it}, \boldsymbol{a}_{it})$ and hence unambiguous, these are suppressed in this and all following equations.

can equally validly be interpreted as the effect of household characteristics and current schooling year on the marginal cost – either psychic or monetary – of an additional schooling year. The current schooling year enters this equation as a spline function $\psi_2(s_{it})$[28] with knots at most schooling years, which has the benefit of allowing for a very flexible relationship between school year and school cost. This specification provides the numerical optimisation procedure with many parameters with which to fit the proportions of people who choose to enter the labour market at various levels of schooling. However, the spline function coefficients are chosen with no regard for the institutional characteristics of the education system so that the added flexibility may actually detract from our understanding of what motivates schooling decisions. For example, the results in Table II (Belzil & Hansen, 2002, p. 2082) suggest that parental transfer are halved during the final year of high school and will jump by 30% in the final year of an undergraduate degree.

There are many reasons other than a sudden and unobserved increase in parental transfers that would make individuals more likely to complete the final year of their undergraduate education relative to the preceding or following schooling years (after controlling for differences in expected wage returns). Indeed, Belzil (2007, p. 51) explains that in such models "observed wage differentials across schooling levels are matched to actual choices by adjusting psychic costs estimates". These parameters are therefore perhaps more sensibly interpreted as any of the non-wage determinants of the schooling decision that varies systematically with schooling year, but the paper does not offer any guidance in how to interpret the large fluctuations in the school utility between successive years suggested by their results.

The vector of family human capital $\boldsymbol{z_i}$ is assumed to be constant despite the fact that it includes variables such as family income that clearly would have varied across periods. It is not clear how this was handled in the implementation of their model. A related but less important concern is the assumption that the logarithm of parental transfers is linear in observable household characteristics such as household income.

[28] The notation regarding this spline is a little confusing since the cost of obtaining the next year of schooling, $s_{it} + 1$, seems to be expressed in terms of the current schooling year, $s_{it}$. It is assumed that the results in Table II (Belzil & Hansen, 2002, p. 2082) ignore this discrepancy when reporting their results, so that (say) the education spline parameter $\delta_{12}$ refers to the change in the schooling utility associated with progressing from the 11th to the 12th year of schooling.

However, parents who always match an absolute increase in their household income with a constant *proportional* increase in parental transfers will soon run into budget problems. Furthermore, the descriptive statistics in Table I appear to have divided family income by 1000 for the calculation of the standard deviation but not for the mean.

The specification of the wage equation in (BH.2) is subject to some of the same objections raised about Mincerian earnings regressions (Heckman, Lochner, & Todd, 2008; Lemieux, 2006): for example, the effect of experience on log earnings is not necessarily quadratic or parallel across successive schooling levels. Perhaps more restrictively, the schooling returns are assumed to be homogenous across individuals. Although it is possible to allow for heterogeneous schooling returns in DCDP models of this type (Belzil & Hansen, 2007) most studies that use this approach assume parameter homogeneity. The most problematic aspect of their wage function is that the results in their Table IV suggest an experience-earnings profile which is substantially more concave than what is usually found in earnings regressions: agents will expect their labour market earnings to peak at 15 years of experience and to fall below their starting salary after accumulating 31 years of experience. This particular result is not discussed in the paper, but is difficult to reconcile with our understanding of on-the-job training models or the descriptive statistics presented in most Mincerian earnings regressions. If their particular schooling-earnings profile can only be squared with the data if combined with such an unusual experience-earnings profile, then this would seriously undermine the credibility of their results.

BH choose a much less flexible specification for their employment function – employment security is assumed to be linear in schooling and unaffected by unobserved labour market or schooling ability – which largely reflects the minor role that un- or underemployment plays amongst males in the US labour market. BH are also vague about how the individual's employment rate is operationalised, but it seems that the employment security rate is $0 < m_{it} < 1$, which would preclude using an employment dummy variable to identify the employment parameters. This specification ignores a potentially important channel

through which schooling and abilities may affect labour market outcomes in South Africa, which is why our own model in section 3.5 adjusts this assumption.

The specification of the $V_{it}^{I}$ function in BH.5 indicates that those whose schooling were interrupted in year $t$ would enter schooling again during year $t+1$, regardless of the draws of $(\varepsilon_{i,t+1}^{m}, \varepsilon_{i,t+1}^{w}, \varepsilon_{i,t+1}^{s})$. However, if an individual only entered schooling in year $t$ due to a very low draw of $(\varepsilon_{it}^{m}, \varepsilon_{it}^{w})$ and a high draw of $\varepsilon_{it}^{s}$, then we would not expect the individual to necessarily make the same decision again in the following year.

All of the BH model's agents are assumed to retire in period $T$, and utility derived after retirement is ignored (as was the case for the Becker-Card model in the previous chapter). The model also ignores the effect of measurement error, which could downwardly bias the average schooling returns. Finally, returns are assumed to be constant across individuals, although heterogeneity in the wage and schooling intercepts may be able to replicate some of the effects of parameter heterogeneity. Although each of these factors could conceivably be incorporated in a DCDP model of this type, it would substantially increase the computational complexity involved in solving the model. For this reason we follow BH in assuming away these issues in our own model

## 3.3 Schooling, earnings and employment in the South African labour market

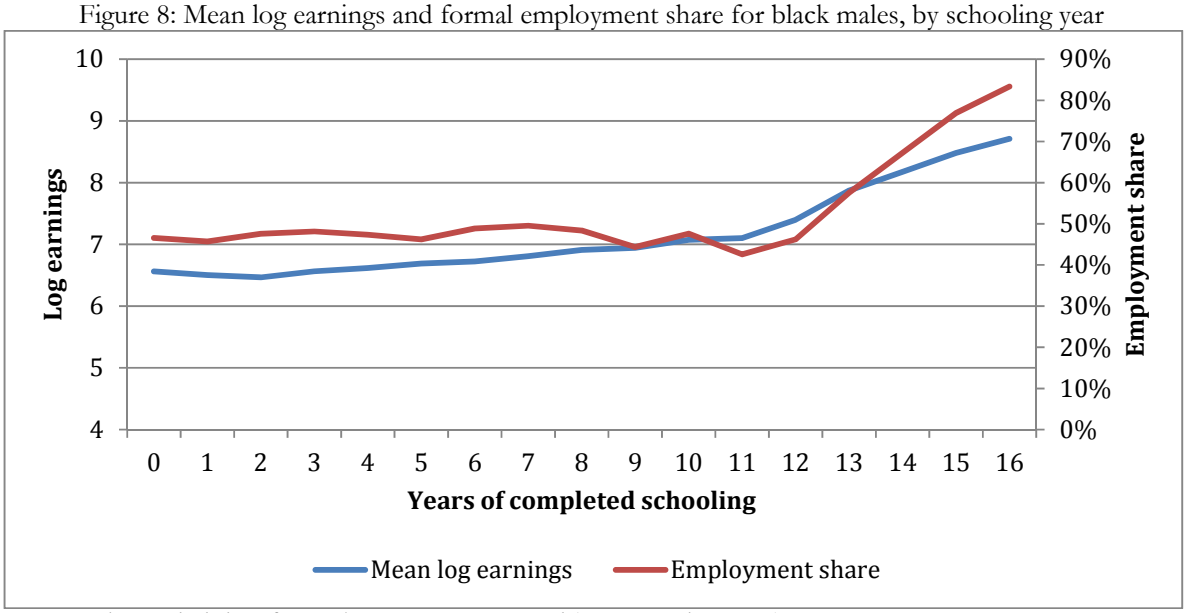
Before delving deeper into the nature of DCDP models we briefly present some descriptive statistics as further motivation for this chapter. Figure 8 plots mean log earnings and the employment share[29] for black males who have completed different schooling years. The South African school system consists of 7 years of primary education followed by 5 years of secondary school. The LFS panel data (discussed in section

[29] The employment share is here defined to be the share of working aged individuals employed in the formal sector. Our model in section 3.4.1 ignores issues of job search and labour market participation, so we do not distinguish between the unemployed and economically inactive individuals.

3.6.1) allow us to distinguish between post-graduate diplomas or certificates (the holders of which are assigned 13 years of completed schooling), an undergraduate university degree (15 years) and post-graduate degree (16 years).

Figure 8 reveals that the mean of log earnings decreases marginally over the first two years of primary school, and then starts increasing at a rate of around 7% for the remaining primary and early secondary school years. The final year of secondary schooling is associated with a 30% increase in mean earnings and post-secondary schooling with an even higher rate of 32%. Given the observed schooling-earnings pattern, it is no surprise that schooling enters the Mincerian earnings as a convex function and that studies that include schooling as a quadratic function suggest that the marginal rate of return increases by between 0.8% (Kingdon & Knight, How Flexible Are Wages in Response to Local Unemployment in South Africa, 2006) and 8% (Keswell & Poswell, 2004) per schooling year. The schooling-employment share profile is similarly convex, although there is no apparent benefit to more schooling until individuals start accumulating post-secondary schooling.

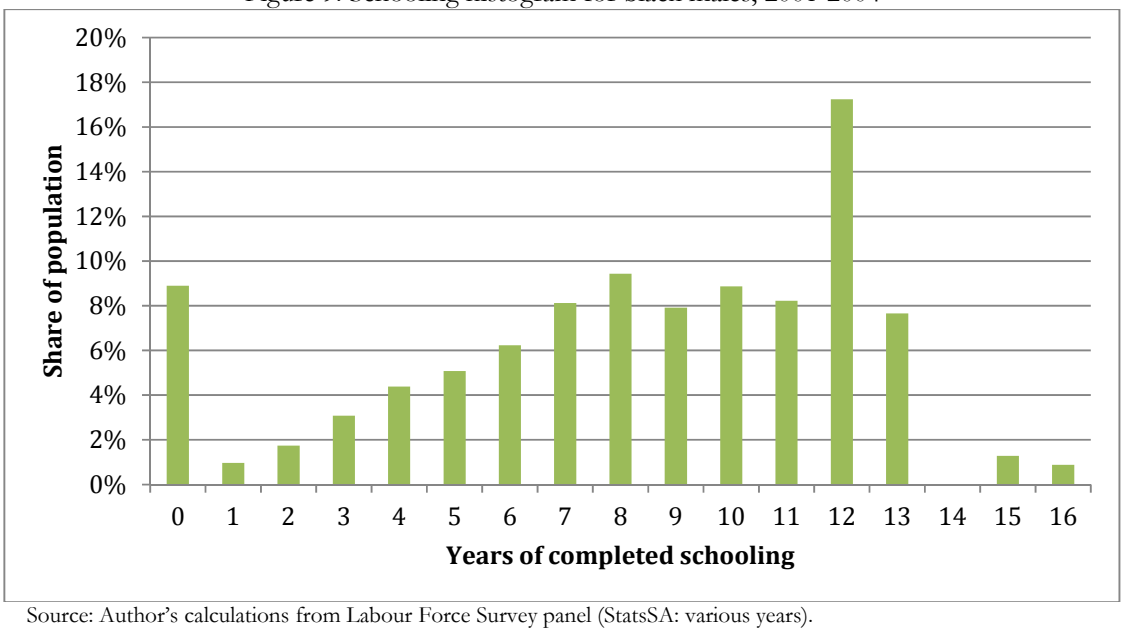
Figure 8: Mean log earnings and formal employment share for black males, by schooling year




Source: Author's calculations from Labour Force Survey panel (StatsSA: various years).

The schooling coefficients from an earnings function are also important for those interested in understanding individual schooling decisions, as schooling's effect on expected labour market outcomes is usually assumed to be the primary benefit of investing in additional schooling. Any model that claims to capture the relationship between schooling and earnings should be able to explain individual schooling decisions in light of the perceived schooling benefits. Figure 9 presents a histogram of the share of black men who left the school system with the different number of completed schooling years. Apart from the 9% of individuals who chose to complete no schooling at all and the 1% who opted to obtain post-graduate qualifications, most individuals chose to exit school during the intermediate schooling years. Very few individuals left school during early primary school, but we observe a stronger outflow by late primary and early secondary school. The most popular point to quit school was after completing secondary education, which almost 20% of black men did. Only about 11% decided to proceed beyond secondary school, most of whom completed diplomas or certificates rather than university degrees. However, it is not clear how these schooling decisions can be reconciled with the mean earnings and employment probability patterns shown in Figure 8. If primary and secondary school added much less to an individual's lifetime earnings capacity than tertiary education, then it is peculiar that so many seemingly rational individuals would choose to complete all the low yielding schooling years and then enter the labour market just as the benefit of additional schooling years starts to take off. Section 3.6.2.2 investigates the extent to which these two outcomes can be explained by a single schooling-earnings model.

Figure 9: Schooling histogram for black males, 2001-2004

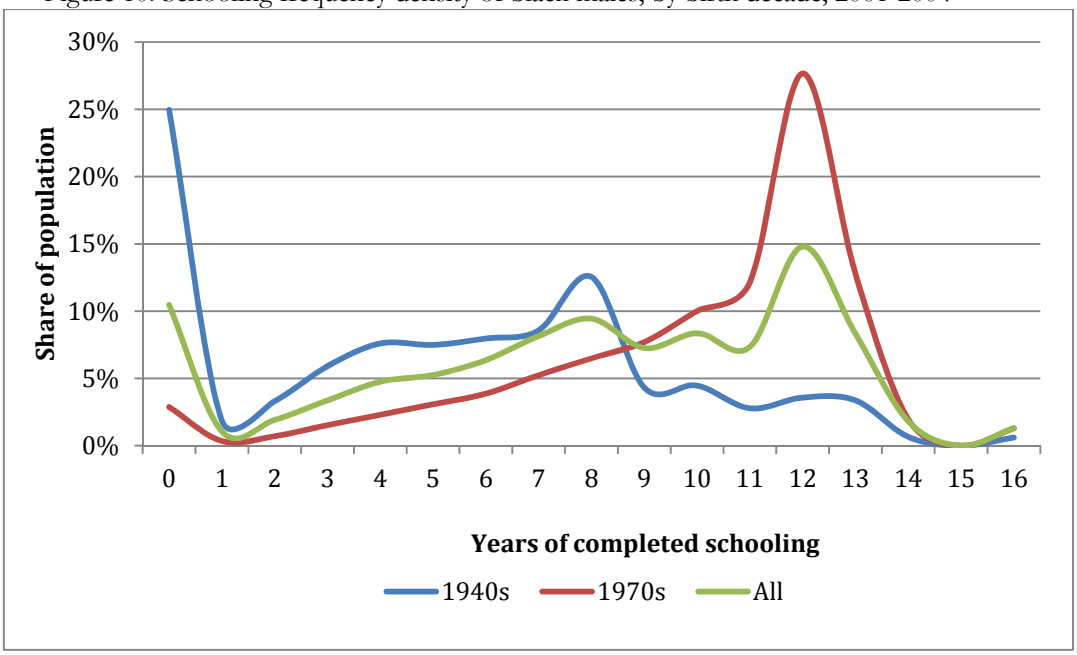



Source: Author's calculations from Labour Force Survey panel (StatsSA: various years).

Figure 10 demonstrates that there is substantial variation in the schooling outcomes of individuals across different generations. The histogram for the total population is reproduced from Figure 9 and juxtaposed to the probability density functions for those born in the 1940s and the 1970s. Starting with the older birth cohort, it can be observed that a much higher share of those born in the 1940s opted to get no schooling at all, and that the shares that completed secondary or tertiary education were lower than for the population as a whole. Amongst those born in the 1970s, obtaining either no or primary schooling, was much less likely whereas leaving school with incomplete or (particularly) completed secondary schooling or tertiary education was more common.

There are a few stylised facts of the schooling distribution that we would like our model to explain. Firstly, people generally choose to obtain intermediate schooling years rather than corner solutions. Secondly, there is bunching at completed secondary schooling, particularly for those born more recently. Thirdly, a substantial share of individuals from older generations decided to obtain no schooling at all, whereas for those from younger cohorts this is was much less common as more individuals proceeded to incomplete or completed secondary school. Finally, some progress was made in increasing tertiary schooling amongst younger generations, but mainly in terms of one year diplomas rather than university degrees.

Figure 10: Schooling frequency density of black males, by birth decade, 2001-2004




Source: Author's calculations from Labour Force Survey panel (StatsSA: various years).

## 3.4 Estimating discrete choice dynamic programming models

### 3.4.1 Structure of discrete choice dynamic programming models

Although DCDP models are generally more complicated to estimate than their programme evaluation counterparts, continued advances in dynamic programming tools and computing power means that such models can increasingly be considered as a feasible alternative to IV techniques when modelling the effect of endogenous variables in the presence of unobserved heterogeneity. We will briefly consider how such models are usually solved and estimated[30], before turning to our own model in section 3.5. Our focus here is restricted to partial equilibrium models without any strategic interactions between agents and which takes place within a finite planning horizon.

[30] For a more comprehensive survey of DCDP methods, see Aguirregabiria & Mira (2010).

Individuals are indexed by $i = 1, \dots, N$ and time by $t = 1, \dots, T$. The economic environment is described by a vector of state variables, $\boldsymbol{a}_{it}$, and choice variable $d_{it} = 0, \dots, J$ (with typical element $j$).[31] Some state variables are observed by the econometrician, $\boldsymbol{q}_{it}$, whereas others are only observable by the agent, $\boldsymbol{\varepsilon}_{it}$, where $\boldsymbol{a}_{it} = (\boldsymbol{q}_{it}, \boldsymbol{\varepsilon}_{it})$. Agents make decisions based on their desire to optimise the discounted stream of period-specific utilities $\sum_{\tau=t}^{T} \beta^{\tau-t} U(\boldsymbol{a}_{i\tau}, d_{i\tau})$, where $\beta \in (0,1)$ is the discount factor and $U(.)$ is a real-valued function that translates the state and decision variables into instantaneous utility levels. In the absence of uncertainty agents would choose the values of $d_{i,t+\tau}$ that maximised utility, but agents typically face some uncertainty regarding future realisation of the state variables. The evolution of these variables is described by the Markov transition distribution function $F(\boldsymbol{a}_{i,t+1}|\boldsymbol{a}_{it}, d_{it})$ which is common knowledge under the rational expectations assumption. Payoff variables $\boldsymbol{y}_{it} = \boldsymbol{Y}(\boldsymbol{a}_{it}, d_{it})$, such as production or earnings functions, contain information about the utility of agents without directly affecting the decisions of individuals. These variables are often not necessary to solve the model, but their presence can increase the efficiency of the estimates.

The typically complicated solution to the agent's optimisation problem means that individual choices cannot generally be related to observable characteristics in a simple and estimable reduced form equation. Instead, estimation of the model parameters requires first solving the agent's dynamic optimising problem for a given set of parameter values $\boldsymbol{\theta}$ before iterating through different values of this coefficient vector to find the values that are most likely to have generated the observed data. These two steps will now be discussed in turn.

[31] This variable is defined as a scalar. If the agent makes more than one decision per period, then the choice variable is redefined so that its different values correspond to various combinations of each decision.

### 3.4.2 Solving the agent's dynamic optimisation problem

The dynamic programming problem requires solving the value function $V(\boldsymbol{a}_{it})$ and the optimal decision rule $\alpha(\boldsymbol{a}_{it})$ for different values of $\boldsymbol{a}_{it}$. These terms are defined by Bellman's principle of optimality as

$$V(\boldsymbol{a}_{it}) = \max_{\boldsymbol{d}_{it}}\{U(\boldsymbol{a}_{it}, d_{it}) + \beta \int V(\boldsymbol{a}_{i,t+1}) dF(\boldsymbol{a}_{i,t+1}|\boldsymbol{a}_{it}, d_{it})\} \quad [10]$$

where $t = 1, \dots, T$ and $T$ is identical across individuals and

$$\alpha(\boldsymbol{a}_{it}) = \arg\max_{\boldsymbol{d}_{it}}\{U(\boldsymbol{a}_{it}, d_{it}) + \beta \int V(\boldsymbol{a}_{i,t+1}) dF(\boldsymbol{a}_{i,t+1}|\boldsymbol{a}_{it}, d_{it})\}$$

Once the model has been solved, it is possible to calculate the conditional choice probabilities as

$$P(d_{it} = j|\boldsymbol{q}_{it}) = \int_{\boldsymbol{\varepsilon}_{it}} 1(\alpha(\boldsymbol{a}_{it}) = j) dG_{\varepsilon}(\boldsymbol{\varepsilon}_{it}) \quad [11]$$

where $dG_{\varepsilon}(\boldsymbol{\varepsilon}_{it})$ is the joint distribution of the unobserved state variables, $1(.)$ is the indicator function and $t = 1, \dots, T$. However, this requires finding a way to solve the *Emax* integral $\int V(\boldsymbol{a}_{i,t+1}) dF(\boldsymbol{a}_{i,t+1}|\boldsymbol{a}_{it}, d_{it})$, which represents the expected value of lifetime utility associated with $(\boldsymbol{a}_{it}, d_{it})$. Solving this formula can require multidimensional integration at every point in the (potentially continuous) state space, and is particularly complicated if the model contains a large number of unobservables – usually a by-product of having many alternatives to choose from or many payoff variables – or if these variables do not enter the integral as normally distributed variables. Since this equation is often computationally intractable, much of the DCDP literature has focused on finding ways to simplify this step of the solution procedure.

At least three methods have been proposed to address this issue within the context of single agent models: (i) make functional form and distributional assumptions for which solutions [10] and [11] do not require integration (Rust, 1987), (ii) use techniques that describe the value function without solving the model explicitly (Hotz & Miller, 1993; Manski, 1988), or (iii) approximate the solution of the multidimensional integral via simulation and interpolation methods (Keane & Wolpin, 1994).

Many of the early DCDP studies were forced to choose their specifications so that their models could be solved using method (i). If the model assumptions correspond to those of the "Rust model", then the value function and conditional choice probabilities have very simple closed-form solutions that do not require solving the *Emax* function. However, the assumptions of the Rust model may be too heroic to use in many applications. It requires, for example, that the unobservable determinants of both utility and the payoff variables should be i.i.d processes. This is problematic in applications such as ours, where a central role is assigned to persistent personal-specific unobservable abilities that affect both schooling and wage outcomes.

Recent advances in computing power and the development of techniques associated with methods (ii) and (iii) are increasingly allowing researchers to choose assumptions based on their appropriateness in the context of the research question rather than considerations of computational complexity. If the Rust model assumptions are inappropriate for a certain research question, then one can turn to a more complex class of model known as "Eckstein-Keane-Wolpin" models. In such cases Keane and Wolpin (1994) suggest using Monte Carlo integration to approximate the value of the *Emax* function at the different points of the state space rather than attempting to solve it analytically. If the state space includes one or more continuous variables then Keane and Wolpin (1994) recommend discretising the state space, approximating the integral using Monte Carlo integration at each of these discrete points and then interpolating between these point to find the value function at the remaining state space values. This approximation will be more accurate the larger the number of simulations at each point of the state space and the larger the number of discrete points on the state space at which the integral is approximated.

### 3.4.3 Estimating the structural parameters

The second step of the dynamic programming problem requires finding the set of parameters for which the joint distribution of the conditional choice probabilities, the payoff variables and the state variables most closely resemble the joint distribution of observed choices and other variables in our data. Estimating the structural parameters therefore requires solving the whole model at each trial vector of coefficient

estimates, and then calculating a metric that evaluates the distance between predicted and observed outcomes – usually a likelihood or simulated moment function. A numerical optimisation routine would therefore usually be implemented via an inner (solution) and outer (estimation) loop. The inner algorithm solves the individual's maximisation problem for a certain set of parameter values $\boldsymbol{\theta}_0$, calculates the conditional choice probabilities for different combinations of state variables and evaluates the likelihood function. The outer algorithm then compares the values of the likelihood function at various values of $\boldsymbol{\theta}$ – each of which requires solving the agent's optimisation problem anew with the inside algorithm – and choosing the parameter values that minimise the chosen objective function. The dataset consists of all the observable state, payoff and choice variables $(\boldsymbol{q}_{it}, d_{it}, \boldsymbol{y}_{it})$ and is used to calculate the log-likelihood according to

$$\log L = \sum_{i=1}^{N} \sum_{t=1}^{T} \log L_{it}\,(\boldsymbol{\theta}) = \sum_{i=1}^{N} \log \Pr\{\boldsymbol{\alpha}(\boldsymbol{q}_{it}, \boldsymbol{\theta}) = d_{it}, \boldsymbol{Y}(d_{it}, \boldsymbol{q}_{it}, \boldsymbol{\theta}) = \boldsymbol{y}_{it}, \boldsymbol{q}_{it} | \boldsymbol{\theta}\}$$

Heterogeneity is most commonly incorporated into such models in the spirit of Heckman and Singer (1984) by allowing individuals to belong to one of $G$ different types (indexed with $g$). Each type faces a different set of parameter values $\boldsymbol{\theta}_g$, so that

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \log \sum_{g=1}^{G} p_g L_{it}(\boldsymbol{\theta}_g)$$

where $p_g$ is the share of the population that belongs to type $g$. The unobservable variables in the payoff and utility functions can now be allowed to contain a type-specific fixed effect as well as an i.i.d. error term: $\boldsymbol{\eta}_{g(i)} + \boldsymbol{\varepsilon}_{it}$. Although the heterogeneous components are not individual specific, the unobservables are allowed to covary in a completely unrestricted way across decisions and with the state variables. This allows for a fairly general class of sample selection and endogeneity in the data generating process.

Econometricians can choose from a number of different numerical optimisation techniques to find the parameter values that maximise the likelihood function, but in reality some procedures work better than others. Simplex downhill search algorithms may take longer to converge, and possibly settle at local rather

than global optima if the likelihood surface is not smooth. If the value function integrals cannot be solved analytically then it is unlikely that closed-form solutions for the gradient or Hessian matrices of the likelihood function can be calculated. In such cases the process of simulated annealing (Corona, Marchesi, Martini, & Ridella, 1987) has many desirable properties. This method starts at an initial parameter vector $\boldsymbol{\theta}$ and then changes element $k$ by a random amount $\theta'_k = \theta_k + r.v_k$ where $r$ is randomly drawn from the $[-1,1]$ interval and $v_k$ is the step length for element $k$ of the parameter vector. If the new parameter vector $\boldsymbol{\theta'}$ yields a higher log likelihood function $\log L(\boldsymbol{\theta'}) > \log L(\boldsymbol{\theta})$, then the change is accepted and the parameter vector updated. However, whereas most numerical optimisers always reject changes in the parameters that result in a lower likelihood, the method of simulated annealing will sometimes accept such changes in order to avoid converging on a local maximum. Specifically, a decrease in the log likelihood function will be accepted if $e^{(\log L(\boldsymbol{\theta'}) - \log L(\boldsymbol{\theta}))/T_0}$ is larger than a uniformly distributed random number drawn from the unit interval. Parameter $T_0$ refers to the "temperature" [32] of the process and can be set to manage the frequency with which downward moves are accepted. The process steps through all the parameter values $N_s$ times, after which the elements of the step length vector $v_k$ are adjusted so as to ensure that half of all parameter trials will be accepted. This process is itself repeated $N_T$ times before the temperature is reduced to $T'_0 = r_T.T_0$, and terminated once the convergence criterion is reached. Although this way of iterating through parameter values requires more computing power than standard numerical optimisation techniques, it has the benefit of being able to avoid getting stuck in local maxima and will be more likely to find the global optimum of functions that are poorly behaved.

In choosing parameter values it is important to opt for a vector of step lengths $v_k$ that will cover any sensible set of coefficient values. Corona, Marchesi, Martini, & Ridella (1987) propose running this algorithm with $r_T = 0.85$, $N_s = 20$ and $N_T = \max(100,5K)$ where $K$ is the number of model parameters. However, Goffe, Ferrier & Rogers (1994) investigate the performance of simulated annealing

[32] The terminology of the simulated annealing process betrays its roots in thermodynamics, where it was conceptualised as a means of cooling down substances in a way that would reduce defects caused by the internal energy of some atoms being stuck at higher than desired levels.

against three alternative numerical optimisation algorithms and find that although it usually takes longer to execute it is also more likely to find the global optimum of complicated functions. Furthermore, their simulations show that the parameter values suggested by Corona, Marchesi, Martini, & Ridella (1987) are often unnecessarily conservative and that execution time can be substantially reduced without any cost in search efficiency. In particular they suggest experimenting with lower values of $r_T$ and $N_T$ in order to reduce search time.

## 3.5 A discrete choice dynamic programming model of the South African schooling-earnings nexus

### 3.5.1 Model assumptions

We now formulate a theoretical model of the schooling decisions and labour market outcomes of South African workers. This model is primarily based on the BH model, although some assumptions are altered in order to avoid some of the shortcomings discussed in section 3.2.2 and to reflect specific features of the South African labour market (as discussed in section 3.3). Utility derived from being in school is expressed as

$$u^s(.) = z_i\psi_1 + \psi_{21}1(7 < s_{it} + 1 \leq 12) + \psi_{22}1(12 < s_{it} + 1) + \eta^s_{g(i)} + \varepsilon^s_{it} \qquad [12]$$

Apart from the difference in interpretation – our school utility equation is stated directly in terms of individual and household characteristics rather than assuming that these factors operate via parental transfers – this equation closely resembles the school utility BH model. One important difference is that the BH spline function $\psi_2(.)$ is replaced by a piecewise constant (or stair) function with discontinuous jumps between primary, secondary and tertiary education. The $\psi_{21}$ and $\psi_{22}$ coefficients represents the marginal utility that individuals expect to experience while in secondary and tertiary education respectively, relative to being in primary schooling. These changes primarily reflect the differential schooling costs

associated with shifting between these different schooling types. This assumption reflects our belief that the coefficients in the schooling equation are most sensibly interpreted as affecting the marginal costs associated with schooling. In estimating our model we use birth year as the school utility shift variable, $z_i$. Figure 10 demonstrated the importance of birth year in determining schooling outcomes, and this is to be expected given South Africa's history of apartheid-era restrictions on black schooling enrolments. Section 2.4.1 in chapter 2 argued that the gradual relaxing of these policies towards the end of apartheid created exogenous variation in the schooling outcomes amongst black men born in different years. Furthermore, our panel dataset (discussed in section 3.6.1) does not contain any of the family human capital variables used in BH, which restricts the range of candidate variables that can be included as schooling shifters.

Labour market participants derive utility according to

$$u^w(.) = m_{it} \ln W_{it} \quad [13]$$

where $m_{it}$ is an employment status dummy indicating whether the individual was employed $(m_{it} = 1)$ or not $(m_{it} = 0)$ and $W_{it}$ is labour market earnings. This process differs from that in the BH model, where individuals were assumed to vary in terms of their employment security but workers were never completely unemployed. The high and persistent levels of unemployment in South Africa would make such an assumption completely inappropriate. Unfortunately this change comes at the cost of considerably complicating the solution and estimation steps of our model. The process that generates log earnings is assumed to be the same as in equation BH.2

$$\ln W_{it} = w_{it} = \chi_1 s_{it} + \chi_2 s_{it}^2 + \chi_3 x_{it} + \chi_4 x_{it}^2 + \eta_{g(i)}^w + \varepsilon_{it}^w \quad [14]$$

Employment is determined as a binary variable according to the following employment equation

$$m_{it} = 1\left(\kappa_1 s_{it} + \kappa_2 s_{it}^2 + \kappa_3 x_{it} + \kappa_4 x_{it}^2 + \eta_{g(i)}^m + \varepsilon_{it}^m > 0\right) \quad [15]$$

The model error terms are each i.i.d normally distributed: $\varepsilon_{it}^s \sim n(0, \sigma_s^2)$, $\varepsilon_{it}^w \sim n(0, \sigma_w^2)$ and $\varepsilon_{it}^m \sim n(0,1)$. This means that the model is completely specified up to a parameter vector

$\boldsymbol{\theta} = (\boldsymbol{\chi}, \sigma_w^2, \boldsymbol{\kappa}, \boldsymbol{\psi}, \sigma_s^2, \boldsymbol{\eta}^w, \boldsymbol{\eta}^m, \boldsymbol{\eta}^s)$. The transition of the schooling state variable is determined according to $s_{it} = s_{i,t-1} + d_{i,t-1}$, but restricted to lie between 0 and 16 (as per the discussion of the schooling data in section 3.3), and the experience variable according to $x_{it} = x_{i,t-1} + (1 - d_{i,t-1})$ and constrained by the restriction that $0 \leq x_{it} + s_{it} \leq T$, where $T = 59$ is the difference between the retirement and minimum school-going age.

Although this model incorporates many features of the South African labour market, there are also a number of aspects that are omitted from our analysis. Firstly, we ignore issues of job search and labour force participation, which implies that all job offers are accepted and that the employment equation [14] is a representation of how firms allocate job offers. Secondly, our model also discards the grade repetition feature of the original BH model. Although South African schools have notoriously high repetition rates, the more relevant question is whether taking this into consideration helps us understand the relationship between schooling and earnings. The BH model does not allow the schooling interruption probability to vary by schooling year or unobserved ability, which means that the possibility of grade repetition essentially acts as a fixed cost to entering any level of schooling. It seems unlikely that this will add much to the model's ability to explain how schooling and earnings interact. A third concern is that schooling in our model is assumed to be a simple ordered variable, whereas individuals can actually choose between two different streams of post-secondary education – non-university or university – that will endow students with either 13 or 15 years of completed schooling. Perhaps the most serious shortcoming of our model is that it assumes stability of economic environment: that the schooling decisions of older generations were made with similar perceived marginal returns to schooling than what is observed in the current labour market. Unfortunately, given the lack of comparable household survey data going back that far, there is no way to test whether this was indeed the case or not. Although we therefore cannot rule out that the convex earnings and employment profiles observed in Figure 8 are very recent phenomena, it is worth noting that the schooling decisions of older cohorts may be even more difficult to reconcile with a concave schooling-earnings profile that implies high returns to early schooling years.

A final and slightly more general objection to the modelling framework is that it supposes that all individuals are dynamically optimising agents with rational expectations whose behaviour is unconstrained by credit market imperfections or any other constraints. There are almost certainly a wide range of alternative behavioural assumptions that could be consistent with the observed outcomes. In fact, it is always possible to explain away any inconsistencies between the investment returns and investment decisions by allowing for incomplete information, irrational expectations or constraints on investment behaviour. Although we do not wish to discard the possibility that the schooling decisions of black males were affected by non-cost constraints not included in our model – such as credit constraints or the limited number of places at tertiary institutions – this paper investigates the extent to which the schooling returns can be reconciled with the schooling decisions of unconstrained dynamically optimising agents with rational expectations.

### 3.5.2 Solving the agent's dynamic optimisation problem

The choice-specific utility functions [12] and [13] can be rewritten as a more general utility function which depends on the individual's schooling investment decision $d_{it}$:

$$U_{it}(\boldsymbol{a}_{it}, d_{it}) = (1 - d_{it})[1(\boldsymbol{q}_{it}\boldsymbol{\kappa} + \eta^m + \varepsilon_{it}^m > 0)(\boldsymbol{q}_{it}\boldsymbol{\chi} + \eta^w + \varepsilon_{it}^w)] + d_{it}[\boldsymbol{q}_{it}\boldsymbol{\psi} + \eta^s + \varepsilon_{it}^s]$$

where $\boldsymbol{q}_{it} = \left(s_{it}, s_{it}^2, x_{it}, x_{it}^2, z_i, 1(7 < s_{it} + 1 \leq 12), 1(12 < s_{it} + 1)\right)$, $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \kappa_3, \kappa_4, 0,0,0)$, $\boldsymbol{\chi} = (\chi_1, \chi_2, \chi_3, \chi_4, 0,0,0)$ and $\boldsymbol{\psi} = (0,0,0,0, \psi_1, \psi_{21}, \psi_{22})$. The individual's optimisation problem requires choosing $d_{it}$ in order to maximise the expected net present value of lifetime utility $E[\sum_{\tau=t}^{T} \beta^{\tau-t} U(\boldsymbol{a}_{i\tau}, d_{i\tau}) \,|\boldsymbol{a}_{it}, d_{it}]$. Under these assumptions the value function associated with state variables $\boldsymbol{a}_{it}$ simplifies to $V_{it}(\boldsymbol{a}_{it}) = \max_{d_{it}}\{u_{it}(\boldsymbol{q}_{it}, d_{it}) + e_{it}^{d_{it}} + \beta \int V_{it}(\boldsymbol{q}_{i,t+1}) dF(e_{i,t+1}^{d_{it}})\}$, where $e_{it}^1 = \varepsilon_{it}^s$ and $e_{it}^0 = e_{it}^w = \{m_{it} - \Phi(\boldsymbol{q}_{it}\boldsymbol{\kappa} + \eta^m)\}(\boldsymbol{q}_{it}\boldsymbol{\chi} + \eta^w) + m_{it}\varepsilon_{it}^w$. Although we therefore have additive separability in our utility function, the resulting error term is not independently distributed over time or independent of the state variables, nor does it follow a common distribution. The changes made to

the original BH model, particularly the way in which employment enters our utility function, come with a substantial cost in terms of the computational complexity involved in solving the model. The choice-specific value functions can be expressed as

$$V_{it}^{w}(s_{it}, \varepsilon_{it}^{m}, \varepsilon_{it}^{w}) = \sum_{\tau=t}^{T} \beta^{\tau-t} \Phi(\boldsymbol{q}_{it}\boldsymbol{\kappa} + \eta^{m})(\boldsymbol{q}_{it}\boldsymbol{\chi} + \eta^{w}) + e_{it}^{w} = \bar{V}_{it}^{w}(\boldsymbol{q}_{it}) + e_{it}^{w}$$

and $V_{it}^{s}(s_{it}, z_i, \varepsilon_{it}^{s}) = \boldsymbol{q}_{it}\boldsymbol{\psi} + \eta^{s} + \varepsilon_{it}^{s} + \beta V_{i,t+1}(s_{it}+1, z_i) = \bar{V}_{it}^{s}(s_{it}, z_i) + \varepsilon_{it}^{s}$, where overbars are used to denote the non-stochastic component of each utility function. The optimal schooling rule for students can be expressed in terms of these value functions, $\alpha(s_{it}, z_i, \boldsymbol{\varepsilon}_{it}) = \left(V_{it}^{s}(s_{it}, z_i, \varepsilon_{it}^{s}) > V_{it}^{w}(s_{it}, \varepsilon_{it}^{m}, \varepsilon_{it}^{w})\right)$, as can the conditional choice probability $P(d = 1|s_{it}, z_i) = P(V_{it}^{s}(\boldsymbol{a}_{it}) > V_{it}^{w}(\boldsymbol{a}_{it})|s_{it}, \boldsymbol{z}_i)$.

Solving the model now proceeds in three steps. First we calculate $\bar{V}_{it}^{w}(s_{it})$, the non-stochastic component of the value of choosing to enter the labour market with $s_{it}$ schooling years. The second step uses backwards induction to calculate the non-stochastic component of the schooling value function $\bar{V}_{it}^{s}(s_{it}, z_i)$. This value is only defined for values of $s_{it} < s_{max}$, since by assumption agents are forced to leave the school system once the maximum number of schooling years is reached. The value of entering the schooling system with $s_{max} - 1$ years of completed schooling can easily be calculated as $\bar{V}_{it}^{s}(s_{max} - 1, z_i) = \boldsymbol{q}_{it}\boldsymbol{\psi} + \eta^{s} + \beta\bar{V}_{it}^{w}(s_{max})$. However, computing the value associated with earlier schooling years is considerably more complicated. Calculating $\bar{V}_{it}^{s}(s_{max} - 2, z_i)$ requires knowing $\beta Emax\{V_{it}^{w}(s_{max} - 1), V_{it}^{s}(s_{max} - 1, z_i)\}$, which is a three-dimensional integral across non-normally distributed random variables. Following the recommendation of Keane and Wolpin (1994) mentioned in section 3.4.3 we approximate this value using the method of Monte Carlo integration. The answer is used to calculate $\bar{V}_{it}^{s}(s_{max} - 2, z_i)$ and this process is then repeated until we have solved $\bar{V}_{it}^{s}(0, \boldsymbol{z}_i)$. The third and final solution step is to construct the conditional choice probability,

$$P(d = 1|s_{it}, z_i) = P(\bar{V}_{it}^{s}(s_{it} - 1, z_i) - \bar{V}_{it}^{w}(s_{it} - 1) > e_{it}^{w} - \varepsilon_{it}^{s}|s_{it}, z_i)$$

Since the Monte Carlo integration procedure already simulates the values of $e_{it}^{w}$ and $\varepsilon_{it}^{s}$, constructing the conditional choice probabilities requires little additional work. These probabilities are combined with the wage density and employment likelihoods to evaluate the likelihood function for a parameter vector $\boldsymbol{\theta}$.

### 3.5.3 Estimating the structural parameters

Armed with the knowledge of how to solve the model for a specific set of parameters $\boldsymbol{\theta}$, we now proceed to discuss the outer or estimation loop of the dynamic programming model. We first consider the case of homogenous agents for whom $\eta_{g(i)}^{l} = \eta^{l}$ where $l = w, s, m$, before allowing for heterogeneous agents. An annual discount rate of $\beta = \frac{1}{1+0.1}$ is assumed[33]. Calculating the wage densities and employment probabilities for a specific set of parameter values $\boldsymbol{\theta}_n$ under our distributional assumptions is straightforward: $L_i^w(\boldsymbol{\theta}_w, \eta^w) = (2\pi e^{\epsilon_w})^{-\frac{T}{2}} \prod_{t=1}^{T} \exp\{(w_{it} - \boldsymbol{q}_{it1}\boldsymbol{\chi} - \eta^w)^2 / 2e^{\epsilon_w}\}$[34], and $L_i^m(\boldsymbol{\kappa}, \eta^m) = \prod_{t=1}^{T} \Phi(\boldsymbol{q}_{it}\boldsymbol{\kappa} + \eta^m)^{m_{it}} [1 - \Phi(\boldsymbol{q}_{it}\boldsymbol{\kappa} + \eta^m)]^{1-m_{it}}$. In order to calculate the likelihoods associated with the schooling outcomes, we first discretise the birth year variable, $z_i$ into $10$ points. The likelihood function is now calculated as $L(\boldsymbol{\theta}) = \prod_i L_i(\boldsymbol{\theta}) = \prod_i L_i^w(\boldsymbol{\theta}).L_i^m(\boldsymbol{\theta}).L_i^s(\boldsymbol{\theta})$ and maximised using the method of simulated annealing described in section 3.4.3. We experimented with various combinations of parameters for the simulated annealing iteration process, and found that when using the values suggested by Goffe, Ferrier, & Rogers (1994) the process still had a tendency to converge on different (local) optima in successive runs. However, when estimating the model with more conservative values of $r_T = 0.75$ and $N_T = 10$, the process terminated at nearly identical coefficient estimates in different runs, which suggests convergence on a global optimum.

[33] Since this parameter is only poorly identified (Rust, 1987), we opt to assume a reasonable value rather than to attempt to estimate it. This is consistent with most other applications of this nature (Aguirregabiria & Mira, 2010, p. 39).

[34] $\sigma_l^2$ is set to $e^{\epsilon_l}$ for $l = w, s$ and $\epsilon_l$ taken as the model parameters in order to prevent the numerical optimisation procedure from assigning negative values to the error term variances.

Allowing for heterogeneous agents merely adds to the computing time required to find the parameter estimates. For any trial value of parameters the homogeneous model is solved separately for each type $g$ after which these type-specific solutions are used to calculate the likelihood function $L_{ig}(\boldsymbol{\theta}_g)$. These values are combined with the unconditional probability that an individual will be of type $g$ to form the finite mixture likelihood function according to $L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \sum_{g=1}^{G} p_g L_{ig}(\boldsymbol{\theta}_g)$. This set of parameter values, which now include the type-specific wage, employment and schooling constants as well as the type probabilities, are iterated through the different trail coefficient values using the same process of simulated annealing described above in order to find the values that maximise $L(\boldsymbol{\theta})$.

## 3.6 Empirical analysis

### 3.6.1 Data

We estimate the model using the LFS panel dataset as released by StatsSA. The LFS is a nationally representative household survey collected twice a year between 2000 and 2007, and was discussed in some depth in 2.3.2. The households sampled between September 2001 and March 2004 were allocated according to a rotating panel design, with 20% of the original sample being replaced by new households in the next round of the survey (apart from September 2002). Although there were substantial problems in matching individuals across the different surveys (Ranchod & Dinkelman, 2008), Banerjee et al. (2007, p. 59) claim that the data can give us a "good estimate of transition patterns in the data". However, it is worth taking note of a few problems with this particular dataset. Firstly, it contains no household characteristics, which seriously restricts our choice of schooling shifters. Furthermore, StatsSA did not publish survey weights for this data, and our estimates are therefore all taken from unweighted regressions. Finally, the data was collected as a rotating panel of dwelling units, which implies that individuals were dropped from our sample once they left the dwelling. This may results in non-random attrition which could potentially bias our estimates (Vermaak, 2010). Furthermore, the post-secondary schooling

outcomes for individuals sometimes evolve in a peculiar manner between successive waves, which may be indicative of coding errors or "over-cleaning" of the data during the matching of individuals from the different cross-sections. However, the LFS panel is currently the only nationally representative panel dataset and one of very few African panel datasets with enough observations to estimate this sort of model. Although we could theoretically identify all of the parameters of the homogeneous agent model with cross-sectional data only, the panel component of the data is vital in identifying the heterogeneous wage, employment and schooling utility intercepts. In order to avoid issues of racial and gender discrimination, as well as the need to model fertility decisions, we restrict our sample to black males. This gives us 38,084 individuals and a total of 110,704 observations.

### 3.6.2 Estimation results

#### *3.6.2.1 Homogeneous agents case*

The model outlined in section 3.5.1 is now estimated using the methods discussed in section 3.5.3. We start by reporting the coefficient estimates from a model in which schooling is assumed to be exogenous in both the wage and employment equations (in column 1 in Table 8). The exogenous schooling model is essentially a static model in which the employment and earnings regressions are estimated simultaneously using a maximum likelihood estimator (MLE), but with uncorrelated errors. This serves as a useful baseline for considering our dynamic programming models. Not surprisingly – given the descriptive statistics seen in section 3.3 – the schooling effect is convex in both the wage and employment functions, whereas the experience profiles are concave. Furthermore, the degree of convexity in the schooling wage returns are very steep: low levels of schooling actually decreases expected earnings but the marginal returns increase by 2.4 percentage points for each additional schooling year.

Table 8: DCDP coefficient estimates

| | Exogenous schooling model | Homogeneous DP model | Heterogeneous DP model |
|---|---|---|---|
| **Wage** | | | |
| Schooling | -0.024*** | -0.023*** | -0.01*** |
| Schooling^2 | 0.0117*** | 0.0114*** | 0.0103*** |
| Experience | 0.083*** | 0.085*** | 0.072*** |
| Experience^2 | -0.0011*** | -0.0011*** | -0.0009*** |
| Constant | 5.017*** | 5.020*** | |
| Std. dev. | 0.867*** | 0.867*** | 0.552*** |
| **Employment** | | | |
| Schooling | -0.063*** | -0.059*** | 0.085*** |
| Schooling^2 | 0.0088*** | 0.0068*** | 0.0005*** |
| Experience | 0.129*** | 0.144*** | 0.214*** |
| Experience^2 | -0.002*** | -0.0023*** | -0.0031*** |
| Constant | -1.803*** | -1.808*** | |
| **School** | | | |
| Sec school | | -1.37*** | -0.766*** |
| Ter school | | -4.000*** | -2.320*** |
| Birth year | | 0.0085*** | 0.0277*** |
| Constant | | 1.40*** | |
| Std. dev. | | 1.766*** | 1.383*** |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.
Standard errors calculated using numerical differentiation, and the delta method for the various error term standard deviations.

The wage, employment and schooling intercepts of the dynamic programming model are initially restricted to be identical across individuals rather than allowing for group-specific heterogeneity. The resulting coefficient estimates are presented in column 2 in Table 8. This model is estimated by maximising the likelihood function consisting of employment and wage densities, and the probabilities associated with observed schooling outcomes. The static exogenous schooling model clearly demonstrates that the employment and wage densities are best fitted with a convex schooling-earnings and schooling-employment share profiles, but these characteristics are difficult to reconcile with the actual schooling choices of individuals. On the other hand, concave schooling-earnings and schooling-employment share profiles do poorly in fitting the labour market outcomes but better in explaining schooling outcomes. The

numerical optimisation technique used to maximise the likelihood function mediates a tug-of-war between the schooling and wage-employment components of our likelihood function, and the resulting estimates in Table 8 can perhaps be viewed as a "compromise" solution between these two forces.

The first thing to note is that – unlike the DCDP models estimated on US data – the average schooling returns are high. Someone with 8.15 years of schooling (the mean for our sample of black males) can expect to earn a 17% return on his next year of schooling investment[35]. Compared to the static baseline case, the effect of schooling on wages is initially higher but less convex, although this difference is relatively small. Controlling for the endogeneity of schooling and the dynamic nature of decision making also leads to a substantial decrease in the convexity of the schooling-employment profile. These results mirror the effect of controlling for endogeneity using IV techniques in chapter 2, where the convexity of both the schooling-earnings and schooling-employment profiles were reduced. Another interesting feature of the DCDP estimates is that the experience-earnings profile is very similar to what we see from the exogenous schooling MLE estimates, and expected earnings will only start to decline after an individual has acquired 39 years of experience. This seems much more closely aligned with our expectations of the lifetime earnings path than the profile estimated by BH.

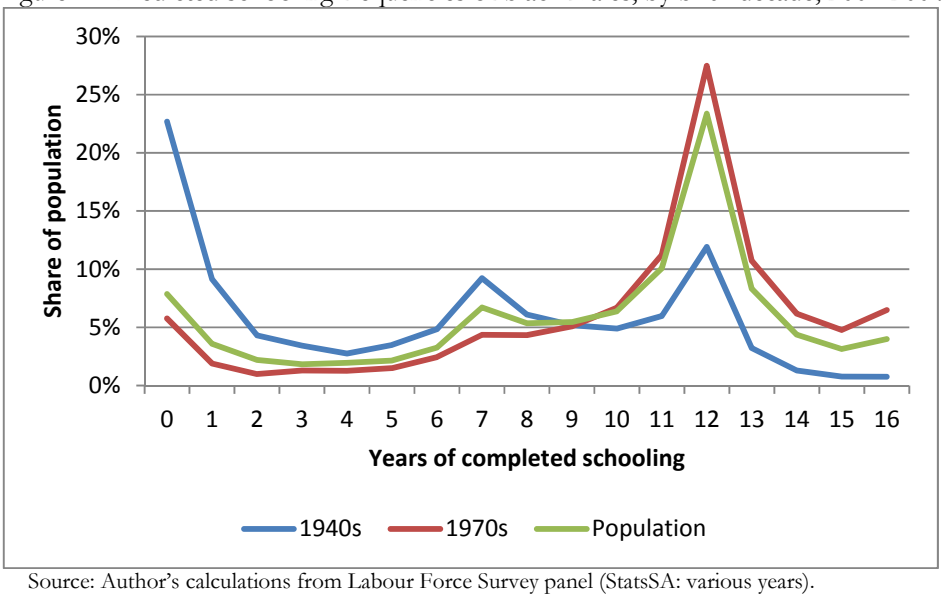
The schooling constant represents the utility derived from being in primary school by someone born in 1930 (the base birth year) and is significantly positive. Being in secondary school can be seen to reduce this utility to very close to zero, whereas tertiary students can expect to experience utility that is significantly negative. These levels are perhaps most sensibly compared to the utility of being in the labour market but not receiving a job offer, which yields a utility level of 0 and hence represents the value of the outside option. Most individuals would therefore prefer attending primary school to being unemployed (abstracting from any future costs and benefits), whereas secondary school is only marginally more desirable (at least for those from the reference birth year) and most individuals will prefer entering the

[35] This return is somewhat higher than the average returns estimated in chapter 2. However, the results from the OLS/MLE estimates also show a higher returns and degree of convexity in the earnings data from LFS panel than was estimated from the pooled OHS/LFS cross-sectional surveys. Some of the difference therefore appears to be ascribable to differences in the sample of wage earners between the two datasets rather than merely to differences between the IV and DCDP approaches.

labour market with no hope of finding work than to attend tertiary education. The magnitudes of these effects are best understood in terms of the differences in the cost of attending school and expressed relative to the utility derived from labour market earnings. Accordingly, transitioning from primary to secondary school implies an increase in this cost equivalent to a 1.37 decrease in one's current log monthly earnings. The cost increase associated with moving to tertiary education is equivalent to a further 2.63 decrease in current period log earnings. If the psychic cost of being in school is constant across the different schooling years, then these values should reflect the monetary cost of schooling, according to which secondary school is predicted to be approximately 137% more expensive than primary school and a year of tertiary education approximately 263% more expensive than a year of secondary schooling.

Individuals born more recently obtain a higher utility for each schooling year and this utility increases by 0.0085 per birth year. This coefficient can be interpreted as the decrease in the cost of attending school for black students who were born when apartheid era schooling restrictions were more relaxed than for the reference birth cohort, as discussed in section 2.4.1.1 of chapter 2.

Figure 11: Predicted schooling frequencies of black males, by birth decade, 2001-2004




Source: Author's calculations from Labour Force Survey panel (StatsSA: various years).

#### 3.6.2.2 The predicted schooling distribution

One way to gauge the validity of our model is to compare the predicted schooling outcomes – as generated by our model – to the schooling choices made by actual individuals. Figure 11 plots the predicted schooling distributions for the population as a whole, for those born in the 1940s and for those born in the 1970s. The model does a reasonable job of reproducing the stylised facts of the schooling distribution as reported in section 3.3 and graphed in Figure 10. Firstly, apart from a peak at no schooling for the older birth cohorts, most people choose to obtain intermediate schooling years rather than corner solutions. Secondly, a high share of individuals is predicted to leave school after completing secondary school. This is particularly true for those born more recently, of whom more than a quarter will choose to leave the school system at this juncture. Furthermore, whereas more than 20% of those born in the 1940s chose not to enter schooling at all, those from younger cohorts were much more likely to leave school during or after secondary school. Finally, although younger generations are much more likely to proceed to the first year of tertiary education, they are similarly unlikely to complete their undergraduate or graduate degrees.

#### 3.6.2.3 The value of additional schooling year

One of the primary benefits of using structural methods is that a completely specified model allows us to investigate the behaviour of individuals in much greater depth than an IV model can. In our model individuals choose the optimal schooling level in a rational, dynamically optimising way and the underlying structural equations that dictate their behaviour produce wage, employment and schooling outcomes that closely resemble those observed in South Africa. A natural next step is therefore to explore the reasoning of our agents in leaving school when they did.

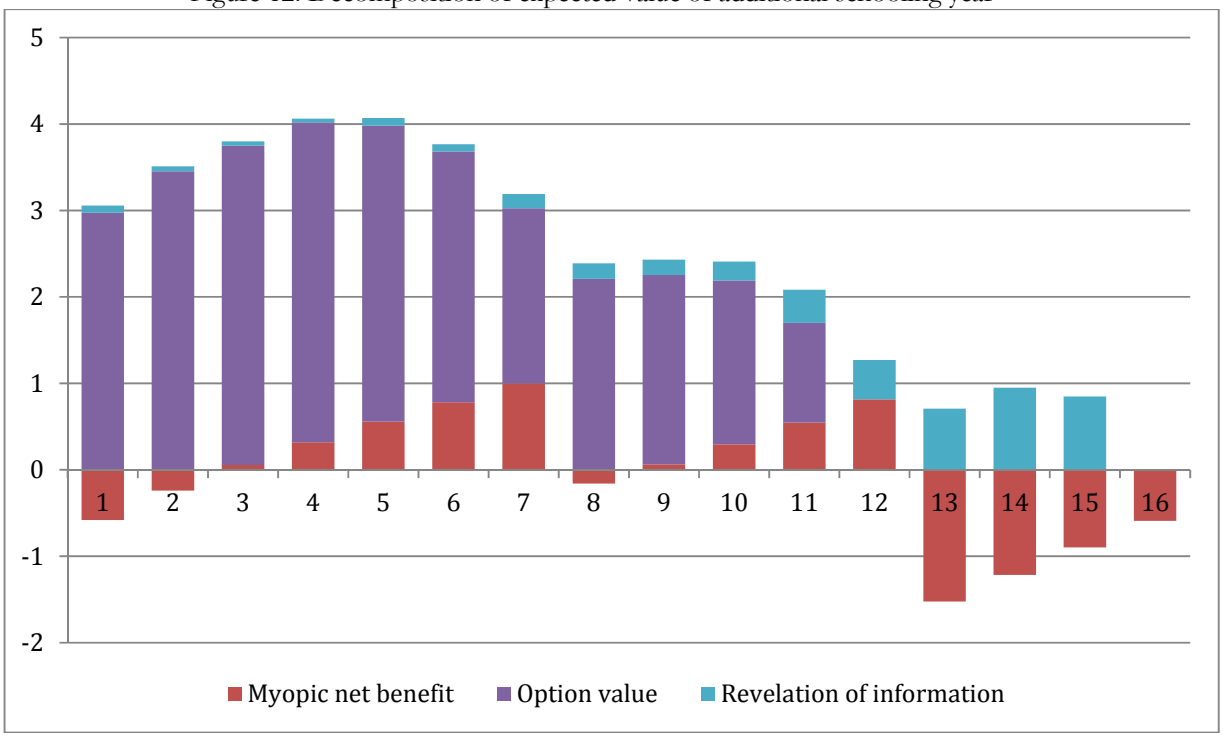
Agents decide whether to invest in another schooling year by comparing the value of staying in school to the value of entering the labour market $V_{it}^{s}(s_{it}, z_i, \varepsilon_{it}^{s}) - V_{it}^{w}(s_{it}, \varepsilon_{it}^{m}, \varepsilon_{it}^{w})$. This term can be expanded into the sum of four components. The first term is the difference in the stochastic labour market and schooling shocks: $\varepsilon_{it}^{s} - e_{it}^{w}$. This term represents the magnitude with which the difference between the received

wage offer and school utility deviates from its mean. This term will be positive if an individual receives a lower than expected (or no) wage offer, or realises that the current schooling year will be less costly than anticipated. Although this component is required to explain why individuals with the same observable characteristics behave differently, it does not help account for the systematic differences in individual behaviour that we are now interested in. The second term is $\boldsymbol{q}_{it}\boldsymbol{\psi} + \eta^s + [\beta\bar{V}_{it}^w(s_{it}+1) - \bar{V}_{it}^w(s_{it})]$ and corresponds to the "myopic" benefit of an additional schooling year. The term in square brackets represents the expected labour market benefit of having one additional schooling year (after implicitly subtracting the opportunity cost of foregone earnings while in school) and this is added to the utility of being in school for that year. In static schooling models with a concave or linear schooling-earnings profile, this term is the only non-stochastic explanation for the variation in schooling outcomes. Of course, some individuals will choose to stay in school for more than just the next schooling year, and the value of having access to further schooling is expected to be particularly important in labour markets with convex schooling-earnings profiles. The third term $\beta max\{\bar{V}_{it}^s(s_{it}+1, z_i), \bar{V}_{it}^w(s_{it}+1)\} - \beta\bar{V}_{it}^w(s_{it}+1)$ represents the value of having access to schooling beyond the next level of schooling. This term is sometimes referred to as the "option value" of schooling and is only positive if the expected value of staying in school for an additional year exceeds the expected value of entering the labour market after the current school year. The fourth and final component $\beta[Emax\{V_{it}^s(s_{it}+1, z_i), V_{it}^w(s_{it}+1)\}] - \beta max\{V_{it}^s(s_{it}+1, z_i), V_{it}^w(s_{it}+1)\}$ captures the benefit of staying in school so as to observe the wage and schooling error draws for another year. Since the expected value of the maximum of two random variables will always be greater than or equal to the maximum of their expected values, this "sequential revelation of information" term will always be non-negative. However, it is especially beneficial for those at the margin of the school-work decision, for whom a slightly larger than expected schooling shock could make it beneficial to invest in another schooling year. Clearly, such individuals stand to gain more from observing next year's wage offer and schooling error term than those who already know that they are highly unlikely to continue on with further schooling.

Plotting the three non-stochastic components of the net schooling benefit across the different schooling levels gives us the graph in Figure 12 (plotted for an individual born during the 1960s). The height of the stacked bars represents the discounted expected net value of investing in a particular schooling year and expressed in terms of current period log earnings. Each individual's stochastic shock term $\varepsilon_{it}^{s} - e_{it}^{w}$ is added to this value in order to determine whether to invest in schooling or not, depending on whether the total is positive or not. The higher the net expected value of a specific schooling year, the larger the share of individuals who will complete this year. The bars in Figure 12 therefore predict that a decreasing share of individuals from this birth cohort will leave the education system after the first few schooling years, but that this outflow will increase after the fifth schooling year and especially after learners enter secondary school. Once they get to tertiary education, the net expected benefit of investing in education is actually negative, which implies that most individuals will expect not to continue with schooling past this point.

Turning now to the composite parts of the schooling benefit, we see that the myopic benefits are negative for the first two schooling years as the convex schooling-earnings and employment profiles predict that the earliest schooling years actually decrease expected labour market earnings. For more advanced primary schooling years the myopic labour market benefit turns positive and then continues to increase as the marginal benefit of schooling grows and the schooling cost remains constant. The myopic benefit again turns negative for the first year of secondary school – due to the higher cost associated with this more advanced schooling level – but increases throughout secondary school before falling and staying below zero for the duration of tertiary education. During primary and most of secondary school, the relatively small myopic net benefit components are dominated by the larger positive option values. By far the most important reason why agents in our model choose to complete primary and early secondary school, despite the relatively low returns associated with these investments, is the promise of higher returns further up the schooling ladder.

Figure 12: Decomposition of expected value of additional schooling year




Source: Author's calculations from Labour Force Survey panel (StatsSA: various years).

During early schooling levels the value attached to the sequential revelation of information regarding wage offers and schooling error terms is still quite small. At this point most agents are fairly certain that they will continue on to the next schooling year, and therefore attach little value to the benefit of observing another period's error terms before making a decision: a very large negative schooling error term (or a very enticing job offer) is required to change their minds and given how infrequently this occurs agents are not willing to sacrifice much for this additional information. As agents approach the final year of secondary school and start to capitalise on the higher yielding educational investments, the option value of further schooling decreases. The high cost of tertiary education means that individuals do not expect to receive a positive return to this investment (unless they draw a high schooling and low wage error term) and hence attach no option value to any schooling beyond the penultimate year of secondary school. The fact that individuals move nearer to the margin of the school-work decision means that the value of additional information also increases. Intuitively, part of the appeal of completing secondary school is because it affords individuals the opportunity to observe their schooling error for the first year of tertiary education. Although most

individuals do not expect to continue past secondary school, they know that a high schooling error term – doing well in the matric exam and being offered funding for university or having someone in the household finding work – will make it worth their while to enrol in tertiary education, particularly if no job offers are forthcoming.

#### *3.6.2.4 Heterogeneous agents case*

Our model also allows us to estimate the dynamic programming model while allowing individuals to belong to groups with different schooling, wage and employment intercepts. The parameter estimates from this model are reported in the final column in Table 8. The main difference between the homogeneous and heterogeneous models is a further decrease in the convexity of the schooling-earnings and schooling-employment share profiles. The schooling returns for someone at the mean schooling level are still very high: approximately 16%. Although the schooling-earnings profile is less convex than was suggested by an exogenous schooling MLE model, this reduction is smaller than what was found by some of the IV methods discussed in chapter 2. On the other hand, after adding heterogeneity the schooling-employment profile is very nearly linear, which implies a much larger reduction in the convexity of this profile than was found using the IV methods in chapter 2.

Table 9: Heterogeneous intercept estimates from DCDP model

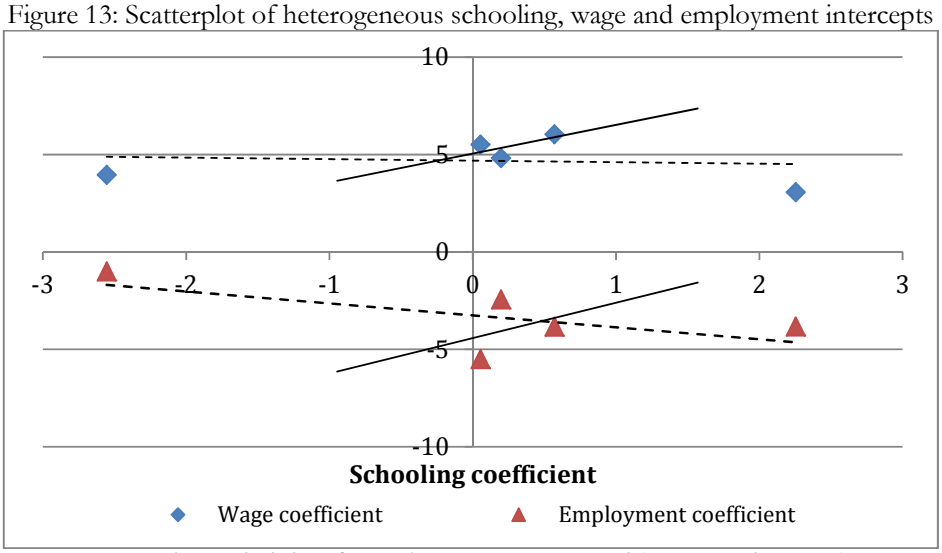
| | Wage | Employment | Schooling | Type shares |
|---|---|---|---|---|
| Type 1 | 3.958*** | -0.998*** | -2.554*** | 10.7% |
| Type 2 | 5.51*** | -5.5*** | 0.054*** | 32.5% |
| Type 3 | 4.814*** | -2.435*** | 0.197*** | 37.7% |
| Type 4 | 6.037*** | -3.838*** | 0.57*** | 16.4% |
| Type 5 | 3.071*** | -3.824*** | 2.254*** | 2.7% |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Heterogeneity also decreases the variance of the wage and schooling error terms, which indicates that this model is better able to explain these outcomes. The type-specific intercepts are reported in Table 9, along with the population shares that each type represents, and plotted in Figure 13. The general pattern does

not appear to reflect the notion of high ability individuals who complete more schooling, earn higher wages and are more likely to be employed. In fact, linear trend lines through the scatterplot (indicated with dotted lines in Figure 13) suggest that high schooling ability individuals are generally slightly less likely to be employed and earn about the same as those with low schooling ability. Upon closer inspection, we see that this is largely driven by two "outlier groups": types 1 and 5. Type 1 individuals are much more likely to obtain very low levels of schooling, while having a higher-than-average employment rate and a slightly lower-than-average earnings. Agents of type 5 are much more likely to achieve high levels of schooling, but have below-average earnings and an average employment likelihood. However, these two groups constitute a relatively small combined share of the total population of 13%. The three largest groups (types 2, 3 and 4) are more narrowly distributed together in terms of academic ability, and for this subgroup of the population higher academic abilities are also associated with higher earnings and employment abilities. These trends are indicated with the solid trend lines in Figure 13.

Figure 13: Scatterplot of heterogeneous schooling, wage and employment intercepts




Source: Author's calculations from Labour Force Survey panel (StatsSA: various years).

## 3.7 Conclusion

This chapter estimates a dynamic programming model of the South African schooling-earnings nexus that attempts to explain the observed patterns in schooling, wages and employment as produced by rational, dynamically optimising agents who face no other restrictions than their budget constraint. The model is able to reproduce most of the salient features of the schooling distributions of individuals from different generations, but requires marginal schooling costs (or disutility) to be steeply increasing across the various schooling categories. In terms of the effect of schooling on earnings, the first important result is that the schooling returns are no lower, on average, than suggested by exogenous schooling models or IV estimators. This is interesting given that DCDP models estimated on US data usually produce return estimates that are substantially lower than what is obtained from OLS or IV methods. The model estimates indicate that the schooling-earnings profile is convex, although the degree of convexity is less than suggested by previous studies that use MLE or OLS to estimate this profile. The effect of additional schooling on the employment likelihood is also an increasing function of schooling, but this effect seems to be grossly overstated when using an exogenous schooling model.

Our analysis also introduces a decomposition technique that allows us to explain the schooling choices of agents in our model. Most individuals choose to complete primary and early secondary schooling years despite the relatively low returns associated with these investments, because this offers them the option of proceeding on to the high yielding final years of secondary school. However, the exorbitant cost of enrolling in tertiary education means that most agents will choose to drop out after completing secondary school. The pattern of unobserved abilities is more complex than the usual notion of high ability individuals who earn more, get more schooling and are more likely to be employed, although this trend can be observed for the bulk of individuals who have neither very low nor very high academic abilities.

# 4. Estimating the schooling-productivity profile using the production function approach


## Abstract

This chapter uses the production function approach to estimate the effect of schooling on the productivity of South African workers. Production functions are estimated for South African industries using a novel industry panel dataset that combines education and employment data from a series of household surveys with output and physical capital data from the South African Reserve Bank Quarterly Bulletins. The pooled ordinary least squares estimates indicate that the returns to education are substantial and concave, but these effects disappear when industry fixed effects are included in the regression. This result is similar to what has been found in other African production function studies. However, we demonstrate that the strong schooling effects are restored when making proper allowance for measurement error or parameter heterogeneity and cross-sectional dependence. The results suggest that industry productivity and the average schooling level of workers are highly correlated and difficult to identify simultaneously using only time series variation in the data.


## 4.1 Introduction

Despite the large number of studies that find a strong schooling effect on individual earnings levels, attempts to identify a similar effect from production function estimates on African firm or sector level data have generally been unsuccessful. The production function approach provides a more direct method of estimating the effect of schooling on labour productivity than the earnings function approach, and can produce consistent estimates of the schooling impact even where human capital externalities or institutional features of the labour market mean that workers are not paid their marginal revenue product.

This chapter estimates industry-level production functions for nine South African industries over a twelve year period in order to determine the total monetary return to investing in human capital, rather than just the return that accrues to the individual. The estimation of these industry production functions are made possible by our unique dataset, which merges physical capital and output data from the SARB Quarterly Bulletins with industry employment and education estimates from a series of twenty-one consecutive StatsSA household surveys. Given the strong empirical evidence against the assumption that all sectors produce according to the same production function (Burnside, 1996; Eberhardt & Teal, 2008) or that production factors are uncorrelated to industry productivity (Söderbom & Teal, 2004), recent studies have preferred to use estimators that allow these features to be incorporated into their econometric models. However, by transforming away much of the variation in the regressors, such estimators may be more susceptible to measurement error and other types of biases not explicitly taken into consideration by the identifying assumptions. In certain cases, this may even lead to estimates that are less reliable than those of the pooled ordinary least squares estimator.

Section 4.2 below reviews the literature regarding the estimation of production functions. This is followed by a brief explanation of the data used in this study in section 4.3. Section 4.4 starts by describing our theoretical model, proceeds to outline the different types of endogeneity that could confound our analysis, and then analyses the characteristics of the various estimators that have been proposed to circumvent these problems. The results for these estimators are discussed in section 4.5, after which section 4.6 concludes.

## 4.2 The effect of education on labour productivity

There exists a substantial literature of Mincerian earnings regression studies for African countries (as discussed in section 2.2 of chapter 2) that find substantial, positive and convex schooling returns. This is usually interpreted as evidence of the strong effect of education on labour productivity, although chapters

2 and 3 discussed some reasons why this is not the only conclusion that is consistent with these results. Surprisingly, and somewhat disappointingly to proponents of greater education expenditure in African countries, production function studies have generally not been able to replicate this result.

Bigsten et al. (2000) estimate firm-level production functions for the manufacturing sectors of five sub-Saharan African countries: Cameroon, Ghana, Kenya, Zambia and Zimbabwe. Despite finding high returns to education in their earnings regressions, the log of average worker education is insignificant in the production function regressions for all five countries[36]. Their results also include a capital coefficient estimate which is almost twice that of the employment coefficient, despite the fact that labour's share in income exceeds that of capital in most countries[37]. Söderbom and Teal (2004) find that schooling is significant in a Cobb-Douglas production function of Ghana's manufacturing sector when estimated using pooled OLS, but that this effect disappears when estimated with a fixed effects estimator. They interpret this as evidence that the schooling level of workers is not very important in determining the productivity of African manufacturing firms. Appleton and Balihuta (1996) review studies that estimated the effect of education on labour productivity in the agricultural industries of African countries, and find that the effect is usually either insignificant or small in magnitude. Their own estimates for Uganda show that although primary education has a significantly positive effect in raising agricultural production, the returns to secondary school are insignificant and the overall returns are much lower than those usually found in earnings regressions.

There are at least three ways to explain the very different results sometimes produced by the earnings function and production function approaches. Firstly, it is possible that the former only captures the private monetary returns to schooling investment, whereas the latter also captures education externalities and spill-overs in the production process. In this case the difference between the two estimates can be interpreted as a reflection of human capital externalities. However, this requires assuming that individuals

[36] The sum of log education and log tenure is significant for four of the five countries, with coefficients varying between 0.1 and 0.29.

[37] If markets are perfectly competitive and production occurs with a constant returns to scale Cobb-Douglas production function, then the share of income accruing to labour and capital will be the labour and capital coefficients, respectively.

are paid their marginal revenue product, and that both methods provide unbiased estimates of schooling's effect on productivity. The second reason why earnings and production function schooling coefficients may diverge is because the former is affected by institutional characteristics of the labour market – such as union bargaining power, minimum wages or public sector wage premia – that drive a wedge between wages and marginal revenue product. Finally, it is also possible that data or specification problems produce biased estimates of the schooling coefficient in either or both approaches. The econometric issues that can arise when estimating the schooling effect in earnings equations were discussed in chapters 2 and 3, whereas the very different identification issues in production functions are considered in section 4.4.

Although a number of studies have attempted to estimate Cobb-Douglas production functions for South Africa[38], data restrictions mean that such studies usually cannot control for human capital and use either time series data at the national level or cross-sectional data at the firm level. One notable exception is Behar (2010) who includes workers employed in the different skills categories as separate inputs in the production function. Although this is potentially informative about skills prices in the South African labour market, the skills categories are defined according to worker occupations rather than schooling levels. His research question also requires the estimation of a translog production function[39] rather than a Cobb-Douglas specification, so that his results are not directly comparable to ours.

The existing South African studies that most closely resemble our own is Fedderke (2002; 2002) which use a panel of South African manufacturing sectors to estimate the effect of human capital variables on total factor productivity growth. These studies use SARB data at a two-digit sector classification, for which output, physical capital and labour data are available, but for which no information on the average schooling level of workers exists. Fedderke (2002, p. 3) acknowledges the problems posed by labour quality differentials, but claims that "data limitations preclude the possibility of pursuing this line of inquiry further". Two approaches are used in an attempt to circumvent this problem: Fedderke (2002) uses

[38] Examples include Arora and Bhundia (2003), Smit and Burrows (2002), Bhorat and Lundall (2004) and Bonga-bonga (2009).
[39] Although the added flexibility of these more sophisticated production functions are desirable, it introduces the issue of substitutability between the different factors of production which reduces the tractability of the model, the comparability with the previous chapters and takes us beyond the scope of this dissertation.

occupational information to classify workers into different skills categories, whereas Fedderke (2002) employs a range of national-level education variables with only time series variation. Although his econometric model allows parameter heterogeneity in how the education variable affects sectoral productivity, this measure is derived as the residual from a regression in which the effects of capital and labour on output are assumed to be constant across all sectors. In this sense his specification is more restrictive than those considered in our own econometric analysis. Fedderke (2002) finds that white school enrolment rates and the number of natural and engineering science degrees are positively correlated to productivity, whereas the logarithm of black enrolment rates is negatively correlated to productivity. This leads him to conclude that "[w]hat matters crucially in the long run is the quality of education" (Fedderke, 2002, p. 29). However, it is also possible that the restrictiveness of his specification or data limitations may be precluding his regressions from accurately identifying the effect of his schooling quantity measures.

The difficulties in identifying a substantial schooling effect on output have also troubled researchers interested in determining the causes of cross-country productivity differentials. Exploring how this puzzle has been addressed in the cross-country literature may therefore offer insights into our own research question. There are at least three reasons why cross-country data would fail to find a large effect of education on labour productivity, whereas the bulk of earnings regressions show high returns to education. Firstly, if education merely serves as a signal of high inherent market ability without actually increasing one's productivity (Spence, 1973), then the highly educated will always earn more than those with lower levels of education, but an increase in a country's average level of education will not enhance the productivity of its workers. In this case the observed low cross-country return estimates will provide a more reliable reflection of the true effect of education on productivity.

Secondly, although cross-country data average away most individual level misreporting of schooling years within countries, this data also introduces other measurement issues such as aggregation bias or international differences in schooling quality. Krueger and Lindahl (2001) review a number of studies that investigate the cross-country relationship between education and economic growth, and find that such

studies almost invariably suffer from schooling being measured with a great deal of error across countries. This produces the usual downward attenuation bias in the schooling coefficient and also – given the high positive correlation between human and physical capital across countries – an upwardly biased capital coefficient. Their hypothesis is supported by the findings of De la Fuente and Doménech (2006), who suggests that correcting for the quality of the schooling data substantially increases the estimated schooling coefficient in cross-country growth regressions. Krueger and Lindahl (2001) propose fixing the capital and labour coefficients to reasonable values, such as their respective shares of total income, in order to estimate more accurately the effect of education on labour productivity.

Thirdly, the effect of education may vary across groups of countries so that a regression model that assumes a constant education effect may suffer from misspecification bias (Temple, 1999, p. 132). Judson (1998) finds evidence that the effect of education on productivity depends on how efficiently educational institutions allocate their resources, which suggests that the education effect should be smaller in developing countries where inefficiencies in the school system are more likely to be an issue. This would also explain the failure of many African countries to convert increased levels of educational attainment into higher output growth. In a similar vein, Pritchett's (2001) observes that in many developing countries the highly educated are drawn to better paying but less productive public sector jobs, which would further reduce the productivity-enhancing character of education. These problems all suggest using a more flexible estimator that allows for parameter heterogeneity when attempting to accurately estimate the average schooling effect on labour productivity.

These problems are suggestive of the type of econometric problems that may confound the estimation of the schooling effect on worker productivity using African data. Properly accounting for measurement error and the differences in production technologies across firms or industries may be crucial in identifying this effect, and failing to do so may explain why production function studies of African data have hitherto failed to find any substantial schooling effect.

## 4.3 Data

Our literature review suggested measurement error and parameter heterogeneity as two potential sources of endogeneity that could bias the estimated effect of worker education on labour productivity. Exploring these issues requires a South African industry or firm-level panel dataset that contains a measure of worker education levels, has a time dimension that is sufficiently long to allow parameter heterogeneity across industries, and should ideally also contain additional variables that can be used as instruments for potentially mismeasured input variables. No such a dataset exist, so we construct an industry level panel using two separate data sources: the SARB Quarterly Bulletin and the StatsSA household surveys.

The StatsSA household surveys discussed in section 2.4.2 of chapter 2 offer the richest source of medium-term South African labour market trends. These surveys include individual responses to questions regarding employment, years of schooling completed and industry of occupation that can be used to estimate the number of formal sector employees (of any race or gender) working in different industries as well as their average years of completed schooling. They also provide information that can serve as instrumental variables for employment, such as the share of industry workers that belong to trade unions or the average industry wage rate. The industry and period specific employment figures are calculated by using the sampling weights provided by StatsSA for each of these surveys. The OHS1995 to February LFS2000 surveys were weighted using the 1996 Census weights, while all those following were weighted using the 2001 Census weights. Studies that compare the observable characteristics of the population (as constructed using these weights) between these surveys find that they are relatively reliable[40] (Yu, 2007). Attempts to improve the consistency of the weights using a cross entropy estimation approach yield results that are very similar than those produced with the original weights (Branson, 2009).

The SARB data are collected from South African firms, usually at either monthly or quarterly intervals. This data includes variables for "gross value added by kind of economic activity" and the "fixed capital stock by kind of economic activity", which we will use as our measures of output and physical capital

[40] Apart from the sampling and questionnaire design issues discussed in section 2.4.2.

respectively. This dataset also contains a measure of the long-term government bond yield which can be used as a measure of the cost of capital and hence as an instrument for the capital stock. The kind of economic activity that firms engage in is classified into nine different industries[41] using the ISO one digit categories. Although these variables are also available at the two-digit sector level (as used by Fedderke (2002; 2002)), constructing the employment and schooling variables at this lower level of aggregation would mean using fewer observations for each estimate and compounding any measurement error problems in the data. These nine industries are therefore used as the cross-sectional units of observation for our production function model. The industry capital stock values[42] are recorded annually for all sectors (as well as quarterly for certain industries) and measured at constant 2000 prices. The industry output values[43] are recorded quarterly for all industries, measured at constant 2000 prices and unadjusted for seasonal fluctuations. These variables are combined with the employment and education data from the household surveys to construct a South African industry panel dataset spanning 13 years[44] and nine industries.

Given the important role assigned to measurement error in explaining certain results in the cross-country human capital-growth literature, it is worth briefly discussing the nature of the measurement issues that affect our data. Some studies have questioned the reliability of the SARB data – for example, Van Walbeek (2006) highlights the large revisions that are periodically made to the SARB data, and the substantial impact that these changes can have on empirical studies – but much more attention has focused on the problems in comparing the Stats SA household surveys. Many papers (Casale, Muller & Posel (2004), Kingdon & Knight (2005), Burger & Yu (2006) and Altman (2008)) discuss the effect that modifications in questionnaire design and sampling methodology may have had on the comparability of the household

[41] The nine industries are 1) agriculture, forestry and fishing, 2) mining and quarrying, 3) manufacturing, 4) electricity, gas and water, 5) construction, 6) wholesale and retail trade, catering and accommodation, 7) transport, storage and communication, 8) finance, insurance, real estate and business services, and 9) community, social and personal services.
[42] SARB time series codes 6140Y-6148Y.
[43] SARB time series codes 6631Y, 6632Y, 6634Y-6636Y, 6638Y-6640Y and 6642Y.
[44] The second bi-annual LFS (usually conducted in September or October) could potentially be used to add an additional observation per industry for each year between 2000 and 2007, although the panel time periods will then no longer be equally spaced. However, the industries for which the SARB only tracked the capital stock at an annual frequency would require the imputation of mid-year capital stock values. In order to avoid this additional complication we choose to use a panel dataset with an annual frequency.

surveys over time. These concerns were discussed in depth in section 2.4.2. It was noted that the most serious comparability problems occur for informal sector or self-employed workers, so that the effect of these inconsistencies can be limited by omitting these workers from the sample and restricting our dataset to formal sector employees only. Since the SARB firm surveys are almost certain to ignore the bulk of industry output arising from (and the capital stock owned by) firms operating in the informal sector, omitting these workers is also likely to improve the internal consistency of our dataset.

Altman (2008) investigates the StatsSA household data by comparing industry employment trends to those obtained from alternative sources, including employment data derived from a series of establishment surveys. This series, known successively as the Survey of Total Employment and Earnings, the Survey of Employment and Earnings and the Quarterly Employment Survey, also underwent numerous changes in surveying methodology and sampling frames, and is therefore unlikely to provide a more accurate measure of industry employment trends[45]. Assuming that these changes were independent of the questionnaire design and sampling adjustments that affected the household surveys, this second employment measure offers a useful benchmark to which the StatsSA employment variable can be compared, and possibly also an instrumental variable that can be used to identify our model parameters in the presence of measurement error. Altman (2008) finds evidence of substantial errors in the industry employment totals derived from the household surveys – particularly in the agriculture, mining and community, social and personal services industries – but her analysis still indicates that these household surveys provide the "most comprehensive and reliable sources of employment trend data for the past decade" (Altman, 2008, p. S127).

## 4.4 Production function and model identification

This chapter is mainly concerned with identifying the average causal effect of additional schooling on the productivity of workers and particularly whether properly addressing measurement error and parameter heterogeneity can explain why so many African production function studies have failed to find a significant

[45] This employment series is used most other South African production function studies, such as Fedderke (2002; 2002).

worker education effect. Our production function is of the human capital augmented Cobb-Douglas form, similar to that used in Hall and Jones (1999) and Bils and Klenow (2000) except that we allow non-constant returns to scale and also for schooling years to enter the production function non-linearly:

$$Y_{nt} = A_{nt}K_{nt}^{\alpha_n}\left(e^{\phi_{1,n}\bar{E}_{nt}+\phi_{2,n}\bar{E}_{nt}^2}L_{nt}\right)^{\gamma_n} \quad [16]$$

The economy consists of $N$ different industries (generically denoted $n$) each of which is observed over $T$ different periods (indexed by $t$). Industries combine physical capital, $K$, the average education of workers, $\bar{E}$, labour, $L$, and Hicks-neutral technology, $A$, in order to produce output, $Y$. The technological parameters include the physical and human capital coefficients, $\alpha$ and $\gamma$, as well as the education coefficients $\phi_1$ and $\phi_2$. Allowing for parameter heterogeneity means that these coefficients will vary across industries.

The form of the human capital link function, $\phi_{1,n}\bar{E}_{nt} + \phi_{2,n}\bar{E}_{nt}^2$, is of particular interest due to its close relationship with its counterpart in the earnings function. Hall and Jones (1999) use this fact to argue that the human capital link function in the "macro-Mincer" equation should be piece-wise linear, with coefficients taken from Mincerian earnings function estimates. Given the micro-economic evidence in favour of using a quadratic specification to model the effect of education on wages (for example Lemieux (2006), and the results presented in chapters 2 and 3), our preferred specification will be as in equation [16], although we also experiment with linear and logarithmic functions. Moreover, unlike Hall and Jones (1999) the education coefficients are estimated directly from the data rather than fixing them to estimates obtained from other studies.

Industry productivity is generated according to $\log A_{nt} = a_{nt} = \bar{a} + \eta_n + \chi_n\tau_t + \varepsilon_{nt}$ where $\bar{a}$ is a constant, $\eta_n$ is the time-invariant industry productivity effect, $\tau_t$ is a universal time shock, $\chi_n$ represents the industry's output response to this shock, and $\varepsilon_{nt}$ is the remaining productivity innovation. We assume, without loss of generality, that $E(\eta_n) = E(\tau_t) = E(\varepsilon_{nt}) = 0$. Writing equation [16] in logarithmic form, gives

$$y_{nt} = \alpha_n k_{nt} + \gamma_n l_{nt} + \gamma_n\left(\phi_{1,n}\bar{E}_{nt} + \phi_{2,n}\bar{E}_{nt}^2\right) + \bar{a} + \eta_n + \chi_n\tau_t + \varepsilon_{nt} \quad [17]$$

where $k_{nt}$ and $l_{nt}$ are the logs of capital and labour. The input and technological parameter vectors are defined more succinctly as $\boldsymbol{x}_{nt} = [k_{nt}, l_{nt}, \bar{E}_{nt}, \bar{E}_{nt}^2]$ and $\boldsymbol{\beta}_n' = [\alpha_n, \gamma_n, \gamma_n\phi_{1,n}, \gamma_n\phi_{2,n}]$. The production parameters are random coefficients drawn from a distribution with $E(\boldsymbol{\beta}_n) = \boldsymbol{\beta}_*$ and the stochastic industry-specific technological parameter deviation is defined as $\boldsymbol{b}_n = \boldsymbol{\beta}_n - \boldsymbol{\beta}_*$.

Whenever we allow for measurement error in the production factors the observed input values can be expressed as $\boldsymbol{x}_{nt} = \boldsymbol{x}_{nt}^* + \boldsymbol{e}_{nt}$, where $\boldsymbol{x}_{nt}^*$ is the vector of actual but unobservable factor input values and $\boldsymbol{e}_{nt}$ is the measurement error. Log output can thus be expressed in terms of the observable input measure, $\boldsymbol{x}_{nt}$, the average industry production coefficient vector, $\boldsymbol{\beta}_*$, and a composite error term, $u_{nt}$, via the following equation $y_{nt} = \boldsymbol{x}_{nt}\boldsymbol{\beta}_* + u_{nt}$ where

$$u_{nt} = \bar{a} + \eta_n + \chi_n\tau_t + \boldsymbol{x}_{nt}\boldsymbol{b}_n - \boldsymbol{e}_{nt}\boldsymbol{\beta}_n + \varepsilon_{nt}. \quad [18]$$

Although the industry productivity coefficients, $\boldsymbol{\beta}_n$, are all of interest in their own right, we are mainly interested in the population averages of these coefficients, $\boldsymbol{\beta}_*$. We would expect $\alpha_n$ and $\gamma_n$ to be strictly larger than zero but smaller than 1 for all $n$, and perhaps also for $E(\alpha_n)$ and $E(\gamma_n)$ not to deviate too much from capital and labour's shares in total South African income: around 0.45 and 0.55 respectively, according to Arora and Bhundia (2003). Although we do not have strong priors about the values of $\phi_{n,1}$ and $\phi_{n,2}$, we would expect them to be such that output is an increasing function of worker schooling level over the range of $\bar{E}_{nt}$. Given that earnings regressions on South African data generally find convex schooling-earnings profiles (as discussed in chapter 2), we may also expect to find that $E(\phi_{2,n}) \geq 0$, although some reasons why these estimates may diverge were discussed in section 4.2.

Various estimators can be used to recover the technological parameters $\boldsymbol{\beta}_*$, depending on which set of identifying assumptions are most appropriate for our model. In section 4.5 we explore the coefficient

estimates produced by a wide range of estimators. Before discussing each of these in turn, we consider a general class of least squares models that can be expressed as:

$$\widehat{\boldsymbol{\beta}}_M = (\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{y} = \boldsymbol{\beta}_* + (\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{u}$$

where $\boldsymbol{X}$ is the matrix of observable production factors (stacked across industries and periods), $\boldsymbol{y}$ is the output vector and $\boldsymbol{u}$ is the vector of unobservable productivity components. The $\boldsymbol{M}$ matrix performs a linear transformation on the data $(\boldsymbol{y}, \boldsymbol{X})$, and is usually chosen to ensure that $\widehat{\boldsymbol{\beta}}_M$ will be an unbiased estimator of $\boldsymbol{\beta}_*$. The objective of this transformation is to discard potentially endogenous variation from the observed production factors in $\boldsymbol{X}$. This logic has been used to justify a wide range of estimators that use different transformation matrices to rid the estimates of various types of endogeneity. However, if the error term contains many different unobserved factors, as in equation [18] above, then it is possible that a transformation can successfully remove the correlation between the regressors and one of the error components, while also increasing the estimator bias $(\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}'E(\boldsymbol{M}\boldsymbol{u}|\boldsymbol{X})$. This will be the case if the estimator transforms away too much of the regressor variation $\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{X}$ that serves to mitigate any remaining covariance between the errors and regressors.

### 4.4.1 Pooled OLS and random effects estimators

A natural point of departure for our econometric analysis is the pooled ordinary least squares (POLS) estimator: $\widehat{\boldsymbol{\beta}}_{POLS} = (\boldsymbol{X}'\boldsymbol{M}_P\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_P\boldsymbol{y}$,[46] where $\boldsymbol{M}_P = \left(\boldsymbol{I}_{NT} - \frac{1}{NT}\boldsymbol{i}_{NT}\boldsymbol{i}'_{NT}\right)$. This estimator will provide unbiased estimates of the technological parameters if $E(\boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{0}$, which requires that all of the observed inputs are uncorrelated with the industry fixed effects $\eta_n$, the time shocks $\tau_t$, the industry-specific time

[46] The POLS coefficients of the regressors can be recovered either by regressing output on the factors of production and a constant, or by regressing demeaned output on the demeaned factors of production, without a constant (as demonstrated by Frisch and Waugh (1933) and Lovell (1963)). The notation used here follows the latter method and is used in order to aid the comparability of the various estimators.

shock response $\chi_n$, the industry technological parameter deviations $\boldsymbol{b}_n$, the measurement errors $\boldsymbol{e}_{nt}$ and any remaining productivity innovations $\varepsilon_{nt}$.

Clearly there are many reasons why the POLS estimator could produce biased results. The estimators discussed in the rest of this section attempt to remove the bias that occurs due to correlation between the factors of production and the different unobserved productivity components. Although the POLS estimator is sensitive to many types of endogeneity, it is also the only estimator that uses all of the available variation in the data with calculate the coefficient estimates. This is a drawback if some of the regressor variation is correlated with unobserved productivity, but a crucial advantage compared to estimators that discard information vital to accurately identifying the true production parameters.

Even if the conditions for POLS unbiasedness are met, the error structure of $u_{nt}$ may still be exploited in order to provide more efficient estimates of the technological parameters if industry productivity differs in a way that does not affect employment or investment decisions. In the absence of measurement error and time shocks, the RE estimator can be estimated as $\widehat{\boldsymbol{\beta}}_{RE} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$, where $\boldsymbol{\Omega}$ is the block diagonal error covariance matrix.

### 4.4.2 Fixed effects, time effects or two-way fixed effects estimators

In a simple model where firms maximise current period profits and all firms face the same factor costs, those in high-productivity industries will employ more workers and invest more in physical capital than those in low-productivity industries. By the same logic high productivity periods will also coincide with employment and investment booms. The results from POLS estimators are therefore viewed with an understandable sense of suspicion in the production function literature (see Söderbom and Teal (2004) for example), and the use of panel data estimators that are unbiased in the presence of correlated industry effects has become standard econometric practice.

The fixed effects (FE) estimator is calculated by transforming away all time-invariant industry-specific variation in the data, using the idempotent within-groups transformation matrix, $\boldsymbol{M}_{WI} = (\boldsymbol{I}_{NT} - \boldsymbol{I}_N \otimes \boldsymbol{i}_T \boldsymbol{i}_T')$[47], so that $\widehat{\boldsymbol{\beta}}_{FE} = (\boldsymbol{X}'\boldsymbol{M}_{WI}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{WI}\boldsymbol{y}$. Since the implied industry-demeaning removes potentially endogenous variation from the error and regressors, the conditions for FE estimator unbiasedness are less stringent than they are for POLS. Specifically, the existence of any correlation between the factors of production and industry fixed effects, $E(\eta_n|\boldsymbol{x}_{nt}) \neq 0$, will bias the POLS but not the FE estimators. The within-groups transformation also removes industry-specific measurement error and correlation between the industry technological coefficients and the factor inputs that could otherwise bias the POLS coefficient estimates. However, if some endogeneity remains after performing the within-group transformation – due to the importance of measurement error or global time shocks, for instance – then there is no guarantee that the FE estimates will be less biased than the POLS estimates. Whether this will be the case or not depends on how much of the regressor variation is removed by the within-groups transformation and how much endogeneity remains afterwards.

It is also possible to remove all the industry-invariant time-specific variation from the data with what can be called the "within-time periods" transformation $\boldsymbol{M}_{WT} = (\boldsymbol{I}_{NT} - \boldsymbol{i}_T \boldsymbol{i}_T' \otimes \boldsymbol{I}_N)$. This estimator – which we will for reasons of symmetry call the time effects (TE) estimator – will be biased in the presence of correlated fixed effects, but will produce unbiased results where correlation between the universal time shocks and the production factors would bias both the POLS and the FE estimates. Since this estimator does not explicitly allow the factor loading $\chi_n$ to vary by industry, it will only be unbiased if $\chi_n = \chi$ or if the effect of the time shocks on factor inputs and productivity is uncorrelated across industries. We will expand on this point in section 4.4.4 below. It also removes period-specific measurement error, something that the discussion of our data (in section 4.3) mentioned as a potentially important source of bias.

If the composite error term contains correlated time and industry effects, then the POLS, FE and TE estimators will all be biased. Combining the time and industry demeaning transformations,

[47] $\boldsymbol{I}_c$ is the identity matrix of dimension $c \times c$ and $\boldsymbol{i}_d$ is a $d \times 1$ vector of ones.

$\boldsymbol{M}_{2FE} = \boldsymbol{M}_{WT}\boldsymbol{M}_{WI}$, yields the two-way fixed effects (2FE) estimator, which now removes both industry and time-specific variation from the data. This estimator will therefore be unbiased in the presence of either correlated industry effects or time shocks, although any industry-specific productivity variation over time can still bias this estimator if it is also correlated with the factors of production.

Another candidate estimator, which requires very similar identifying conditions as the FE estimator, is the first-differenced estimator (FD), which differences the data before regressing production factors on output. In a panel data setup this estimator is defined as $\boldsymbol{M}_{FD} = \boldsymbol{I}_N \otimes \boldsymbol{M}_\Delta$ where $M_\Delta$ is the first differencing transformation matrix. Since the FD estimator differences away the between-industry variation in the data, the resulting estimates will suffer from the same decrease in estimator accuracy as the FE estimator if much of the regressor variation occurs across industries rather than time.

### 4.4.3 Parameter heterogeneity

The estimators discussed above all implicitly impose the restriction of slope homogeneity, but Burnside (1996) warns against exploiting this assumption when estimating cross-sector production functions. Using data for the U.S. manufacturing industry he finds that although this restriction seemingly improves estimator precision, the data strongly rejects its validity. We therefore introduce the mean group (MG) estimator – which explicitly allows for parameter heterogeneity – before considering how this affects the estimators discussed above.

The MG estimates are calculated by first obtaining estimates of $\boldsymbol{\beta}_n$ using OLS separately for each industry and then averaging these coefficient vectors across industries: $\widehat{\boldsymbol{\beta}}_*^{MG} = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{X}_n{}'\boldsymbol{M}_P\boldsymbol{X}_n)^{-1}\boldsymbol{X}_n{}'\boldsymbol{M}_P\boldsymbol{y}_n$. Pesaran and Smith (1995) note that, unlike the estimators discussed above, the MG estimator explicitly estimates the expected value of the production parameters $\boldsymbol{\beta}_*$ as the average of the individual $\boldsymbol{\beta}_n$s, which is conceptually more consistent with the heterogeneous production process outlined at the beginning of this section. The variance for this MG estimator can be expressed as $Var\left(\widehat{\boldsymbol{\beta}}_n^{MG}\right) = \frac{1}{N(N-1)}\sum_{n=1}^{N}(\widehat{\boldsymbol{\beta}}_n^{MG} - \widehat{\boldsymbol{\beta}}_*^{MG})^2$, and

the estimator bias as $\frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{X_n M_P X_n})^{-1}\boldsymbol{X_n}'\boldsymbol{M_P E}(\boldsymbol{u_n}|\boldsymbol{X_n})$, where the transformed error term now no longer contains the industry specific elements $\boldsymbol{x_{nt}b_n}$, $\eta_n$ or industry-specific measurement error. The MG industry-specific estimators only use data from single industries to estimate $\boldsymbol{\beta_n}$, so that correlation between the production factors and the industry fixed effects or coefficient deviations will no longer bias the individual industry estimates of $\boldsymbol{\beta_n}$ or, by implication, the average coefficient vector $\boldsymbol{\beta_*}$. On the other hand, the fact that only time series variation is exploited to estimate the production coefficients means that time dummies cannot be included in the regressions and the estimator will therefore be vulnerable to any correlation between global output shocks and the regressors. Discarding cross-sectional variation will also amplify any remaining endogeneity in the regressors in the same way that it does for the FE and FD estimators. If time series variation in the data is relatively unimportant compared to cross-industry variation, then this will leave us with very little confidence in the individual $\widehat{\boldsymbol{\beta}}_{\boldsymbol{n}}^{\boldsymbol{MG}}$'s, and only somewhat more in the cross-industry estimate of $\widehat{\boldsymbol{\beta}}_{*}^{\boldsymbol{MG}}$.

If the parameters are heterogeneous but distributed independently of the regressors, then heterogeneity merely adds a specific type of heteroscedasticity to the error terms and will not affect the consistency properties of our estimators. Puzzlingly, Coakley, Fuertes, & Smith (2006) supposedly consider the effect of parameter heterogeneity on various heterogeneous panel data estimators, but never go beyond this rather restrictive type of heterogeneity. Not surprisingly, they conclude that parameter heterogeneity on its own is of little consequence. However, a more interesting case occurs if we allow "sorting on gains" as in the CRC model discussed in section 2.3.3.1 of chapter 2. If a firm's investment and employment decisions are affected by the returns to human or physical capital (as suggested by profit maximising behaviour), then this will induce a positive correlation between the error term in equation [18] and the untransformed model regressors via $\boldsymbol{x_{nt}b_n}$, which violates the assumption of the (uncorrelated) random coefficient

model. In this case the POLS and TE models will be biased, but any estimator that allows for industry-specific intercepts (such as FE, 2FE or MG) will be unbiased[48].

Much of the heterogeneous panel data literature is concerned with modelling dynamic features of the data generating process and producing consistent estimates of the model parameters with data that is characterised by non-stationarity. As explained in section 4.2, this study is in large part motivated by the African production function literature that has been unable to find schooling effects of a comparable magnitude to those estimated from earnings regressions. The production functions in these studies generally ignore dynamics in the production process and non-stationarity, and this is also the approach taken in our econometric analysis below. However, we do report tests of the time series properties of the different estimator residuals, and also discuss a few estimators that will produce consistent estimates with non-stationary data.

### 4.4.4 Cross-sectional dependence

A recent strand of the heterogeneous panel data literature considers the effect of cross-sectional dependence, with a particular focus on global time shocks that affect individual units differently. In our application that would imply industry heterogeneity not only in terms of how inputs affect production, but also in how common latent time shocks affect production. We briefly consider how the resulting cross-sectional dependence will affect the estimators discussed above, before introducing four estimators that could provide more reliable estimates under such conditions.

It is instructive to consider the process that generated the industry factor inputs; we follow Coakley, Fuertes and Smith (2006) in assuming that $\boldsymbol{x}_{nt}^{*} = \boldsymbol{\theta}_n \tau_t + \boldsymbol{q}_{nt}$, where $E(\boldsymbol{\theta}_n \tau_t | \boldsymbol{q}_{nt}) = 0$.[49] The $\boldsymbol{\theta}_n$ vector now captures the effect of a global productivity shock $\tau_t$ on industry employment and investment

[48] However, if the production parameters are correlated to non-linear functions of employment or investment levels, such as the variance, then none of the estimators we consider will be unbiased.

[49] This process can also be conceived as a variation decomposition, in which case $\boldsymbol{\theta}_n$ is the regression coefficients $\frac{cov(\tau_t, \boldsymbol{x}_{n,t}^{*})}{var(\tau_t)}$, and $E(\boldsymbol{\theta}_n \tau_t \boldsymbol{q}_{nt}) = 0$ by design.

decisions, and $\boldsymbol{q}_{nt}$ represents the remaining determinants of these inputs. In this case the POLS, FE and MG estimators will only be unbiased if either $\chi_n = 0$ or $\boldsymbol{\theta}_n = \mathbf{0}$, that is if the universal time shocks either had no productivity effect, or had no effect on the input decisions of industries. The TE and 2FE estimators control for time effects without allowing for industry-specific factor loadings, so require somewhat weaker consistency conditions: $\chi_n = \bar{\chi}$ or $\boldsymbol{\theta}_n = \overline{\boldsymbol{\theta}}$ or $E\left((\chi_n - \bar{\chi})(\boldsymbol{\theta}_n - \overline{\boldsymbol{\theta}})\right) = \mathbf{0}$. These estimators therefore require one of three conditions to hold[50]: the effect of the time shocks on productivity should be constant across industries, its effect on industry inputs must be constant, or its effect on productivity should be uncorrelated to its effect on each of the inputs. In case the same industries are more responsive to global time shocks (like the business cycle) both in terms of output and employment or investment decisions, then none of the estimators considered so far will produce unbiased estimates of the model parameters.

If the POLS-FE-MG consistency conditions specified above are met, but the time shocks induce cross-sectional correlation in the model error terms, then Coakley et al. (2006) suggest combining the MG estimator with Zellner's (1962) seemingly unrelated regression (SUR) approach – estimating the $\boldsymbol{\beta}_n$'s for the different industries jointly, while allowing for non-zero cross-sectional covariance between the errors – in order to improve the efficiency of the estimates. This estimator will be no more robust to the various types of endogeneity that could bias the MG estimator, but the increased efficiency potentially afforded by the MG-SUR estimator could be important in small samples such as ours.

However, the MG-SUR estimator will still be biased in case $\chi_n \neq 0$ and $\boldsymbol{\theta}_n \neq \mathbf{0}$. Coakley et al. (2006) recommend using the demeaned mean group (DMG) estimator, which is calculated by estimating the MG estimator from period-demeaned data. The DMG-industry estimator will then be $\widehat{\boldsymbol{\beta}}_*^{DMG} = \frac{1}{N}\sum_{n=1}^{N}(\sum_{t=1}^{T}\tilde{\boldsymbol{x}}_{nt}'\tilde{\boldsymbol{x}}_{nt})^{-1}\sum_{t=1}^{T}\tilde{\boldsymbol{x}}_{n,t}'\tilde{y}_{nt}$, where $\tilde{\boldsymbol{x}}_{nt}$ is the period-demeaned $\boldsymbol{x}_{nt}$, and similarly for $\tilde{y}_{nt}$. This estimator will be unbiased under the same conditions as 2FE, but has the benefit of explicitly

[50] These conditions have not always been clearly stated in the heterogeneous panel literature (Coakley, Fuertes and Smith (2006) for example).

recognising the heterogeneity of the production parameters as part of the estimation procedure. Pesaran (2006) suggests using the common correlated effects mean group (CMG) estimator to provide consistent estimates of the model parameters under these conditions, which entails estimating the output equation separately for each industry using OLS but including output and input time means as regressors, i.e. estimating the following augmented regression equation:

$$y_{nt} = c_1\bar{y}_t + \boldsymbol{x}_{nt}\boldsymbol{\beta}_n + \bar{\boldsymbol{x}}_t\boldsymbol{c} + u_{nt}$$

where $\boldsymbol{c}' = [c_2, c_3, c_4, c_5]$. In the absence of measurement error, the average effect of the time shocks can be expressed as $\bar{\chi}\tau_t = \bar{y}_t - \bar{\boldsymbol{x}}_t\bar{\boldsymbol{\beta}} + v_{nt}$, where $v_{nt}$ is sampling noise that will tend to zero as $N \to \infty$. The inclusion of $\bar{\boldsymbol{x}}_t$ and $\bar{y}_t$ as regressors will tend to absorb the effect of the time shocks, assuming that the cross-sectional dimension of the dataset is large enough. When adding measurement error to the model, the inclusion of these two terms will also control for any time-specific measurement errors. Unfortunately, this only controls for the homogeneous component of the time shocks, and the heterogeneous component $(\chi_n - \bar{\chi})\tau_t$ can still induce bias in the estimates in the same way that it can for the DMG or 2FE models. However, the simulation results in Coakley et al. (2006) suggests that the CMG model performs better in smaller samples, and is more robust to the type of cross-sectional dependence that violates the identifying assumptions of the 2FE-DMG-CMG models. It bears noting that both the DMG and CMG estimators only exploit the within-industry variation in the data, and will therefore suffer the same estimator inaccuracy associated with the FE, 2FE or FD estimators.

Another estimator that will provide consistent estimates of $\boldsymbol{\beta}_*$ in the case of slope heterogeneity and cross-sectional dependence is the cross-section (CS) or between-groups estimator, which requires regressing the cross-sectional average of output on the cross-sectional average of the inputs, that is estimating $\bar{y}_n = \bar{\boldsymbol{x}}_n\boldsymbol{\beta}_* + \bar{u}_n$, where $\bar{u}_n = \bar{a} + \eta_n + \bar{\boldsymbol{x}}_n\boldsymbol{b}_n - \bar{\boldsymbol{e}}_n\boldsymbol{\beta}_n + \bar{\varepsilon}_n$. This estimator will be biased by correlated industry fixed effects, correlated random coefficients and industry-specific measurement error in the same way as the POLS or TE estimators. On the other hand, it is the only estimator considered so far that will

not be biased by time shocks (even if $E\left((\chi_n - \bar{\chi})(\boldsymbol{\theta}_n - \bar{\boldsymbol{\theta}})\right) = \mathbf{0}$) or two-way demeaned measurement error. Pesaran and Smith (1995) show that this estimator provides consistent estimates of $\boldsymbol{\beta}_*$ even if the error terms are distributed I(1). Since this estimator ignores all within-industry variation in the data, we would expect it to be fairly inaccurate in a dataset with as few cross-sectional observations as ours. The fact that it does not discard the between-industry variation means that it provides an interesting benchmark for our analysis, especially if we have reason to suspect that much of the informative variation in our data occurs along the cross-section dimension.

### 4.4.5 Measurement error

As discussed in section 4.2, cross-country studies of the human capital-growth nexus have sometimes ascribed the low estimated schooling effect to problems in obtaining comparable measures of educational attainment. However, few production function studies on African or South African data have attempted to consider the effects of measurement error, which may explain the surprisingly low estimated schooling effects. Section 4.3 discussed some of the reasons why the variables in our dataset may have been inaccurately measured, and speculated that the industry employment variable compiled from the StatsSA surveys is likely to be less accurately measured than the capital or schooling variables. Although worker schooling levels are compiled from the same set of inconsistently sampled surveys, some of the sampling problems will be mitigated in variables that are constructed as averages rather than the totals. If these measurement errors mainly derive from comparability problems across the different surveys, then no within-industry transformation will remove much of the resulting endogeneity. In fact, our discussion in section 4.3 suggests that $\boldsymbol{e}_{nt}$ will have a strong time-specific component, the elimination of which would require controlling for or averaging across time effects.

The classical errors-in-variables model suggests that the coefficient of the mismeasured regressor will suffer a downward attenuation bias proportional to the measurement error variance and inversely

proportional to the regressor variation that is orthogonal to the other regressors. In the case where a second, correctly measured regressor is positively correlated to the mismeasured regressor, Thiel (1961) shows that the bias in the coefficient of the correctly measured variable will be higher the stronger the correlation between the two explanatory variables, and of the same sign as the mismeasured variable's coefficient.

Given that all of our production factors are positively correlated (not shown) and presumably measured with some error, this suggests that the coefficients for each variable will be ambiguously affected: the own-error bias will exert the usual attenuation bias while the other-variable-errors will tend to upwardly bias the estimates. The structure of the regressor covariance and the suspected magnitudes of the different measurement errors suggest that the own-error downward bias will be relatively larger for the labour coefficient, whereas the capital coefficient – which is fairly consistently measured over time, and highly correlated to both mismeasured labour and education – may well be upwardly biased. The net effect on the education coefficients is difficult to gauge without deeper analysis. These biases will tend to be exacerbated when using an estimator that transforms away much of the regressor variation without removing the measurement error.

There are at least three methods that we can use to identify the education effect in the presence of measurement error: the Grilliches-Hausman estimator, fixing certain production parameters to their "known" values (as per Krueger and Lindahl's (2001) suggestion) and instrumental variable techniques. The first approach was pioneered by Griliches and Hausman (1986), who noted that if variables are measured with serially uncorrelated errors, then estimators based on longer differences will be less severely biased than those based on "short" differenced results. If measurement error is important in explaining why the different estimators are producing contrasting results, then we would expect the results to converge as we make use of longer differencing.

The second method of attempting to circumvent the effects of measurement errors is to follow Krueger and Lindahl (2001) in fixing the labour and capital parameter coefficients to their shares of national

income. We can check the robustness of the resulting human capital estimates by varying the capital and labour parameter values within a broad range of potentially realistic values.

Finally, if our dataset contains variables that are correlated to the true values of our factor inputs and uncorrelated to the measurement error, then IV estimators such as 2SLS can also recover consistent estimates of the production parameters. This estimator attempts to remove the endogeneity associated with measurement error by transforming away all of the endogenous variation in the regressors according to $\boldsymbol{M_{IV} = Z(Z'Z)^{-1}\,Z'}$ so that $\boldsymbol{\widehat{\beta}_{2SLS} = (X'ZZ'X)^{-1}X'ZZ'y}$. As with the estimators above, the downside of this approach is that the decrease in regressor variation sacrifices estimator accuracy and could amplify any remaining correlation between the regressors and unobserved productivity that was not successfully transformed away. The 2SLS estimator is known to be inaccurate in small samples, and this problem is exacerbated by using many or weak instruments (as discussed in Bound, Jaeger and Baker (1995), and Staiger and Stock (1997)). In this case the IV estimates will be biased in the direction of the POLS estimates and the actual size for hypothesis tests will be inflated relative to the stated test size

Where IVs are uncorrelated to the measurement error but correlated to industry fixed effects or other error components this IV transformation can be combined with other transformations in order to remove multiple sources of endogeneity. For example, the IV estimates on industry-demeaned data (which we will refer to as the IV-FE estimator) uses $\boldsymbol{M_{IV-FE} = M_{IV}M_{WI}}$ and will be unbiased in the presence of measurement error and correlated industry effects. Furthermore, this estimator can produce unbiased estimates of the model parameters where it would otherwise be affected by the simultaneity of output and factor inputs.

One popular IV candidate is the lagged values of the mismeasured production factors. The one-period lagged terms would be valid and informative if production factors were generated as an AR(1) processes – due, for example, to the gradual adjustment of employment levels that results from quadratic adjustment costs – and measurement error is serially uncorrelated. Other potential instruments include industry labour costs, unionisation rates and the cost of capital. Section 4.3 also alluded to an alternative measure of

employment which suffered substantial measurement problems but in a way that was unrelated to the problems that plagued our own industry employment variable.

## 4.5 Empirical results

We now estimate the industry-level production function using the data discussed in section 4.3 and the estimators introduced in section 4.4. Table 10 reports the coefficient estimates obtained from using the POLS, TE, FE, 2FE, RE and FD estimators (as discussed in sections 4.4.1 and 4.4.2). The POLS estimates indicate that the marginal return to employing better educated workers is very high on average: about 31% at the mean schooling years. This is even higher than the large earnings regression estimates obtained in chapters 2 and 3. The schooling coefficients also imply that returns are steeply decreasing in schooling years, which is surprising given the convex schooling-earnings profiles reported in most South African earnings regressions (as discussed in section 4.4.1). The capital and labour coefficients are slightly below 0.3 and above 0.6 respectively, which is close to their shares of total income and is similar to what has been found for other countries and for South Africa using different approaches or data in the past (for example, the studies cited in footnote 38). These coefficients suggest decreasing returns to scale (RTS) for the typical industry.

The second column in Table 10 reports the results from a TE regression, and demonstrates that controlling for homogenous time shocks has little impact on the coefficient estimates and only marginally increases the regression R-squared. This suggests that global productivity shocks (or time-specific measurement error) induce a negligible amount of coefficient bias in the POLS estimates, at least under the maintained hypothesis of parameter homogeneity.

Table 10: Estimates of production function coefficients, using various panel data estimators

| | POLS | TE | RE | FE | 2FE | FD |
|---|---|---|---|---|---|---|
| **Dependent variable** | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* | *D(Log output)* |
| Observations | 117 | 117 | 117 | 117 | 117 | 108 |
| R-squared | 0.957 | 0.958 | 0.984 | 0.991 | 0.992 | 0.114 |
| **Coefficient estimates** | | | | | | |
| Constant | 1.922*** | 1.623*** | 7.409*** | 5.895*** | 8.032*** | 0.019*** |
| | *(0.485)* | *(0.493)* | *(0.991)* | *(1.106)* | *(1.136)* | *(0.007)* |
| Log capital stock | 0.283*** | 0.285*** | 0.411*** | 0.86*** | 0.654*** | 0.831*** |
| | *(0.016)* | *(0.016)* | *(0.059)* | *(0.110)* | *(0.106)* | *(0.181)* |
| Log employment | 0.601*** | 0.605*** | 0.279*** | 0.039 | 0.062 | -0.025 |
| | *(0.018)* | *(0.018)* | *(0.075)* | *(0.084)* | *(0.072)* | *(0.077)* |
| Average education | 0.784*** | 0.798*** | 0.104 | -0.016 | -0.05 | 0.047 |
| | *(0.089)* | *(0.085)* | *(0.123)* | *(0.080)* | *(0.082)* | *(0.092)* |
| Average education^2 | -0.037*** | -0.038*** | 0.003 | 0.008* | 0.006 | -0.003 |
| | *(0.005)* | *(0.005)* | *(0.007)* | *(0.005)* | *(0.005)* | *(0.005)* |
| **Period dummies** | No | Yes | No | No | Yes | No |
| **Industry dummies** | No | No | No | Yes | Yes | No |
| **DW statistic** | 0.568 | 0.509 | 1.050 | 0.775 | 0.686 | 1.518 |
| **Implied RTS** | 0.885 | 0.890 | 0.690 | 0.898 | 0.716 | 0.806 |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

The estimators in Table 10 and the rest of this chapter use the quadratic education specification in the human capital link function. Although this aids the comparability with the results presented in previous chapters and is therefore our preferred specification, we would still like to compare these results to those obtained when using alternative human capital link functions. Table 21 in the appendix presents the TE estimates with different specifications for average worker schooling years. The first column, which includes no measure of human capital, produces labour and capital coefficients that are very close to labour and capital's shares of total income. This result is consistent with previous South African production function studies that also failed to control for human capital (as discussed in section 4.4.1). The next three columns show the coefficient estimates for functions in which schooling was included logarithmically, linearly and quadratically, and in all three cases the human capital terms are highly significant and their inclusion leads to a marked improvement in the explanatory power of the model.

The education coefficient from the linear specification implies a rate of return in excess of 32% for each additional year of education, which is much higher than the estimates obtained from international individual-level studies (Psacharopoulos (1994), Case and Yogo (1999)) but close to the upper end of the

range of such studies for South Africa (as surveyed in Keswell and Poswell (2004)). The coefficients of the logarithmic and quadratic specifications imply schooling returns at the mean schooling level of 32% and 31% respectively. The fact that the logarithmic function – which restricts the schooling effect to be a decreasing function of schooling year – is a better fit than the linear function suggests a concave schooling-production profile, and this is confirmed by the coefficients of the quadratic specification which also yields the highest R-squared.

Controlling for average worker schooling level does not substantially alter the coefficient on labour, although the physical capital coefficient is reduced. This suggests that the failure to control for the effect of human capital on production in previous South African studies may have over-estimated the effect of capital in the production process. However, it is worth bearing in mind that 80% of the variation in the average industry schooling variable occurs between 6.4 and 11.2 years of education: a much narrower range than what is observed for individuals. This variable may therefore be too crude to accurately identify the exact shape of the schooling-production profile for schooling years at sparser parts of this distribution.

Returning to the estimates in Table 10, the RE estimator allows for uncorrelated industry effects and exploits the error variance structure to provide more efficient production coefficient estimates. Surprisingly, the RE estimates (column 3) are markedly different from the POLS results, despite the fact that the two models require the same identifying assumptions for unbiasedness. The capital coefficient is now much closer to its share in total income, but the labour coefficient drops to below 0.3 and the education terms are both insignificant (although the implied returns are still sizable). Section 4.4.2 mentioned a few reasons why the industry fixed effects could be correlated with the production factors, in which case all three of the estimators discussed above would be biased. In contrast, the FE, 2FE and FD estimates could be unbiased even if industry effects are correlated to inputs. These estimates (reported in columns 4, 5 and 6) are very different from those of the POLS and TE estimators: the returns to education are now insignificant and more or less negligible in size. Furthermore, the capital coefficients are much

higher than capital's share in income or the estimates produced in other South African studies, whereas the labour coefficients fall to something close to zero.

We are therefore confronted with two very different pictures of what the South African production process looks like. According to the POLS model the total returns to education (including production externalities) are extremely high and concave which – if true – would offer empirical evidence supporting policies to allocate more resources to improving access to education, and early education years in particular. On the other hand, the FE model suggests that education itself contributes almost nothing to productivity: some industries use production technologies that make them more productive and perhaps require hiring more educated workers, but merely hiring more educated workers has little effect on output.

From a purely mechanical perspective, removing the cross-sectional variation from our data clearly reduces the roles assigned to labour and schooling in the production process. One interpretation of this result would be that high levels of human capital merely reflect the important industry fixed effects, so that these variables are falsely assigned an important role when not allowing for different industry intercepts. Alternatively, removing the cross-sectional variation in the data could make the FE estimator very sensitive to any remaining endogeneity that arises from measurement error, parameter heterogeneity or cross-sectional dependence. If this noise makes it impossible to identify the schooling effect, then the regression will tend to overstate the importance of its correlates, notably physical capital and the industry fixed effects. This same pattern has emerged for other African countries as well (Söderbom & Teal (2004), for example) where it is typical to interpret the FE estimator as producing more credible results. This despite the fact that the capital coefficients produced by such regressions (Bigsten, et al., 2000) are – as in the FE model in Table 10 – much higher than capital's share of income. Furthermore, the FD estimator is the only model in Table 10 not to suffer from residual autocorrelation, as represented by the Durbin-Watson (DW) statistics. Although the presence of autocorrelation should not affect the consistency of the coefficient estimates in our inherently static production function (Wooldridge, 2002, p. 174), the fact that our residuals appear to follow a random-walk process suggests that estimators that use the differencing

transformation – such as the Grilliches-Hausman estimator (Table 12), the FD estimators with parameter restrictions (Table 13) or the IV-FD estimators (Table 25) – may be more appropriate to model South African production in the presence of fixed effects.

Before turning to estimators that yield consistent estimates when industries are heterogeneous or inputs are measured with error, it is worth briefly considering the composition of the regressor variation in our data. As discussed in section 4.4, the FE estimator may be more susceptible to remaining sources of endogeneity *if* the within-groups transformation leads to a substantial reduction in regressor variation. Unless this is the case, there is little reason to think that the POLS results would be more robust than the FE model. Table 11 reports the results from decomposing the variation into cross-sectional, time and the remaining two-way variation. All of the production factors can be observed to have a very high proportion of industry-specific variation (99% for capital, 96% for labour and 94.5% for education), so that any estimator that transforms away the cross-industry variation is bound to suffer a large decrease in accuracy. If this transformation does not rid the data of all the covariance between the regressors and the error, then the POLS may have a lower relative bias than the FE model. In order to further explore this hypothesis, we now turn to methods that could provide unbiased estimates when endogeneity is not completely transformed away by industry- or time-demeaning.

Table 11: Variance decomposition for factors of production

| | **Industry-specific** | **Period-specific** | **Two-way demeaned** |
|---|---|---|---|
| **Log capital** | 99.0% | 0.4% | 0.6% |
| **Log labour** | 96.0% | 0.9% | 3.1% |
| **Education** | 94.5% | 3.8% | 1.7% |

Source: Author's calculations from October Household Surveys 1995-99, Labour Force Surveys (March & September rounds) 2000-2007 (StatsSA: various years) and Quarterly Bulletin data 1995-2007 (South African Reserve Bank, various years)

### 4.5.1 Measurement error

We next investigate whether measurement error could be biasing our estimators by using the three approaches mentioned in section 4.4.5. Firstly, Griliches and Hausman (1986) showed that if the model errors are serially uncorrelated, then the measurement error bias in the FD estimates should diminish as

longer differences are used. Columns 2, 3 and 4 in Table 11 compare the results from the FD estimator using 3 year, 5 year and 12 year lags, to those obtained from one year differenced data in column 1 (replicated from the FD coefficients in Table 10).

Table 12: Panel data estimates of production function coefficients

| | **FD** | **FD** | **FD** | **FD** |
|---|---|---|---|---|
| Dependent variable | ***$D^i$(Log output)*** | ***$D^i$(Log output)*** | ***$D^i$(Log output)*** | ***$D^i$(Log output)*** |
| Differencing period | **1 year (i = 1)** | **3 years (i = 3)** | **5 years (i = 5)** | **12 years (i = 12)** |
| Observations | 108 | 90 | 72 | 9 |
| R-squared | 0.197 | 0.422 | 0.421 | 0.280 |
| Constant | 0.066** | 0.058** | 0.112** | -0.029 |
| | *(0.033)* | *(0.028)* | *(0.052)* | *(0.306)* |
| $D^i$(Log capital stock) | 0.844*** | 0.725*** | 0.604*** | 0.414 |
| | *(0.184)* | *(0.130)* | *(0.117)* | *(0.459)* |
| $D^i$(Log employment) | -0.012 | 0.037 | 0.098 | 0.398 |
| | *(0.073)* | *(0.049)* | *(0.070)* | *(0.364)* |
| $D^i$(Average education) | 0.031 | -0.095 | -0.052 | 0.533 |
| | *(0.080)* | *(0.107)* | *(0.079)* | *(0.403)* |
| $D^i$(Average education^2) | -0.002 | 0.007 | 0.007 | -0.011 |
| | *(0.004)* | *(0.006)* | *(0.005)* | *(0.022)* |
| Period dummies | Yes | Yes | Yes | Yes |
| Industry dummies | No | No | No | No |
| DW statistic | 1.502 | 0.525 | 0.427 | . |
| Implied RTS | 0.832 | 0.762 | 0.702 | 0.812 |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

The capital coefficient can be seen to decrease as the data is differenced over longer lags, whereas the coefficient on labour increases. Using 12 year lags consumes almost all of our degrees of freedom, so that very little confidence can be placed in the precision of these point estimates, but it is still interesting to note how close the results from this regression – including the education coefficients – are to those obtained from the POLS estimator. This is supportive of the notion that measurement error is a source of bias in the one-period FD (and by implication the FE) estimates.

The second approach to estimating the human capital coefficients in the presence of measurement error (as discussed in section 4.4.5) is to fix the coefficient values of labour and capital to reasonable values. This will prevent the schooling coefficients from being biased by the correlation between education and other

mismeasured inputs. Table 13 reports the FE and FD regression estimates from production functions in which the employment coefficient has been fixed to values of 0.5, 0.6 and 0.7.

Table 13: FE and FD production function regressions with fixed employment coefficient values

| | **2FE** | **2FE** | **2FE** | **FD** | **FD** | **FD** |
|---|---|---|---|---|---|---|
| **Dependent variable** | *Log output* | *Log output* | *Log output* | *D(Log output)* | *D(Log output)* | *D(Log output)* |
| Observations | 117 | 117 | 117 | 108 | 108 | 108 |
| **Coefficient estimates** | | | | | | |
| Constant | 5.504*** | 4.926*** | 4.349*** | 0.141* | 0.156* | 0.171* |
| | *(0.085)* | *(0.093)* | *(0.102)* | *(0.080)* | *(0.090)* | *(0.100)* |
| Log capital stock | 0.281*** | 0.196* | 0.111 | 0.38 | 0.289 | 0.198 |
| | *(0.106)* | *(0.115)* | *(0.126)* | *(0.327)* | *(0.381)* | *(0.436)* |
| Log employment | 0.5 | 0.6 | 0.7 | 0.5 | 0.6 | 0.7 |
| | . | . | . | . | . | . |
| Average education | 0.139 | 0.183 | 0.226 | 0.179 | 0.208 | 0.237 |
| | *(0.127)* | *(0.139)* | *(0.152)* | *(0.143)* | *(0.168)* | *(0.193)* |
| Average education^2 | -0.001 | -0.003 | -0.004 | -0.01 | -0.012 | -0.014 |
| | *(0.007)* | *(0.008)* | *(0.009)* | *(0.008)* | *(0.009)* | *(0.010)* |
| **Period dummies** | Yes | Yes | Yes | Yes | Yes | Yes |
| **Industry dummies** | Yes | Yes | Yes | No | No | No |
| **DW test statistic** | 1.196 | 1.275 | 1.333 | 1.625 | 1.658 | 1.682 |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

In case of both the FE and FD estimators, fixing this coefficient has the effect of producing physical and schooling coefficients that are rather similar to the POLS estimates in Table 10. The human capital coefficients indicate substantial returns and concavity, although the coefficients themselves are not significant. A higher employment coefficient is associated with a lower capital coefficient and schooling returns that are more concave, but the underlying picture that emerges from these regressions is fairly robust to the exact value of the employment coefficient. This evidence supports the notion that the POLS estimates are relatively accurate compared to the FE and FD estimates that are biased by the measurement error in the employment variable.

Finally, Table 23-Table 25 in the appendix report the results of a series of IV regressions aimed at addressing the problems of measurement error (and perhaps also simultaneity).[51] Table 23 presents the results of five different IV-TE regressions (including time, but not industry dummies in the regression). Columns 1, 2 and 3 in Table 23 all treat industry employment as the only endogenous regressor, since this variable is considered to be the most severely affected by measurement error. The first regression instruments for our chosen employment measure using the alternative measure of industry employment (discussed in section 4.3), whereas the second regression adds a one-period lag of this alternative measure as well as the one- and two-period lags of our own employment series to the instrument set. The third column exploits the one- and two-period lags of the average industry wage rate as IVs. Compared to the TE estimates in Table 10, instrumenting has little effect on either the capital or labour coefficient (both are increased marginally). Furthermore, in all three regressions the schooling effect remains large and concave, but somewhat lower at the mean schooling level than suggested by the TE estimator. The two remaining regressions also consider the other production factors as endogenous and use as instruments the one-period lag of all the inputs (column 4) as well the average wage rate, the bond yield and the one-period lag of the industry unionisation rate (column 5). The results from these IV-TE regressions are virtually indistinguishable from the TE regressions that took all the regressors as exogenous. The Hansen J-test and F-test statistics[52] indicate that the instrument sets are far from weak in all cases, and that there is little evidence that our instruments are themselves endogenous.

The IV regressions in Table 24 add industry dummies to the IV-TE regressions in Table 23 in order to control for correlated industry productivity differentials, and hence represent the IV-2FE estimates discussed in section 4.4.5. The implied data transformation now removes any variation from the data that is industry or period-specific, or that does not covary with the instrumental variables. This combination means that any weak instrument problems or remaining endogeneity will be exacerbated, something that is

[51] In all cases the limited information maximum likelihood estimation technique was used rather than the 2SLS discussed in section 4.4.5, since Stock and Yogo (2005) find that the results for this estimator are less sensitive to the effects of weak instruments.

[52] In the regressions in columns 4 and 5, and all subsequent regressions where more than one endogenous variables are instrumented for, the F-statistic refers to the Cragg-Donald F-statistic discussed in footnote 21.

reflected in the low F-tests and inflated standard errors. On the other hand, this estimator can provide consistent estimates of the production parameters even where the POLS model was biased by time shocks, industry effects and measurement error. In general the capital, labour and human capital coefficients are relatively sensitive to the exact choice of instruments. However, the results also show that instrumenting for the regressors can produce large schooling effects, even when allowing for correlated industry effects.

Finally, Table 25 also reports the results of the IV-FD estimators corresponding to the endogenous variables and instruments sets in Table 23 and Table 24. This implied data transformation is very similar to that of the IV-2FE estimator, and we therefore expect the resulting coefficients to share the same strengths and weakness as the IV-2FE estimates. Again, this is confirmed by the low F-statistics and few significant coefficients. Much the same as for the IV-2FE model, the IV-FD estimator produces capital and labour coefficients that are often inaccurate and volatile. Although the schooling coefficients are generally insignificant, the implied returns are substantial (although lower than estimated by the POLS model) and concave. In general, the results from our instrumental variable regressions therefore support the notion that the POLS results actually offer a relatively accurate reflection of the schooling effect in the production process, but that this effect is hidden by the problems in measuring the industry input levels (particularly employment).

### 4.5.2 Parameter heterogeneity

The preceding results were produced with estimators that were developed under the assumption of parameter homogeneity, but the discussion in section 4.4.3 explained how randomness in the production parameters can bias these estimates. In order to investigate the validity of the constant coefficient assumption, Table 22 in the appendix reports the coefficient estimates obtained by running OLS on each industry separately (the MG-industry estimators) as well as the average and standard deviations of these industry coefficients (the MG estimator). The MG-industry results could scarcely provide less support for the notion of identical production parameters across industries. However, given the fact that these

estimates are likely to be very inaccurate – due to the short time period and how little variation in the data occurs along the time series dimension – and sensitive to measurement error and common correlated time shocks, it is difficult to conclude with any confidence that the cross-industry variation in the production coefficients reflect true parameter heterogeneity. The fact that the industry point estimates of the labour and capital coefficients lie outside the unit interval as often as not further detracts from our faith in these results.

The MG estimator averages away some of the sampling error and is hence likely to be more accurate than the industry-specific estimates. However, the final column in Table 22 demonstrates the cost of attempting to estimate so many different parameters from so little variation in the data: although the production coefficients are often sizable, none of them differ significantly from zero. Although this is disconcerting, it is in line with the experience of Burnside (1996), who found that imposing slope homogeneity was without grounds and artificially increased estimate precision. Compared to the FE estimator, allowing for parameter heterogeneity slightly decreases the point estimate on the physical capital coefficient while the coefficients on labour and human capital both remain small and insignificant. The MG coefficient estimates more closely resemble the FE than the POLS estimates, though the large standard errors also demonstrate how ambitious it is to estimate the average of potentially heterogeneous coefficients using only within-industry variation in such a small data set.

Like the FE, 2FE and FD estimators, the MG estimator is unbiased in the presence of correlated industry fixed effects, correlated industry production coefficients or industry-specific measurement error but will be similarly vulnerable to other types of measurement error or common correlated time shocks due to the relatively small amount of variation in the data that is used to calculate the estimates. Additionally, the MG-estimator cannot control for pure time effects and will therefore be biased by time shocks that are correlated with production factors, even where these effects impact homogeneously across all industries. Nonetheless, there are two conclusions that can be drawn from the MG estimates in Table 22. Firstly, allowing for parameter heterogeneity but not cross-sectional dependence does not substantially alter the

estimates produced by the FE estimator, and would seem to support the notion that POLS tends to over-estimate the effect of human capital in production because of its very restrictive identifying assumptions. Secondly, the results also highlight the importance of viewing the coefficient estimates obtained from POLS or FE as providing a summary statistic of the distribution of the industry (or firm)-specific production parameters, rather than a constant parameter that is equally valid in every industry.

### 4.5.3 Cross-sectional dependence

Table 14 compares the results from the 2FE and MG estimators – which did not allow for cross-sectional dependence – to those produced with the MG-SUR, DMG, CMG and CS estimators (as discussed in section 4.4.4). The MG-SUR estimates are broadly similar to those of the MG and 2FE estimators. Allowing for cross-industry correlation in the error terms – while still assuming that these shocks are uncorrelated to the regressors – has the effect of slightly increasing all of the production function coefficients, including the effect of schooling. However, these estimates still suffer from the large standard errors that also characterised the MG estimates. The DMG estimator effectively controls for homogenous global time shocks, and is therefore more robust than either the MG or the MG-SUR estimators. The production function coefficient estimates also have smaller standard errors, although these are still much larger than those of the "homogeneous parameter" estimators. Its high physical and low human capital estimates are consistent with the FE interpretation of the data, although our discussion in section 4.4.4 warned that this model will also be inconsistent unless the time shocks have a homogenous effect across industries, its effect on industry inputs is constant, or its effect on productivity is uncorrelated to its effect on each of the inputs. Although none of the estimators we consider will be consistent unless at least one of the three conditions above are satisfied, the CMG model has been shown to perform better in small samples, and to be less sensitive to the violation of these conditions. In terms of evaluating whether cross-sectional dependence in the data can explain the low education returns reported in production function studies for Africa, we are therefore particularly interested in the results from this estimator. The

coefficients (Table 14, column 5) show that the labour coefficient is still lower than what we would expect, but that the schooling coefficients are virtually the same as for the POLS estimator[53].

Finally, the CS estimator will also be unbiased under fairly general conditions, including parameter heterogeneity and I(1) errors, although the averaging transformation sacrifices a lot of efficiency and degrees of freedom in order to achieve this. These estimates, reported in column 6 of Table 14 closely resemble those of the POLS estimator. This may be because, apart from the POLS and TE estimators it is one of the few estimators that also exploits the (very informative) between-industry variation in the data.

Table 14: Various heterogeneous parameter panel data estimates of production function coefficients

| | **FE2** | **MG** | **MG-SUR** | **DMG** | **CMG** | **CS** |
|---|---|---|---|---|---|---|
| **Dependent variable** | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* |
| Observations | 117 | 117 | 117 | 117 | 117 | 9 |
| R-squared | 0.992 | 0.998 | 0.996 | 0.998 | 0.9998 | 0.964 |
| **Coefficient estimates** | | | | | | |
| Constant | 8.032*** | 9.738 | 7.436 | -0.056 | -5.028 | 0.695 |
| | *(1.077)* | *(9.604)* | *(8.859)* | *(0.497)* | *(4.079)* | *(1.994)* |
| Log capital stock | 0.654*** | 0.474 | 0.609 | 0.867 | 0.498 | 0.297** |
| | *(0.098)* | *(0.629)* | *(0.516)* | *(0.528)* | *(0.545)* | *(0.073)* |
| Log employment | 0.062 | 0.059 | 0.069 | 0.127* | 0.128* | 0.628*** |
| | *(0.055)* | *(0.075)* | *(0.076)* | *(0.066)* | *(0.068)* | *(0.073)* |
| Average education | -0.05 | -0.044 | 0.069 | -0.009 | 0.824* | 0.97** |
| | *(0.078)* | *(0.733)* | *(0.745)* | *(0.280)* | *(0.471)* | *(0.297)* |
| Average education^2 | 0.006 | 0.006 | 0.001 | 0.000 | -0.045* | -0.048* |
| | *(0.004)* | *(0.037)* | *(0.036)* | *(0.016)* | *(0.023)* | *(0.018)* |
| **DW statistic** | 0.511 | 2.102 | 1.876 | 1.627 | 1.964 | . |
| **Implied RTS** | 0.716 | 0.533 | 0.678 | 0.993 | 0.626 | 0.925 |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Like most other firm or industry-level datasets available for Africa, ours does not have the dimensions required to conclusively determine whether cross-sectional dependence has downwardly biased the estimated education effect in other studies. However, the fact that the CMG and CS education estimates are so similar to those obtained from the POLS and measurement error models certainly suggest that

[53] The R-squared statistics reported are the coefficients of determination from the respective regressions. However, as discussed in section 4.4, these regressions are estimated with independent variables – such as time and industry dummies or the average output level per period – that can remove the coefficient bias from our parameters of interest but do not contribute to our understanding of the underlying production process. This artificially inflates the R-squared statistics and means they cannot be interpreted in the usual way.

cross-sectional dependence may have biased – and potentially severely biased – the FE, FD and 2FE results reported in other studies.

## 4.6 Conclusion

This paper has studied the way in which nine broadly defined South African industries used human capital in the production process during the twelve years following political transition. It does so by constructing a dataset from SARB Quarterly Bulletins and StatsSA household surveys. This novel dataset allows us to explore issues that previous studies on African and particularly South African data were unable to do. It also investigated the results from a wide range of estimators, in order to investigate the effect of measurement error and cross-industry differences in production technology on the predicted schooling effects.

The POLS results suggests labour and physical capital coefficients of approximately 0.6 and 0.3 respectively, and that the returns to education were high and concave. The FE and FD estimates, on the other hand, found that the returns to both labour and schooling were much lower, and the effect of physical capital somewhat higher. This is consistent with the pattern observed in other African production function studies. Our analysis of the data shows that the industry-demeaning transformation discards between 95 and 99% of the variation in the regressors, which makes the FE and FD estimates much less accurate and highly sensitive to any biases (including measurement error bias or differences in production technologies across industries) that are not successfully transformed away. The POLS estimates, on the other hand, may be more robust to the presence of these types of biases.

Using longer differences in the FD estimator or fixing the employment coefficients to reasonable values, both suggest that the FE and FD estimates suffered from large measurement error bias, and that the POLS estimates provide a more reliable account of the production process in the typical industry. Using instrumental variables to explicitly account for measurement error also produced results that generally,

although not uniformly, supported the original POLS estimates rather than the FE and FD estimators. The fact that the IV-FE estimator assigns a smaller role to schooling in the production process than POLS suggests that average worker education levels are correlated with industry productivity, which could complicate efforts to simultaneously identify both of these two effects.

Although the FE and FD estimators allow industries to produce with different levels of productivity, they still restrict output to respond identically to changes in the inputs or global time shocks. Using estimators that exploit less restrictive identifying assumptions, we show that accounting for parameter heterogeneity and (particularly) cross-sectional dependence also produces estimates that more closely resemble the POLS coefficients estimates that those from the FE and FD estimators. The results also highlight the importance of viewing the coefficient estimates obtained from POLS or FE as providing a summary statistic of the distribution of the industry specific production parameters, rather than a constant parameter that is equally valid in every industry.

# 5. Conclusion

This study estimated the effect of schooling on individual employment, earnings and productivity using three very different approaches and a whole range of estimators. We explored whether the diverging results produced by instrumental variable and dynamic programming earnings functions, and production function studies can be reconciled when addressing a number of common econometric shortcomings in these approaches. Particularly, we investigated how the presence of heterogeneity, endogeneity, sample selection and non-linearities in the schooling-earnings profile affects the estimation of the schooling returns.

Chapter 2 used various instrumental variable techniques to estimate the causal effect of schooling on the earnings of black South African men in the period directly following political transition. Our model of earnings allows for endogenous schooling, individual heterogeneity in the returns to schooling, non-linearities in the schooling-earnings profile, and a selective process through which individuals enter wage employment. By using estimators that explicitly or implicitly incorporate all four of these issues, this model is more general than has been estimated for South Africa (and any other country, to our knowledge) before.

We take two different approaches to identifying the shape of the schooling-earnings function: comparing the LATEs of different IVs and estimating the parameters from a more restrictive Becker-Card model. Our empirical analysis demonstrated that schooling returns amongst black South African men are high on average, probably between 7% and 12% for those at the mean level of schooling attainment. The convexity of the schooling-earnings profile that has also been found in numerous other studies is relatively robust to most of the estimators that we consider. However, there is some evidence that ordinary least squares and two-stage least squares tend to overstate the degree of convexity in the schooling-earnings profile. This is particularly true when correcting for sample selection bias or allowing for non-linear relationships between schooling and market unobservables. Our results also show that addressing the

endogeneity of schooling in an employment probit produces a schooling-employment profile which is slightly less convex than what is suggested under the exogenous schooling assumption.

Chapter 3 estimated a dynamic programming model and showed that observed schooling decisions were consistent with a convex schooling-earnings profile in an environment in which individuals make rational, unconstrained, dynamically optimising decisions. The model estimates indicates that the schooling-earnings profile is convex, although the degree of convexity is slightly less than suggested by previous studies that use exogenous schooling models. This result is very different from that produced by similar models for the US , which usually found much lower and more convex returns to schooling than suggested by either OLS or IV estimates.

The schooling-employment profile is also convex, but much less so than suggested by OLS or MLE models that do not control for the endogeneity of schooling or the forward-looking nature of schooling decisions. Most agents in our model choose to complete primary and early secondary schooling years despite the relatively low returns associated with these investments, because this offers them the option of proceeding on to the high yielding final years of secondary school. However, the exorbitant cost of enrolling in tertiary education means that most agents in our model find it optimal to drop out after completing secondary school. The pattern of unobserved abilities is more complex than the usual notion of high ability individuals who earn more, get more schooling and are more likely to be employed, although this trend can be observed for the bulk of individuals who have neither very low nor very high academic abilities.

Chapter 4 estimated production functions for nine South African industries on a dataset constructed from multiple sources. POLS estimates indicate that the returns to schooling are high and concave. Although the schooling effect disappears when the FE or FD estimators are used instead, allowing for either measurement error or parameter heterogeneity and cross-sectional dependence suggests that the POLS estimates produce a fairly accurate description of the South African production process. The results also show that better educated workers are more inclined to work in more productive industries, which

explains why other studies have often struggled to identify the human capital effect while controlling for industry fixed effects.

There are a number of conclusions that can be drawn from this investigation. The empirical evidence overwhelmingly confirms that the returns to investing in schooling in the South African labour market are high. Results from dynamic programming structural models have been used to argue that the returns in the US market are much lower than IV and OLS methods would lead us to believe. The same conclusion seems to follow from the low schooling coefficients usually obtained from African production function studies. However, chapters 3 and 4 argue that once we properly account for some of the econometric problems that can bias these estimates the results from both models are consistent with high schooling returns.

The results regarding the exact shape of the schooling-earnings profile are sensitive to which technique is used, but throughout the analysis a picture started to unfold which suggested that the highly convex schooling-earnings profile obtained using the OLS or 2SLS estimates were prone to over-estimating the degree of convexity. The alternative IV techniques employed in chapter 2 generally produced profiles that were considerably less convex, especially when explicitly modeling schooling as an ordered probit or allowing the schooling error to be correlated to market ability in a non-linear way. The estimated earnings profile from the dynamic programming model in chapter 3 was also less convex, although only marginally so. The production function approach produced schooling-earnings profiles that were concave, although the results may not be strictly comparable since the employment variable in chapter 4 is constructed using workers from all races and genders, whereas the analysis in chapters 2 and 3 only focused on black males.

The IV and DCDP models from chapters 2 and 3 also estimated employment functions which showed that the schooling-employment profile was convex. Controlling for the endogeneity of schooling in this employment process decreased the degree of convexity in both models, although more so for the latter than the former.

The various models also allow us to draw some conclusions about the interaction between the various model unobservables. The IV models suggested that schooling ability is positively correlated to the ability to transform schooling into earnings. There is also some evidence that this relationship may be non-linear. The DCDP explicitly estimated the various unobserved ability coefficients, and found that although we do not find the standard pattern according to which high ability individuals earn more, obtain more schooling and are more likely to be employed, this trend can be observed for the bulk of individuals who have neither very low nor very high academic abilities. The production function estimates indicate that high ability individuals are more likely to be employed in more productive industries.

# 6. Bibliography


Aguirregabiria, V., & Mira, P. (2010, May). Dynamic discrete choice structural models: A survey. *Journal of Econometrics, 156*(1), 38-67.

Altman, M. (2008). Revisiting South African Employment Trends in the 1990s. *South African Journal of Economics, 76*(S2), S126-S147.

Angrist, J. D., & Krueger, A. B. (1991, November). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics, 106*(4), 979-1014.

Appleton, S., & Balihuta, A. (1996). Education and agricultural productivity: Evidence from Uganda. *Journal of International Development, 8*(3), 415 - 444.

Appleton, S., Hoddinott, J., & MacKinnon, J. (1996). Education and Health in Sub-Saharan Africa. *Journal of International Development, 8*(3), 307-339.

Arabsheibani, G. R., & Mussurov, A. (2007, April). Returns to schooling in Kazakhstan: OLS and instrumental variables approach. *Economics of Transition, 15*(2), 341–364.

Arora, V., & Bhundia, A. (2003, September). Potential Output and Total Factor Productivity Growth in Post-Apartheid South Africa. *IMF Working Paper, WP/03/178*. Washington, DC: IMF.

Ashenfelter, O., Harmon, C., & Oosterbeek, H. (1999, November). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics, 6*(4), 453-470.

Baloyi, R. (2004). The role of the state in the establishment of a culture of learning and teaching in South Africa (1910-2004). *Unpublished doctoral dissertation*. Pretoria: UNISA.

Banerjee, A. V., & Duflo, E. (2005). Growth Theory through the Lens of Development Economics. In P. Aghion, & S. N. Durlauf (Eds.), *Handbook of Economic Growth* (Vol. 1, pp. 473-552). Amsterdam: Elsevier.

Banerjee, A., Galiani, S. F., Levinsohn, J., McLaren, Z., & Woolard, I. (2007). Why has unemployment risen in the New South Africa? *NBER Working Paper, No. 13167*. Cambridge, MA: NBER.

Baptist, S., & Teal, F. (2008). Why do South Korean firms produce so much more output per worker than Ghanaian ones? *CSAE Working Paper*. Oxford: Centre for the Study of African Economies.

Beaudry, P., & DiNardo, J. (1991, August). The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data. *The Journal of Political Economy, 99*(4), 665-688.

Becker, G. S. (1964). *Human Capital*. New York: Columbia University Press.

Becker, G. S. (1967). *Human Capital and the Personal Distribution of Income: An Analitical Approach*. Ann Arbor: University of Michigan, Institute of Public Administration.

Becker, G. S., & Chiswick, B. R. (1966, March 1). Education and the Distribution of Earnings. *The American Economic Review, 56*(1/2), 358-369.

Behar, A. (2010). Would More Skills Raise Demand for Those Who Do Not Get Them? Evidence from South African Manufacturing. *Journal of African Economies, 19*(4), 496-535.

Belzil, C. (2007, July ). The return to schooling in structural dynamic models: a survey. *European Economic Review, 51*(5), 1059-1105.

Belzil, C., & Hansen, J. (2002, September). Unobserved Ability and the Return to Schooling. *Econometrica, 70*(5), 2075-2091.

Belzil, C., & Hansen, J. (2007, October). A structural analysis of the correlated random coefficient wage regression model. *Journal of Econometrics, 140*(2), 827-848.

Benhabib, J., & Spiegel, M. M. (1994, October). The role of human capital in economic development: evidence from aggregate cross-country data. *Journal of Monetary Economics, 34*(2), 143-173.

Bennell, P. (1996). Rates of Return to Education: Does the Conventional Pattern Prevail in Sub-Saharan Africa? *World Development, 24*(1), 183-199.

Bhorat, H., & Lundall, P. (2004, December). Employment, Wages And Skills Development: Firm-Specific Effects - Evidence From A Firm Survey In South Africa. *The South African Journal of Economics, 72*(5), 1023-1056.

Bigsten, A., Isaksson, A., Söderbom, M. M., Collier, P., Zeufack, A., Dercon, S., et al. (2000, July). Rates of Return on Physical and Human Capital in Africa's Manufacturing Sector. *Economic Development and Cultural Change, 48*(4), 801-27.

Bils, M., & Klenow, P. J. (2000, December). Does Schooling Cause Growth? *American Economic Review, 90*(5), 1160-1183.

Binelli, C. (2008). Returns to Education and Increasing Wage Inequality in Latin America. *Working Paper Series 30-08*. Rimini Centre for Economic Analysis.

Bonga-bonga, L. (2009, June). The South African aggregate production function: estimation of the constant elasticity of substitution function. *South African Journal of Economics, 77*(2), 332–349.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995, Jun). Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American Statistical Association, 90*(430), 443-450.

Branson, N. (2009). *Re-weighting the OHS and LFS National household Survey Data to create a consistent series over time: A Cross Entropy Estimation Approach.* Cape Town: SALDRU / DataFirst Working Paper No. 38.

Bunting, I. (2002). The higher education landscape under apartheid. In N. Cloete, R. Fehnel, P. Maassen, T. Moja, H. Perold, & T. Gibbon (Eds.), *Transformation in Higher Education: Global Pressures and Local Realities in South Africa* (pp. 35-52). Cape Town: Centre for Higher Education Transformation.

Burger, R. P., & Yu, D. (2006). Wage trends in post-apartheid South Africa: Constructing an earnings series for South Africa from household survey data. *Labour Market Frontiers, 8*, 1-8.

Burnside, C. (1996). Production function regressions, returns to scale, and externalities. *Joumal of Monetary Economics, 37*, 177-201.

Card, D. (1999). The causal effect of education on earnings. In O. Ashenfelter, & D. Card, *Handbook of Labor Economics* (Vol. 3(1), pp. 1801-1863). Amsterdam: Elsevier Science.

Card, D. (2001, September). Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica, 69*(5), 1127-1160 .

Carnoy, M. (1995). Rates of Return to Education. In M. Carnoy, *International Encyclopaedia of the Economics of Education* (2nd ed., pp. 364-369). Oxford: Pergamon.

Casale, D., & Posel, D. (2002, March ). The continued feminisation of the labour force in South Africa: An analysis of recent data and trends. *South African Journal of Economics, 70*(1), 156–184.

Casale, D., Muller, C., & Posel, D. (2004). `Two million net new jobs': A reconsideration of the rise in employment in South Africa, 1995-2004. *South African Journal of Economics, 72*(5), 1-20.

Case, A., & Yogo, M. (1999). Does School Quality Matter? Returns to Education and the Characteristics of Schools in South Africa. *NBER Working Paper No. 7399*. Cambridge: NBER.

Chen, G., & Hamori, S. (2009, June). Economic returns to schooling in urban China: OLS and the instrumental variables approach. *China Economic Review, 20*(2), 143-152.

Christie, P. (1985). *The right to learn: the struggle for education in South Africa.* Braamfontein: Ravan Press.

Coakley, J., Fuertes, A.-M., & Smith, R. P. (2006). Unobserved heterogeneity in panel time series models. *Computational Statistics & Data Analysis, 50*, 2361-2380.

Cohn, E., & Addison, J. T. (1998, December ). The Economic Returns to Lifelong Learning in OECD Countries. *Education Economics, 6*(3), 253 - 307.

Collier, P., & Gunning, J. W. (1999, March). Explaining African Economic Performance. *Journal of Economic Literature, 37*(1), 64-111.

Corona, A., Marchesi, M., Martini, C., & Ridella, S. (1987). Minimizing Multimodal Functions of Continuous Variables with the 'Simulated Annealing' Algorithm. *ACM Transactions on Mathematical Software, 13*, 262-280.

Cragg, J. G., & Donald, S. G. (1993). Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory, 9*, 222-240.

De la Fuente, A., & Doménech, R. (2006). Human Capital in Growth Regressions: How Much Difference Does Data Quality Make? *Journal of the European Economic Association, 4*(1), 1-36.

Deaton, A. S. (1985). Panel Data from Time Series of Cross-Sections. *Journal of Econometrics, 30*, 109-126.

Deaton, A. S. (2009). *Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development.* Cambridge, Mass.: NBER.

Denison, E. F. (1964). Measuring the contribution of education (and the residual) to economic growth. In E. F. Denison, *The residual factor and economic growth.* Paris: OECD.

Department of Education. (1998). Age requirements for admission to an ordinary public government school. *Government Gazette no. 2433 of 1998*.

Deschenes, O. (2004). Unobserved Ability, Comparative Advantage, and the Rising Return to Education in the United States 1979-2002. *Unpublished paper*.

Duflo, E. (2001, September). Schooling And Labor Market Consequences Of School Construction In Indonesia: Evidence From An Unusual Policy Experiment. *American Economic Review, 91*, 795-813.

Eberhardt, M., & Teal, F. (2008). Modeling Technology and Technological Change in Manufacturing: How do Countries Differ? *CSAE Working Paper Series 2008-12*. Centre for the Study of African Economies, University of Oxford.

Eckstein, Z., & Wolpin, K. I. (1989). Dynamic Labour Force Participation of Married Women and Endogenous Work Experience. *Review of Economic Studies, 3*, 375-390.

Eyal, K., & Woolard, I. (2011). Female Labour Force Participation and the Child Support Grant in South Africa. *CSAE 25th Anniversary Conference, 20th - 22nd March 2011*. Oxford, England.

Fedderke, J. W. (2002). Technology, Human Capital and Growth: evidence from a middle income country case study applying dynamic heterogeneous panel analysis. *Presented at the Royal Economic Society Annual Conference*. Warwick: University of Warwick.

Fedderke, J. W. (2002, March). The Structure of Growth in the South African Economy: Factor Accumulation and Total Factor Productivity Growth 1970-97. *South African Journal of Economics, 70*(4), 282–299.

Frisch, R., & Waugh, F. V. (1933, October). Partial Time Regressions as Compared with Individual Trends. *Econometrica, 4*, 387-401.

García, J., Hernández, P. J., & López-Nicolás, A. (2001). How wide is the gap? An investigation of gender wage differences using quantile regression. *Empirical Economics, 26*(1), 149-167.

Garen, J. (1984, September). The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable. *Econometrica, 52*(5), 1199-1218.

Goffe, W. L., Ferrier, G. D., & Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics, 60*(1-2), 65-99.

Goodey, J. S. (1988). Multicultural education in South Africa: is it possible? *Journal of Abstracts in International Education, 17*(1), 7-23.

Griliches, Z. (1977, January). Estimating the returns to schooling: some econometric problems. *Econometrica, 45*(1), 1–22.

Griliches, Z., & Hausman, J. A. (1986, February). Errors in variables in panel data. *Journal of Econometrics, 31*(1), 93-118.

Gronau, R. (1974, Nov. - Dec.). Wage Comparisons--A Selectivity Bias. *The Journal of Political Economy, 82*(6), 1119-1143.

Guluza, X., & Hoadley, U. K. (1998). *Learner Progress and Achievement Study: Research report 2.* Braamfontein: Joint Education Trust.

Hall, R. E., & Jones, C. I. (1999, February). Why Do Some Countries Produce So Much More Output Per Worker Than Others? *The Quarterly Journal of Economics, 114*(1), 83-116.

Hansen, L. P. (1982, July). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica, 50*(4), 1029-1054.

Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature, 46*(3), 607-668.

Harmon, C., & Walker, I. (1995, December). Estimates of the Economic Return to Schooling for the United Kingdom. *The American Economic Review, 85*(5), 1278-1286.

Hartshorne, K. (1992). *Crisis and Challenge: Black education 1910-1990.* Cape Town: Oxford University Press.

Heckman, J. J. (1974). Shadow Prices, Market Wages and Labor Supply. *Econometrica, 42*(4), 679-694.

Heckman, J. J. (1979, January). Sample Selection Bias as a Specification Error. *Econometrica, 47*(1), 153-161.

Heckman, J. J. (1990). Varieties of Selection Bias. *American Economic Review, Papers and Proceedings, 80*(2), 313-318.

Heckman, J. J. (2010, June). 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, March). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica, 52*(2), 271-320.

Heckman, J. J., & Urzúa, S. (2010, May). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics, 156*(1), 27-37.

Heckman, J. J., & Vytlacil, E. (1998, Autumn). Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling When the Return is Correlated with Schooling. *The Journal of Human Resources, 33*(4), 974-987.

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In E. A. Hanushek, & F. Welch, *Handbook of the Economics of Education* (Vol. 1, pp. 307-458). Amsterdam: Elsevier.

Heckman, J. J., Lochner, L. J., & Todd, P. E. (2008). Earnings Functions and Rates of Return. *Journal of Human Capital, 2*(1), 1-31.

Heckman, J. J., Lochner, L., & Taber, C. (1998, January). Explaining Rising Wage Inequality: Explanations With A Dynamic General Equilibrium Model of Labor Earnings With Heterogeneous Agents. *Review of Economic Dynamics, 1*(1), 1-58.

Hotz, V. J., & Miller, R. A. (1993, July). Conditional Choice Probabilities and the Estimation of Dynamic Models. *The Review of Economic Studies, 60*(3), 497-529.

Imbens, G. W. (2010, June). Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature, 48*(2), 399-423(25).

Imbens, G. W., & Angrist, J. D. (1994, March). Identification and Estimation of Local Average Treatment Effects. *Econometrica, 62*(2), 467-475.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature, 47*(1), 5–86.

Judson, R. (1998). Economic Growth and Investment in Education: How Allocation Matters. *Journal of Economic Growth, 3*(4), 337-359.

Kane, T. J., & Rouse, C. E. (1995, June). Labor-Market Returns to Two- and Four-Year College. *The American Economic Review, 85*(3), 600-614.

Keane, M. P. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics, 156* , 3-20.

Keane, M. P., & Wolpin, K. I. (1994, November). The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence. *The Review of Economics and Statistics, 76*(4), 648-672.

Keane, M. P., & Wolpin, K. I. (1997, June). The Career Decisions of Young Men. *The Journal of Political Economy, 105*(3), 473-522.

Keswell, M., & Poswell, L. (2004, September). Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. *The South African Journal of Economics, 72*(4), 834 - 860.

Kingdon, G. G., & Knight, J. (2005). Unemployment in South Africa, 1995-2003: Causes, Problems and Policies. *Paper presented at the `10 year review of the South African Economy' conference, 28-29 October 2005.* Stellenbosch University, Stellenbosch.

Kingdon, G. G., & Knight, J. (2006, April). How Flexible Are Wages in Response to Local Unemployment in South Africa. *Industrial and Labor Relations Review, 59*(3), 471-495.

Kraak, A. (2008). The education-economy relationship in South Africa, 2001-2005. In A. Kraak, & K. Press (Eds.), *Human Resources Development Review 2008* (pp. 1-25). Cape Town: HSRC Press.

Krueger, A. B., & Lindahl, M. (2001, December). Education for Growth: Why and for Whom? *Journal of Economic Literature, 39*(4), 1101-1136.

LaLonde, R. J. (1986, September). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review, 76*(4), 604-620.

Lee, D. (2005). An Estimable Dynamic General Equilibrium Model of Work, Schooling and Occupational Choice. *International Economic Review, 46*(1), 1-34.

Lee, D., & Wolpin, K. I. (2006). Intersectoral Labor Mobility and the Growth of the Service Sector. *Econometrica, 74*(1), 1-46.

Lemieux, T. (2006). The Mincer Equation Thirty Years after Schooling, Experience, and Earnings. In S. Grossbard-Shechtman (Ed.), *Jacob Mincer, A Pioneer of Modern Labor Economics.* Springer Verlag.

Lovell, M. (1963). Seasonal adjustment of economic time series. *Journal of the American Statistical Association, 58*, 993-1010.

Malherbe, E. G. (1977). *Education in South Africa, Volume II: 1923-75.* Cape Town: Juta & Company Limited.

Manski, C. (1988). *Path Utility Analysis of Dynamic Choice.* SSRI Working Paper No. 8805, University of Wisconsin, Madison.

Mincer, J. (1958). Investment in Human Capital and Personal Income Distribution. *Journal of Political Economy, 66*(4), 281-302.

Mincer, J. (1974). *Schooling, Experience and Earnings.* New York: NBER.

Mincer, J. (1996). Changes in Wage Inequality, 1970-1990. *NBER Working Paper No. 5823*. Cambridge: NBER.

Moffitt, R. (1999). New developments in econometric methods for labor market analysis. In O. Ashenfelter, & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3A, pp. 1367-1397). Amsterdam: Elsevier.

NEPAD. (2001). *The New Partnership for Africa's Development.*

Nielsen, H. S., & Westergard-Nielsen, N. (2001, January). Returns to Schooling in Less Developed Countries: New Evidence from Zambia. *Economic Development and Cultural Change, 49*(2), 365-94.

Perry, H., & Arends, F. (2003). Public Schooling. In HSRC, *Human Resources Development Review 2003* (pp. 303-325). Cape Town: HSRC Press.

Pesaran, M. H. (2006, July). Estimation and inference in large heterogeneous panels with multifactor error structure. *Econometrica, 74*(4), 967-1012.

Pesaran, M. H., & Smith, R. P. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics, 68*, 79-113.

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

Pritchett, L. (2001). Where Has All the Education Gone? *The World Bank Economic Review, 15*(3), 367-391.

Psacharopoulos, G. (1973). *Returns to Education: An International Comparison.* Amsterdam: Elsevier.

Psacharopoulos, G. (1985). Returns to Education: A Further International Update and Implications. *Journal of Human Resources, 20*(4), 583-604.

Psacharopoulos, G. (1994, September ). Returns to investment in education: A global update. *World Development, 22*(9), 1325-1343.

Psacharopoulos, G., & Patrinos, H. A. (2004, August). Returns to investment in education: a further update. *Education Economics, 12*(2), 111-134.

Ranchod, V., & Dinkelman, T. (2008). *Labour Market Transitions in South Africa: What can we learn from matched Labour Force Survey data?* Cape Town: Southern Africa Labour and Development Research Unit.

Republic of South Africa. (1995). *Meeting the Commitment to Free and Compulsory General Education. Chapter 13. White Paper on Education and Training. Notice 196 of 1995. Department of Education. Available [Online]:.* Retrieved from http://www.polity.org.za/html/govdocs/white_papers/educ5.html

Rivers, D., & Vuong, Q. H. (1988, November). Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics, 39*(3), 347-366.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy, 98*(5), S71-S102.

Rust, J. (1987, September). Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica, 55*(5), 999-1033 .

Siphambe, H. K. (2000, June). Rates of return to education in Botswana. *Economics of Education Review, 19*(3), 291-300.

Skyt-Nielsen, H., & Westergaard-Nielsen, N. (1998, September). Returns to Schooling in LDCs: New Evidence from Zambia. *CLS Working Paper No. 98-010* . Aarhus: University of Aarhus.

Smit, B. W., & Burrows, L. R. (2002). *Estimating Potential Output and Output Gaps for the South African Economy.* Stellenbosch: Stellenbosch University.

Söderbom, M. M., & Teal, F. J. (2004, February). Size and efficiency in African manufacturing firms: evidence from firm-level panel data. *Journal of Development Economics, 73*(1), 369-394.

Söderbom, M. M., Teal, F., Wambugu, A., & Kahyarara, G. (2006, June). The Dynamics of Returns to Education in Kenyan and Tanzanian Manufacturing. *Oxford Bulletin of Economics and Statistics, 68*(3), 261-288.

South African Reserve Bank. (various years). Quarterly Bulletin. Pretoria: South African Reserve Bank.

Spence, M. (1973, Augustus). Job Market Signaling. *The Quarterly Journal of Economics, 87*(3), 355-374.

Staiger, D., & Stock, J. H. (1997, May). Instrumental Variables Regression with Weak Instruments. *Econometrica, 65*(3), 557-586.

Statistics South Africa. (1996). *October Household Survey 1995 Statistical release.* Pretoria: Government printer.

Statistics South Africa. (1999). *October Household Survey 1997 Statistical release.* Pretoria: Government printer.

Statistics South Africa. (2000). *October Household Survey 1998 Statistical release.* Pretoria: Government printer.

Statistics South Africa. (2002). *Labour Force Survey February 2002 Statistical release.* Pretoria: Government printer.

Statistics South Africa. (various years). Labour Force Surveys. Pretoria: Statistics South Africa.

Statistics South Africa. (various years). October Household Surveys. Pretoria: Statistics South Africa.

Stock, J. H., & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. W. Andrews, & J. H. Stock, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 80–108). Cambridge: Cambridge University Press, 2005.

Teal, F. (2001). *Education, Incomes, Poverty and Inequality in Ghana in the 1990's.* CSAE Working Paper 2001-21, University of Oxford.

Temple, J. (1999, October). A positive effect of human capital on growth. *Economics Letters, 65*(1), 131-134.

Theil, H. (1961). *Economic forecasts and policy* (2nd revised ed.). Amsterdam: North-Holland.

UNESCO. (2011). *Financing Education in Sub-Saharan Africa.* Montreal: UNESCO Institute for Statistics.

United Nations. (2000). *United Nations Millennium Declaration.* New York: United Nations.

Van der Berg, S. (2002). Issues in South African Social Security. *Working Papers 01/2002*. Stellenbosch University, Department of Economics.

Van Walbeek, C. P. (2006). Official revisions to South African national accounts data: Magnitudes and implications. *South African Journal of Economics, 74*(4), 745-765.

Vella, F. (1998). Estimating Models with Sample Selection Bias: A Survey. *The Journal of Human Resources, 33*(1), 127-169.

Vermaak, C. (2010). Low-wage mobility in the South African labour market. *Paper presented at DPRU/TIPS Conference.* Johannesburg, October 2010.

Von Fintel, D. (2007). Dealing With Earnings Bracket Responses In Household Surveys - How Sharp Are Midpoint Imputations? *South African Journal of Economics, 75*(2), 293-312.

Western Cape Education Department. (2003). *Admission Of Over-Age Learners To Public Schools.* Retrieved from http://wced.wcape.gov.za/circulars/2003/e240_03.html

Whaba, J. (2000). Returns to Education and Regional Earnings Differentials in Egypt. *Discussion Papers in Economics and Econometrics.* Southampton: University of South Hampton.

Williamson, J. (1990). What Washington Means by Policy Reform. In J. Williamson (Ed.), *Latin American Adjustment: How Much Has Happened.* Washington, D.C.: Institute for International Economics.

Wooldridge, J. M. (1997, October). On two stage least squares estimation of the average treatment effect in a random coefficient model. *Economics Letters, 56*(2), 129-133.

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

Wooldridge, J. M. (2003, May). Further results on instrumental variables estimation of average treatment effects in the correlated random coefficient model. *Economics Letters, 79*(2), 185-191.

Wooldridge, J. M. (2007). *Control Functions and Related Methods.* Retrieved from http://video.google.com/videoplay?docid=-3103055655385103797#

Yu, D. (2007). The comparability of the Statistics South Africa October Household Surveys and Labour Force Surveys. *Stellenbosch Working Papers No. 2007/13*. Stellenbosch: University of Stellenbosch.

# 7. Appendix

## 7.1 First-stage schooling regression

Table 15 reports the results of a series of schooling regressions in which age is included as a spline function and birth year as a cubic function. Unlike the quadratic function used in Table 1, including age as a spline does not impose the awkward restriction that an individual's expected schooling years will decrease after a certain age is reached. Adding birth year as an additional determinant may capture the effect of unobservable shifts in how different generations perceive the costs and value of schooling that otherwise may be erroneously captured by our instrumental variables. Since age, birth year and sample year are perfectly multicolinear, we choose to preclude the year effects from capturing any long-run trends in the schooling outcome by restricting their coefficients to be orthogonal to time itself (as suggested for cohort panels by Deaton (1985)). In this case the birth year variables will capture the generational variation and the sample year dummies the cyclical variation in schooling outcomes.

The completed schooling years variable is now regressed on the new set of conditioning variables and the IVs associated with the first policy shift (column 1), the second policy (column 2) and both policies (column 3). The age-spline coefficients show a strong effect of age on completed schooling years during the late teens, a smaller but still substantial effect in the twenties and then a much smaller effect during the remainder of working life. The year effects reveal a cyclical schooling pattern that is relatively small but appears to vary in the same direction as the business cycle. The birth year coefficients expose a strong generational trend according to which members of older birth cohorts can expect to have as much as five years of schooling fewer than those from younger generations. This trend towards increased schooling levels amongst younger generations is consistent with the pattern observed in other groups that were unaffected by apartheid schooling restrictions, including workers in most African countries and South African whites, and therefore cannot be ascribed merely to the relaxing of apartheid-era restrictions.

Table 15: First-stage OLS regression on years of education (sample of black males), 1995-2007

| | | | |
|---|---|---|---|
| Grade 1 (cohort %) | -0.639 | | -0.603 |
| Grade 2 (cohort %) | 1.299* | | 1.322* |
| Grade 3 (cohort %) | 0.744 | | 0.751 |
| Grade 4 (cohort %) | 1.682*** | | 1.622*** |
| Grade 5 (cohort %) | 1.887*** | | 1.895*** |
| Grade 6 (cohort %) | 1.394*** | | 1.392*** |
| Grade 7 (cohort %) | 1.723*** | | 1.727*** |
| Grade 8 (cohort %) | 1.764*** | | 1.799*** |
| Grade 9 (cohort %) | 1.493*** | | 1.481*** |
| Grade 10 (cohort %) | 2.098*** | | 2.017*** |
| Grade 11 (cohort %) | 2.187*** | | 2.259*** |
| Grade 12 (cohort %) | 3.565*** | | 3.435*** |
| Post-school dipl. /cert. (cohort %) | 4.748*** | | 4.451*** |
| Degree (cohort %) | 4.006*** | | 3.687*** |
| P(allow overage), age = 16 | | 0.997 | 0.886 |
| P(allow overage), age = 17 | | 0.86 | 0.608 |
| P(allow overage), age = 18 | | 2.468*** | 2.096*** |
| P(allow overage), age = 19 | | 1.014* | 0.491 |
| P(allow overage), age = 20 | | 1.295*** | 0.858* |
| P(allow overage), age = 21 | | -1.369** | -1.159* |
| P(allow overage), age = 22 | | 0.128 | -0.036 |
| P(allow overage), age = 23 | | 1.121** | 0.695 |
| P(allow overage), age = 24 | | 1.233*** | 1.114** |
| P(allow overage), age = 25 | | 1.293* | 0.91 |
| P(allow overage), age = 26 | | -0.034 | 0.095 |
| Age spline (15-20) | 0.874*** | 1.167*** | 0.947*** |
| Age spline (20-30) | 0.15*** | 0.165*** | 0.15*** |
| Age spline (30-40) | 0.026*** | 0.019*** | 0.027*** |
| Age spline (40-50) | -0.043*** | -0.049*** | -0.041*** |
| Age spline (50-60) | -0.017 | -0.031** | -0.022* |
| Age spline (60-65) | -0.08* | -0.112** | -0.079* |
| Year = 1997 (restricted) | 0.29*** | 0.297*** | 0.29*** |
| Year = 1998 (restricted) | -0.214*** | -0.265*** | -0.218*** |
| Year = 1999 (restricted) | -0.115** | -0.155*** | -0.119** |
| Year = 2000 (restricted) | -0.192*** | -0.248*** | -0.196*** |
| Year = 2001 (restricted) | -0.114*** | -0.151*** | -0.116*** |
| Year = 2002 (restricted) | -0.05 | -0.063 | -0.048 |
| Year = 2003 (restricted) | -0.067 | -0.069 | -0.065 |
| Year = 2004 (restricted) | -0.077* | -0.082** | -0.076* |
| Year = 2005 (restricted) | 0.078* | 0.09** | 0.079* |
| Year = 2006 (restricted) | 0.178*** | 0.225*** | 0.18*** |
| Year = 2007 (restricted) | 0.073* | 0.099** | 0.072* |
| Birth year | -0.019 | -0.025 | -0.01 |
| Birth year^2 | 0.005*** | 0.005*** | 0.004*** |
| Birth year^3 | -0.00005*** | -0.00005*** | -0.00004*** |
| Western Cape | -0.824*** | -0.027 | -0.252 |
| Eastern Cape | 0.09 | 0.001 | 0.131 |
| Northern Cape | -1.923*** | -0.388 | -0.66** |
| Free State | -1.133*** | -0.882*** | -0.894*** |
| Kwazulu-Natal | -0.21*** | 0.151 | 0.066 |
| North-West | -0.756*** | 0.048 | -0.1 |
| Gauteng | 0.146** | 0.668*** | 0.328** |
| Mpumalanga | -0.985*** | -1.059*** | -0.935*** |
| Constant | -14.715*** | -25.155*** | -20.449*** |
| Observations | 129659 | 129659 | 129659 |
| R-squared | 0.184 | 0.182 | 0.184 |
| F-test (IV sets 1 & 2) | 13.888 | 6.773 | 9.591 |
| F-test (IV set 1) | 13.888 | | 12.074 |
| F-test (IV set 2) | | 6.773 | 3.84 |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

The pattern of the IV coefficients are similar to what was observed in Table 1, but adding birth year as an additional control further reduces the explanatory power of the instruments (as observed by their smaller magnitudes and F-test statistics). This means that 2SLS estimates based on this specification will be somewhat more vulnerable to weak instrument bias. However, it remains a useful specification to test whether our instruments are affecting earnings through channels other than the one proposed as part of our instrumenting strategy.

## 7.2 Earnings regressions

### 7.2.1 The local average treatment effect interpretation

Table 16 investigates the sensitivity of the results in Table 4 to using a more flexible specification. Specifically, it allows age to enter the earnings regression via a more flexible age-spline specification, and for birth year to determine earnings through a third-order polynomial function. In order to avoid perfect multicolinearity, the year coefficients are restricted to be orthogonal to the sample year. The conditioning variables therefore correspond to those used in the schooling regressions reported in Table 15. The OLS schooling returns are almost identical to those in Table 4, but the 2SLS estimates are higher than when not controlling for birth year. The IVs associated with the second policy are again greater than those associated with the first policy, with the combined estimates lying in-between but somewhat closer to the lower end of the spectrum. The fact that birth year controls do not eliminate or even reduce the estimated schooling returns suggest that the correlation between our IVs and earnings is not due to other omitted birth year specific factors, such as generational trends in the quality of schooling or macro-economic environment.

Table 16: OLS and 2SLS earnings regressions using different IVs (sample of black males), 1995-2007

| | **OLS** | **2SLS** | **2SLS** | **2SLS** |
|---|---|---|---|---|
| Education | 0.107*** | 0.125*** | 0.228*** | 0.146*** |
| | *(0.0008)* | *(0.0140)* | *(0.0298)* | *(0.0132)* |
| Observations | 129659 | 129659 | 129659 | 129659 |
| R-squared | 0.298 | 0.292 | 0.042 | 0.271 |
| CD F-test | | 24.918 | 9.308 | 16.311 |
| J-test | | 67.826 | 122.779 | 201.731 |
| p-value | | 0.000 | 0.000 | 0.000 |
| Instrumental variables | None | IV sets 1 | IV sets 2 | IV sets 1 & 2 |
| Also control for: | Sample year (restricted), province, birth year (cubic polynomial) & age (spline) | | | |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

### 7.2.2 Average schooling-earnings profile interpretation

The CF regression in Table 6 requires that the expected values of the unobserved relative return and relative ability variables in the earnings regression are both linear in the schooling error, and also that the schooling outcome variable is not "so discrete" as to invalidate the assumption that the schooling error is independent of the regressors. However, in section 2.3.3.1 we discussed alternative regressions that will provide consistent estimates of the earnings coefficients while relaxing these two sets of assumptions. Table 17 below replicates the CF estimates from Table 6 in column 1, and then re-estimates this model with alternative specifications of this schooling residual: a third order polynomial in column 2 and as calculated from an ordered probit schooling regression in column 3.

The significance of the higher order schooling residual terms in column 2 indicates that market ability is actually a non-linear function the schooling error. Accounting for this more complex relationship between the unobservable components has the effect of turning the schooling-earnings profile concave. The coefficients on the ordered probit residual and its interaction with schooling in column 3 are both significantly positive and large in magnitude. Estimating the schooling regression with OLS implicitly assumes that the different thresholds that learners need to cross in order to complete the various grades are equidistant, but this assumption is relaxed when estimating an ordered probit schooling equation. The

expected schooling error from this model (conditional on having completed the different grades) rises much faster for each year of completed education at more advanced schooling levels. In a sense this specification is therefore also able to replicate the non-linear relationship between the OLS schooling and wage error terms found in column 2.

Table 17: CF earnings regressions with different education residual controls (sample of black males), 1995-2007

| | **CF** | **CF** | **CF** |
|---|---|---|---|
| Education | 0.073*** | 0.138*** | 0.05*** |
| | *(0.0099)* | *(0.0203)* | *(0.0082)* |
| Education^2 | 0.003*** | -0.0017* | -0.0025*** |
| | *(0.0004)* | *(0.0010)* | *(0.0004)* |
| Education residual | -0.0563*** | -0.159*** | |
| | | *(0.0202)* | |
| Education residual^2 | | -0.0213*** | |
| | | *(0.0013)* | |
| Education residual^3 | | -0.001*** | |
| | | | |
| Education residual*Educ | 0.007*** | 0.02*** | |
| | *(0.0004)* | *(0.0019)* | |
| Education residual^2*Educ | | 0.0022*** | |
| | | *(0.0001)* | |
| Education residual^3*Educ | | -0.0001*** | |
| | | *(0.00001)* | |
| Education oprobit residual | | | 0.102*** |
| | | | *(0.0266)* |
| Education oprobit residual*Educ | | | 0.0348*** |
| | | | *(0.0015)* |
| Age | 0.0878*** | 0.0775*** | 0.1032*** |
| | *(0.0023)* | *(0.0028)* | *(0.0024)* |
| Age^2 | -0.0008*** | -0.0006*** | -0.0011*** |
| | *(0.00004)* | *(0.00005)* | *(0.00003)* |
| Observations | 129659 | 129659 | 129659 |
| R-squared | 0.334 | 0.341 | 0.345 |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level. Standard deviations in brackets.

Table 18: Earnings regressions comparison of OLS, 2SLS and CF estimators (sample of black males), 1995-2007

| | OLS | 2SLS | CF |
|---|---|---|---|
| Education | -0.027*** | -0.097** | 0.083*** |
| | *(0.0024)* | *(0.0423)* | *(0.0176)* |
| Education^2 | 0.009*** | 0.0161*** | 0.0041*** |
| | *(0.0002)* | *(0.0025)* | *(0.0005)* |
| Education residual | | 0.07 | -0.075*** |
| | | *(0.0428)* | *(0.0163)* |
| Education^2 residual | | -0.0071*** | |
| | | *(0.0025)* | |
| Education residual*Educ | | | 0.006*** |
| | | | *(0.0005)* |
| Observations | 129659 | 129659 | 129659 |
| R-squared | 0.334 | 0.334 | 0.336 |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level. Standard deviations in brackets.

## 7.3 Sensitivity analysis

Table 19: CF earnings regressions with different birth year controls (sample of black males), 1995-2007

| | CF | CF | CF | CF | CF |
|---|---|---|---|---|---|
| Education | 0.023** | 0.087*** | 0.083*** | 0.08*** | 0.07*** |
| | *(0.0092)* | *(0.0173)* | *(0.0157)* | *(0.0154)* | *(0.0159)* |
| Education^2 | 0.0035*** | 0.0032*** | 0.0041*** | 0.0044*** | 0.0045*** |
| | *(0.0005)* | *(0.0005)* | *(0.0005)* | *(0.0004)* | *(0.0005)* |
| Education residual | -0.0098 | -0.0716*** | -0.0748*** | -0.0733*** | -0.0652*** |
| | *(0.0070)* | *(0.0159)* | *(0.0148)* | *(0.0151)* | *(0.0139)* |
| Education residual*Educ | 0.006*** | 0.007*** | 0.006*** | 0.005*** | 0.005*** |
| | *(0.0004)* | *(0.0004)* | *(0.0005)* | *(0.0004)* | *(0.0005)* |
| Observations | 129659 | 129659 | 129659 | 129659 | 129659 |
| R-squared | 0.334 | 0.334 | 0.336 | 0.336 | 0.337 |
| Birth year control: | None | Linear | Cubic polynomial | 5th order polynomial | Dummies |
| Also control for: | Sample year (restricted), province & age (spline) | | | | |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level. Standard deviations in brackets.

Table 20: CF earnings regressions with different control and instrumental variables (sample of black males), 1995-2007

| | **CF** | **CF** | **CF** | **CF** |
|---|---|---|---|---|
| Education | 0.073*** | 0.05*** | 0.091*** | 0.106*** |
| | *(0.0103)* | *(0.0108)* | *(0.0128)* | *(0.0132)* |
| Education^2 | 0.003*** | 0.003*** | 0.004*** | 0.0033*** |
| | *(0.0004)* | *(0.0004)* | *(0.0005)* | *(0.0005)* |
| Education residual | -0.0563*** | -0.0329*** | -0.0825*** | -0.0911*** |
| | *(0.0083)* | *(0.0086)* | *(0.0113)* | *(0.0116)* |
| Education residual*Educ | 0.007*** | 0.007*** | 0.006*** | 0.006*** |
| | *(0.0004)* | *(0.0005)* | *(0.0005)* | *(0.0005)* |
| Prim. school interest rate | | | -0.027*** | -0.028*** |
| | | | *(0.0058)* | *(0.0059)* |
| Sec. school interest rate | | | -0.003 | -0.003 |
| | | | *(0.0040)* | *(0.0040)* |
| Ter. school interest rate | | | 0.009*** | 0.008*** |
| | | | *(0.0021)* | *(0.0022)* |
| Prim. school growth rate | | | 1.942*** | 1.863** |
| | | | *(0.7135)* | *(0.7323)* |
| Sec. school growth rate | | | -0.015 | -0.033 |
| | | | *(0.5970)* | *(0.6089)* |
| Ter. school growth rate | | | 0.023 | -0.005 |
| | | | *(0.3115)* | *(0.3192)* |
| Cohort employment rate | | | | 0.118*** |
| | | | | *(0.0111)* |
| Observations | 129659 | 129659 | 127669 | 122228 |
| R-squared | 0.334 | 0.334 | 0.337 | 0.328 |
| Also control for: | Sample year, province, age | | | |
| IV variation | Birth year, province, sample year | Birth year, sample year | Birth year, province, sample year | Birth year, province, sample year |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level. Standard deviations in brackets.

## 7.4 Production function estimates

Table 21: POLS estimates of production function coefficients, with various human capital link functions

| | TE | TE | TE | TE |
|---|---|---|---|---|
| **Dependent variable** | *Log output* | *Log output* | *Log output* | *Log output* |
| Observations | 117 | 117 | 117 | 117 |
| R-squared | 0.822 | 0.944 | 0.930 | 0.958 |
| **Coefficient estimates** | | | | |
| Constant | | 3.421*** | 5.329*** | 1.623*** |
| | *(0.558)* | *(0.325)* | *(0.352)* | *(0.530)* |
| Log capital stock | 0.416*** | 0.236*** | 0.226*** | 0.285*** |
| | *(0.031)* | *(0.021)* | *(0.025)* | *(0.021)* |
| Log employment | 0.552*** | 0.571*** | 0.557*** | 0.605*** |
| | *(0.040)* | *(0.023)* | *(0.025)* | *(0.020)* |
| Log average education | | 1.492*** | | |
| | | *(0.099)* | | |
| Average education | | | 0.182*** | 0.798*** |
| | | | *(0.014)* | *(0.076)* |
| Average education^2 | | | | -0.038*** |
| | | | | *(0.005)* |
| **Period dummies** | Yes | Yes | Yes | Yes |
| **Industry dummies** | No | No | No | No |
| **Durbin-Watson statistic** | 0.071 | 0.287 | 0.224 | 0.422 |
| **Implied returns to scale** | 0.97 | 0.81 | 0.78 | 0.89 |

Table 22: OLS time series regression estimates of production function, by industry

| | **Agri-culture** | **Mining** | **Manu-facturing** | **Con-struction** | **Utilities** | **Internal trade** | **Transport-ation** | **Financial services** | **Community services** | **Total Average** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Dependent variable** | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* |
| Observations | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 117 |
| R-squared | 0.594 | 0.737 | 0.965 | 0.474 | 0.969 | 0.984 | 0.974 | 0.954 | 0.930 | 0.998 |
| **Coefficient estimates** | | | | | | | | | | |
| Constant | 67.686* | 7.636** | -5.932 | 22.998** | 18.801*** | 12.584 | -13.464 | -37.23* | 14.558 | 9.738 |
| | *(30.242)* | *(2.728)* | *(7.038)* | *(7.296)* | *(4.732)* | *(9.556)* | *(10.470)* | *(18.799)* | *(19.978)* | *(9.604)* |
| Log capital stock | -4.207 | 0.65*** | 1.393*** | -0.208 | 0.463*** | 1.157*** | 2.023*** | 1.896** | 1.097*** | 0.474 |
| | *(2.554)* | *(0.144)* | *(0.232)* | *(0.329)* | *(0.105)* | *(0.233)* | *(0.263)* | *(0.594)* | *(0.182)* | *(0.629)* |
| Log employment | -0.42*** | -0.035 | 0.405*** | 0.068 | 0.174 | 0.046 | 0.007 | 0.246 | 0.041 | 0.059 |
| | *(0.116)* | *(0.026)* | *(0.109)* | *(0.090)* | *(0.110)* | *(0.079)* | *(0.159)* | *(0.248)* | *(0.090)* | *(0.075)* |
| Average education | 0.693 | 0.526 | 0.077 | -1.094 | -2.426** | -1.893 | 0.852 | 4.786 | -1.919 | -0.044 |
| | *(0.630)* | *(0.420)* | *(1.060)* | *(0.898)* | *(0.996)* | *(1.638)* | *(2.003)* | *(3.050)* | *(3.245)* | *(0.733)* |
| Average education^2 | -0.072 | -0.032 | 0.001 | 0.057 | 0.16** | 0.101 | -0.035 | -0.21 | 0.082 | 0.006 |
| | *(0.056)* | *(0.026)* | *(0.055)* | *(0.044)* | *(0.063)* | *(0.083)* | *(0.099)* | *(0.133)* | *(0.143)* | *(0.037)* |
| **DW statistic** | 2.579 | 2.189 | 2.479 | 2.055 | 2.674 | 1.858 | 1.567 | 1.041 | 0.976 | 2.102 |
| **Implied RTS** | -3.934 | 1.142 | 1.875 | -1.235 | -1.790 | -0.690 | 2.882 | 6.928 | -0.781 | 0.489 |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 23: IV-TE estimates of production function coefficients

| | **IV_TE** | **IV_TE** | **IV_TE** | **IV_TE** | **IV_TE** |
|---|---|---|---|---|---|
| **Dependent variable** | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* |
| Observations | 104 | 88 | 99 | 108 | 108 |
| R-squared | 0.970 | 0.974 | 0.970 | 0.965 | 0.965 |
| **Coefficient estimates** | | | | | |
| Constant | 3.754*** | 3.035*** | 1.513** | 1.638*** | 1.625*** |
| | *(0.811)* | *(0.814)* | *(0.634)* | *(0.520)* | *(0.520)* |
| Log capital stock | 0.317*** | 0.316*** | 0.292*** | 0.286*** | 0.286*** |
| | *(0.018)* | *(0.018)* | *(0.019)* | *(0.020)* | *(0.020)* |
| Log employment | 0.611*** | 0.614*** | 0.619*** | 0.618*** | 0.618*** |
| | *(0.017)* | *(0.017)* | *(0.032)* | *(0.020)* | *(0.020)* |
| Average education | 0.306* | 0.492*** | 0.811*** | 0.806*** | 0.806*** |
| | *(0.161)* | *(0.160)* | *(0.077)* | *(0.076)* | *(0.076)* |
| Average education^2 | -0.014* | -0.024*** | -0.038*** | -0.038*** | -0.038*** |
| | *(0.008)* | *(0.008)* | *(0.005)* | *(0.005)* | *(0.005)* |
| F-test | 1000.725 | 1133.293 | 22.749 | 193.115 | 108.492 |
| J-test | 0.000 | 3.739 | 0.876 | 0.000 | 2.271 |
| p-value | . | 0.291 | 0.349 | . | 0.518 |
| **Period dummies** | Yes | Yes | Yes | Yes | Yes |
| **Industry dummies** | No | No | No | No | No |
| **Implied RTS** | 0.929 | 0.931 | 0.911 | 0.904 | 0.904 |
| **Endogenous regressors** | Log employment | Log employment | Log employment | All factors of production | All factors of production |
| **Instrumental variables** | Log employment (alternative measure) | Log employment (alternative measure, lags 0 & 1), log employment (lags 1 & 2) | Log wage (lags 1 & 2) | All factors of production (lag 1) | All factors of production (lag 1), log wage, bond yield, union membership (lag 1) |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 24: IV-FE estimates of production function coefficients

| | **IV-2FE** | **IV-2FE** | **IV-2FE** | **IV-2FE** | **IV-2FE** |
|---|---|---|---|---|---|
| **Dependent variable** | *Log output* | *Log output* | *Log output* | *Log output* | *Log output* |
| Observations | 104 | 88 | 99 | 108 | 108 |
| R-squared | 0.983 | 0.993 | 0.994 | 0.991 | 0.945 |
| **Coefficient estimates** | | | | | |
| Constant | 3.898 | 6.613*** | 5.623*** | 5.191*** | -0.8 |
| | *(2.431)* | *(1.418)* | *(1.289)* | *(1.861)* | *(7.387)* |
| Log capital stock | -0.076 | 0.087 | 0.27 | 0.402* | -0.549 |
| | *(0.334)* | *(0.269)* | *(0.178)* | *(0.238)* | *(1.077)* |
| Log employment | 0.841*** | 0.584*** | 0.542*** | 0.463* | 1.523 |
| | *(0.289)* | *(0.188)* | *(0.162)* | *(0.246)* | *(1.175)* |
| Average education | 0.457 | 0.401 | 0.087 | -0.007 | 0.451 |
| | *(0.364)* | *(0.289)* | *(0.097)* | *(0.145)* | *(0.569)* |
| Average education^2 | -0.016 | -0.018 | -0.001 | 0.012* | 0.006 |
| | *(0.018)* | *(0.015)* | *(0.005)* | *(0.007)* | *(0.017)* |
| F-test | 8.208 | 5.574 | 7.882 | 1.345 | 0.808 |
| J-test | 0.000 | 8.337 | 1.091 | 0.000 | 5.346 |
| p-value | . | 0.040 | 0.296 | . | 0.148 |
| **Period dummies** | Yes | Yes | Yes | Yes | Yes |
| **Industry dummies** | Yes | Yes | Yes | Yes | Yes |
| **Implied RTS** | 0.764 | 0.671 | 0.811 | 0.866 | 0.974 |
| **Endogenous regressors** | Log employment | Log employment | Log employment | All factors of production | All factors of production |
| **Instrumental variables** | Log employment (alternative measure) | Log employment (alternative measure, lags 0 & 1), log employment (lags 1 & 2) | Log wage (lags 1 & 2) | All factors of production (lag 1) | All factors of production (lag 1), log wage, bond yield, union membership (lag 1) |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 25: IV-FD estimates of production function coefficients

| | **IV-FD** | **IV-FD** | **IV-FD** | **IV-FD** | **IV-FD** |
|---|---|---|---|---|---|
| **Dependent variable** | *D(Log output)* | *D(Log output)* | *D(Log output)* | *D(Log output)* | *D(Log output)* |
| Observations | 96 | 80 | 108 | 99 | 99 |
| R-squared | 0.157 | 0.420 | 0.119 | 0.007 | 0.215 |
| **Coefficient estimates** | | | | | |
| Constant | 0.028 | 0.024 | 0.017 | 0.024 | 0.015 |
| | *(0.022)* | *(0.016)* | *(0.020)* | *(0.021)* | *(0.014)* |
| D(Log capital stock) | 0.49 | 0.603** | 0.691*** | 1.001*** | 1.09*** |
| | *(0.507)* | *(0.272)* | *(0.261)* | *(0.268)* | *(0.195)* |
| D(Log employment) | 0.245 | 0.162 | 0.157 | 0.097 | -0.084 |
| | *(0.494)* | *(0.213)* | *(0.166)* | *(0.347)* | *(0.091)* |
| D(Average education) | 0.127 | 0.163 | 0.08 | 0.2 | 0.188 |
| | *(0.100)* | *(0.159)* | *(0.089)* | *(0.169)* | *(0.153)* |
| D(Average education^2) | -0.008 | -0.009 | -0.005 | -0.018 | -0.013 |
| | *(0.006)* | *(0.008)* | *(0.005)* | *(0.012)* | *(0.009)* |
| F-test | 0.277 | 0.973 | 4.270 | 0.180 | 1.401 |
| J-test | 0.000 | . | 0.000 | 0.000 | 1.863 |
| p-value | . | 0.485 | . | . | 0.601 |
| **Period dummies** | Yes | Yes | Yes | Yes | Yes |
| **Industry dummies** | No | No | No | No | No |
| **Implied RTS** | 0.735 | 0.765 | 0.848 | 1.098 | 1.005 |
| **Endogenous regressors** | Log employment | Log employment | Log employment | All factors of production | All factors of production |
| **Instrumental variables** | D(Log employment) (alternative measure) | D(Log employment) (alternative measure, lags 0 & 1), D(log employment) (lags 1 & 2) | D(Log wage) | D(All factors of production) (lag 1) | All factors of production (lag 1), D(log wage), D(bond yield), D(union membership) |

Notes: *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

## 7.5 List of frequently used abbreviations

| | |
|---|---|
| 2FE | Two-way fixed effects |
| 2SLS | Two-stage least squares |
| ATE | Average treatment effect |
| BH | Belzil and Hansen (2002) |
| CF | Control function |
| CMG | Common correlated effects mean group |
| CRC | Correlated random coefficient |
| CS | Cross-section |
| CSG | Child support grant |
| DCDP | Discrete choice dynamic programming |
| DMG | demeaned mean group |
| DoE | Department of Education |
| DW | Durbin-Watson |
| EA | Enumerator area |
| FD | First-differenced |
| FE | Fixed effects |
| FET | Further Education and Training |
| IMR | Inverse Mills ratio |
| IV | Instrumental variable |
| LATE | Local average treatment effect |
| LFS | Labour Force Survey |
| MG | Mean groups |
| MLE | Maximum likelihood estimator |
| NTC | National technical certificate |
| OAP | Old-age pension |
| OHS | October Household Survey |
| OLS | Ordinary least squares |
| POLS | Pooled ordinary least squares |
| RTS | Returns to scale |
| SARB | South African Reserve Bank |
| StatsSA | Statistics South Africa |
| SUR | Seemingly Unrelated Regression |
| TE | Time effects |