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

Rulof Burger[1]

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 extends that of Belzil and Hansen (2002) by allowing 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 decomposition technique that operationalises the notion of an option value that arises due to non-linearities in the schooling-earnings profile or the sequential resolution of uncertainty, as discussed in Heckman, Lochner, & Todd (2006). 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 also reduced, and is shown to be very nearly linear once we allow for agent heterogeneity. 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.

[1] Department of Economics, University of Stellenbosch and The Centre for the Study of African Economies, University of Oxford

1. Introduction

The internal rate of return to schooling 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 parameters 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 used to estimate this parameter are fundamentally flawed and that the returns parameter remains "widely sought after and rarely obtained".

A number of concerns have been raised with respect to interpreting the Mincerian earnings regression schooling coefficients as the benefit to an additional schooling year, including the endogeneity of schooling, the role of dynamics and uncertainty in the schooling process and the fact that such estimates are not required to be consistent with observed educational investment decisions. The latter problem is exacerbated by recent findings that the schooling-earnings profile has become increasingly convex in a number of countries. If individuals choose schooling so as to maximise lifetime earnings net of schooling costs, then 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 a corner solution.

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 individual heterogeneity and the underlying process that determines schooling outcomes. However, such studies only consider the effect of schooling on earnings, and may therefore produce estimates that are inconsistent with individual schooling decisions. In fact, structural econometricians have reported that the high return estimates produced by this "experimentalist" approach are difficult to reconcile with so many individuals leaving school as early as they do (Keane, 2010, p. 18). A second and very different approach uses dynamic programming methods to explicitly model the individuals' schooling investment decisions and labour market outcomes. Unlike the IV approach, this method requires the schooling returns to be consistent with both the earnings distribution and observed schooling outcomes. Not surprisingly, such 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, it 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. Secondly, individual schooling decisions are analysed using a decomposition technique that calculates the relative importance of different benefits of additional schooling, including concepts such as the *option value* and the *sequential resolution of uncertainty* that are emphasised in Heckman, Lochner & Todd (2008). Thirdly, 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. 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.

2. Discrete choice dynamic programming studies of schooling returns

Accurate estimates of the rate of return to schooling are vital for those attempting 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)).

Concerns regarding the endogeneity of schooling in an earnings equation (Denison, 1964; Spence, 1973; Griliches, 1977; Willis & Rosen, 1979; Card, 1999) have led a popular strand of the empirical schooling returns literature to use supply side variation of the schooling decision as instrumental variables in order to produce consistent estimates of the effect of schooling on wages. Such studies are generally uninformative about the shape of the schooling-earnings profile, and effectively treat individual heterogeneity as a nuisance parameter. The fact that IV studies generally find schooling returns that are somewhat higher than suggested by OLS (Ashenfelter et al. (1999)) is often interpreted as evidence that ability bias is a relatively minor problem compared to measurement error, although Card (1999) offers an alternative explanation which is based on the assumption of a concave schooling-earnings profile. The perceived robustness of the experimentalist approach partly derives from the fact that the econometrician does not need to know anything about how schooling decisions are made, apart from making the instrument validity assumption and testing instrument relevance. Unfortunately, this agnosticism about the schooling decision means that there is no framework to consider whether the estimated schooling benefit is consistent with observed schooling decisions. In fact, structural models that jointly estimate the earnings function and schooling decision process usually find it difficult to reproduce the high returns estimates from IV studies given the large share of individuals who choose to obtain relatively low schooling levels.

A number of recent descriptive studies suggest that the schooling-earnings profile has turned increasingly 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 possibility that the schooling return is low initially, and then increases as students progress up the schooling ladder would make observed schooling decisions even more puzzling. Most individuals choose to cease investing in education somewhere during the intermediate schooling years, which would imply that most agents obtain the low-yielding schooling levels but then drop out just as the benefits start to materialise.

What we understand as the value of an additional schooling year also changes when schooling investment is viewed as a dynamic process rather than a one-time static decision. Heckman, Lochner and Todd (2006) demonstrate how the returns to schooling can include an option value if i) there is uncertainty about future wage offers or schooling costs that are only revealed as one advances through the schooling years, or ii) the schooling-earnings profile is convex. Of course, the standard Mincerian earnings function that assumes earnings to be log-linear in schooling and ignores the role of uncertainty will be neglect this option value. In such cases, the Mincerian schooling returns parameter may be unable to square the apparent benefit of an additional schooling year with the revealed schooling preferences of individuals.

The last two decades have seen the emergence of a modelling approach that allows us to incorporate all of the above issues in a single model of the schooling-earnings nexus. 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 weighted 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.

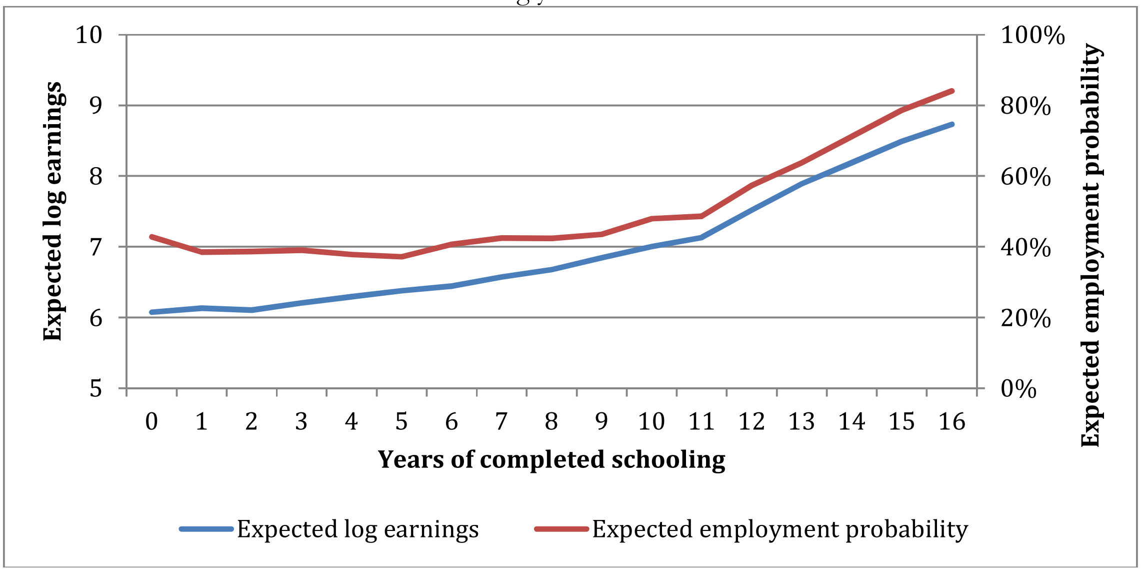
3. Schooling, earnings and employment in the South African labour market

Before delving deeper into the nature of DCDP models we briefly present some descriptive statistics as further motivation for this analysis. Figure 1 plots the conditional expectations of log earnings and the employment share[2] for black males who have completed different schooling years, after correcting for experience and experience squared. 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 5) 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).

[2] 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.

Figure 1 reveals that expected log earnings increases by approximately 7% for each additional year of primary school, by 14% for each year of incomplete secondary education, and by 32% for completing secondary schooling and each year of post-secondary education. Given the observed schooling-earnings pattern, it is no surprise that schooling enters the Mincerian earnings equation as a convex function[3]. The schooling-employment share profile is similarly convex: there is no increase in the expected employment likelihood for primary school, a 2% increase for each year of incomplete secondary education and 7% for completed secondary education and higher.

Figure 1: Conditional expectation of log earnings and probability of being employed for black males, by schooling year: 2001-2004



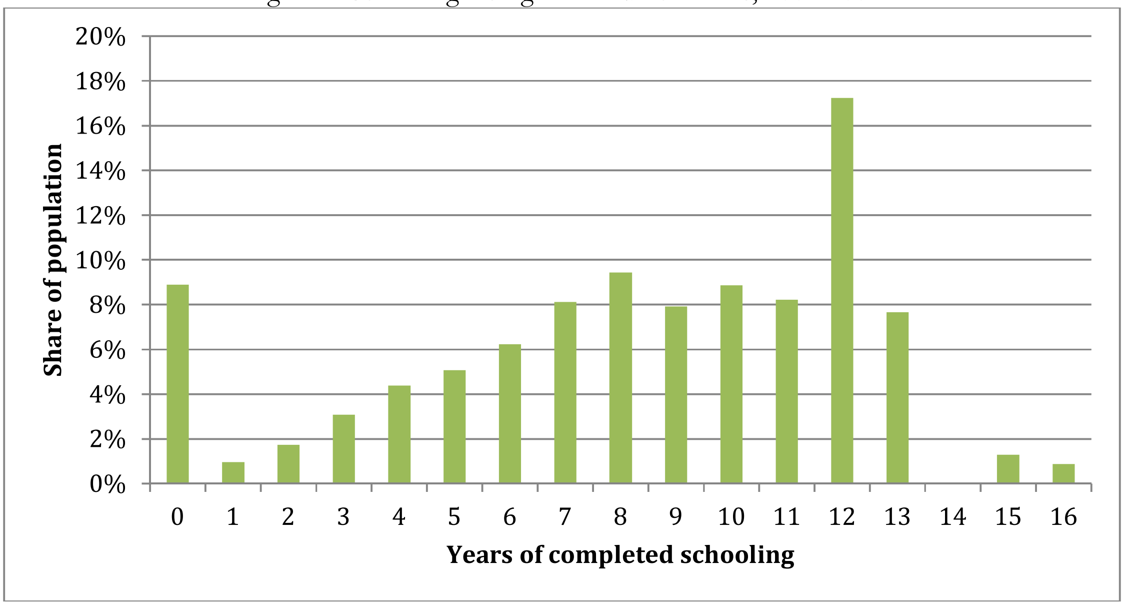
Source: Author's calculations from Labour Force Survey panel (StatsSA: various years).

As discussed in section 2, the schooling coefficients from an earnings function are also important for 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. Figure 2 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

[3] Studies that include schooling as a quadratic function find that the marginal rate of return increases by between 0.8% (Kingdon & Knight, 2006) and 8% (Keswell & Poswell, 2004) per schooling year

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 1. 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 6.1.1 investigates the extent to which these two outcomes can be explained by a single schooling-earnings model.

Figure 2: Schooling histogram for black males, 2001-2004

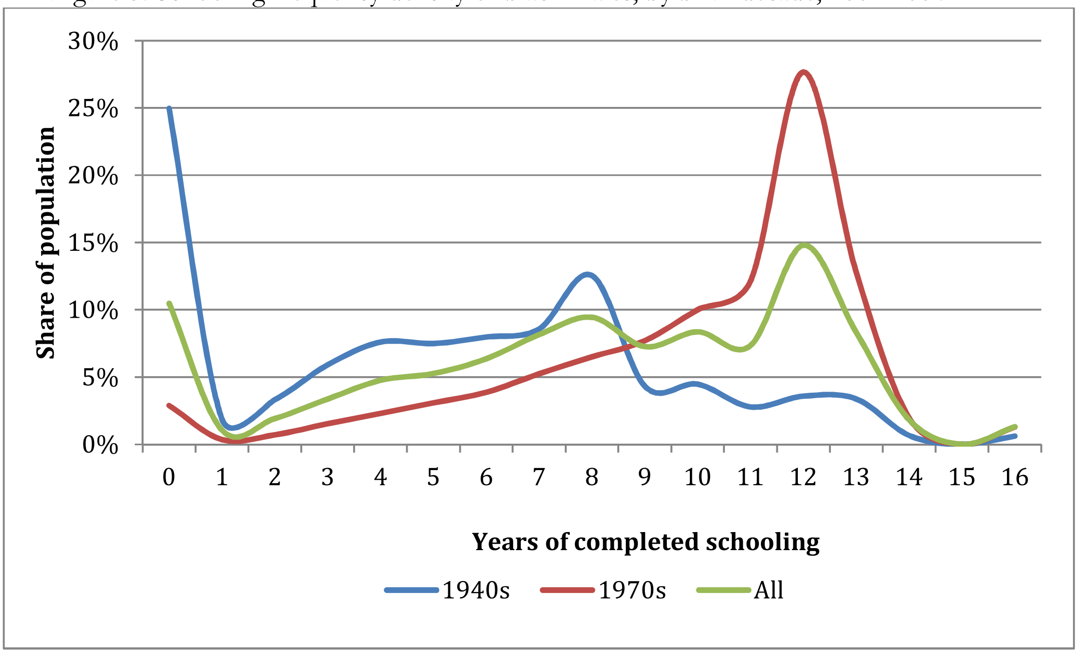


Source: Author's calculations from Labour Force Survey panel (StatsSA: various years).

Figure 3 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 2 and juxtaposed those for individuals 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 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 3: Schooling frequency density of black males, by birth decade, 2001-2004



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

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

4.1 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 Belzil-Hansen model, although some assumptions are altered in order to reflect specific features of the South African labour market (as discussed in section 3) and to aid the interpretability of the model parameters. Utility derived by individual i from being in school in period t is expressed as

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

where z_i is the individual's birth year, s_{it} is schooling years completed, $\eta^s_{g(i)}$ is the schooling utility intercept which is specific to the individuals of type g, and ε^s_{it} is a schooling utility shock.

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 Belzil-Hansen model. One important difference is that whereas schooling enters their schooling utility through a spline function, this is replaced by a piecewise constant (or stair) function with discontinuous jumps between primary, secondary and tertiary education. Their spline function allows for a very flexible relationship between school year and school cost, and 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. The ψ_{21} and ψ_{22} coefficients in equation (1) represent the marginal utility that individuals expect to experience while in secondary and tertiary education respectively, relative to being in primary schooling. These changes are presumed to primarily reflect the differential schooling costs – either monetary or psychic – associated with shifting between these different schooling types.

In estimating our model we use birth year as a school utility shift variable. Figure 3 demonstrated the importance of birth year in determining schooling outcome. This is to be expected given South Africa's history of apartheid-era restrictions on black schooling enrolments, where the gradual relaxing of these policies towards the end of apartheid created variation in the schooling outcomes amongst black men born in different years. Furthermore, our panel dataset (discussed in section 5) does not contain any of the family human capital variables used in Belzil-Hansen, which restricts the range of candidate variables that can be included as schooling shifters.

Our model allows for persistent individual heterogeneity by allowing individuals to belong to one of G different types (indexed with g) in the spirit of Heckman and Singer (1984). 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.

Labour market participants derive utility according to

$$u^w(.) = m_{it} \ln W_{it} \qquad [2]$$

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 Belzil-Hansen 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 inappropriate for our analysis. 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 the Belzil-Hansen model

$$\ln W_{it} = w_{it} = \chi_1 s_{it} + \chi_2 s_{it}^2 + \chi_3 x_{it} + \chi_4 x_{it}^2 + \eta_{g(i)}^w + \varepsilon_{it}^w \qquad [3]$$

where x_{it} is potential years of work experience, $\eta_{g(i)}^w$ is a type g-specific earnings intercept and ε_{it}^w represents a stochastic earnings shock.

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

$$m_{it} = 1\left(\kappa_1 s_{it} + \kappa_2 s_{it}^2 + \kappa_3 x_{it} + \kappa_4 x_{it}^2 + \eta_{g(i)}^m + \varepsilon_{it}^m > 0\right) \qquad [4]$$

where $\eta_{g(i)}^m$ represents the type g-specific employment intercept and ε_{it}^m is a stochastic employment shock. Unlike the Belzil-Hansen model, we allow schooling to enter this function as a quadratic function, and also for heterogeneity to affect the employment likelihood.

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), 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. Individuals are not allowed to re-enter the education system once they have entered the labour market.

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 (3) is a representation of how firms allocate job offers. Secondly, our model also discards the grade repetition feature of the original Belzil-Hansen model. Although South African schools have notoriously high repetition rates, the more relevant question is whether incorporating this feature helps us understand the relationship between schooling and earnings. The Belzil-Hansen 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, the lack of comparable household survey data going back that far means that the generational variation in earnings are required to identify the experience coefficients so that there is no way to test whether this assumption is valid or not. Although we therefore cannot rule out that the convex earnings and employment profiles observed in Figure 1 are very recent phenomena, it is worth noting that the schooling decisions of older cohorts (many of whom chose to obtain very low levels of education) is easier to reconcile with the observed convex schooling-earnings profile than that of their younger counterparts.

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 restrictions other than a lifetime budget constraint. 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 liquidity 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 assumptions of the standard human capital earnings model that schooling decisions are made by dynamically optimising agents with rational expectations.

4.2 Solving the agent's dynamic optimisation problem

The choice-specific utility functions [1] and [2] 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 Belzil-Hansen 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 deterministic 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 $P(d = 1|s_{it}, z_i) = P(V_{it}^{s}(\boldsymbol{a}_{it}) > V_{it}^{w}(\boldsymbol{a}_{it})|s_{it}, \boldsymbol{z}_i)$.

The deterministic component of the schooling value function $\bar{V}_{it}^{s}(s_{it}, z_i)$ is solved using backwards induction and Monte Carlo integration in order to calculate the three-dimensional integral across non-normally distributed random variables required to calculate $\beta Emax\{V_{it}^{w}(s_{it} + 1), V_{it}^{s}(s_{it} + 1, z_i)\}$ (following the recommendation of Keane and Wolpin (1994)). The conditional choice probabilities

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

are then easily computed since the Monte Carlo integration procedure already simulates the values of e_{it}^{w} and ε_{it}^{s}. These probabilities are combined with the wage density and employment likelihoods to evaluate the likelihood function for a parameter vector $\boldsymbol{\theta}$.

4.3 Estimating the structural parameters

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. Calculating the wage densities $L_i^{w}(\boldsymbol{\theta}_w, \eta^{w})$ and employment probabilities $L_i^{m}(\boldsymbol{\kappa}, \eta^{m})$ for a specific set of parameter values $\boldsymbol{\theta}_n$ under our distributional assumptions is straightforward. In order to calculate the likelihoods associated with the schooling outcomes, we first discretise the birth year variable, z_i into 10 equidistant 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 (Corona, Marchesi, Martini, & Ridella, 1987). 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, the process terminated at nearly identical coefficient estimates in different runs, which suggests convergence on a global optimum.

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 $L(\boldsymbol{\theta})$.

5. 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. 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 (as noted in Vermaak (2010)). 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. 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 South African panel dataset and one of very few African panel datasets with enough observations to estimate this sort of model. 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.

6. Estimation results

6.1 Homogeneous agents case

The model outlined in section 4.1 is now estimated using the methods discussed in sections 4.2 and 4.3. 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 1). The exogenous schooling model is essentially a static version of our model in which only the employment and earnings equations 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 – 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 decrease expected earnings but the marginal returns increase by 2.4 percentage points for each additional schooling year.

Table 1: 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 1. 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 (as was also reported by Keane (2010) for the US data). 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 1 can perhaps be viewed as a "compromise" solution between these two forces.

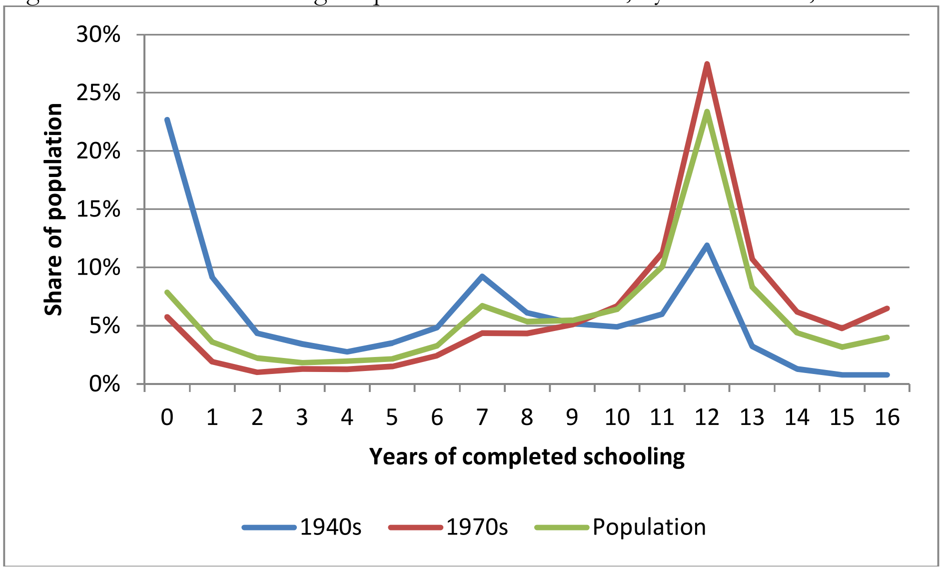
The first thing to note is that – unlike the DCDP models estimated on US data – the average schooling returns are high. Someone with 8.15 years of schooling (the mean for our sample of black males) can expect to earn a 17% return on his next year of schooling investment. Compared to the static baseline 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.

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 the effect on 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 derive more immediate utility from entering the labour market with no hope of finding work than to enter tertiary education. We believe 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 relaxed compared to those faced by the reference birth cohort.

Figure 4: Predicted schooling frequencies of black males, by birth decade, 2001-2004



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

6.1.1 The predicted schooling distribution

One way to test 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 4 plots the predicted schooling distributions for the population as a whole, for those born in the 1940s and for those born in the 1970s. Despite the relatively parsimonious specification of our model, it is able to replicate the stylised facts of the schooling distribution as reported in section 3 and graphed in Figure 3. 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.

6.1.2 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 reasoning behind individual schooling decisions. 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.

Heckman, Lochner and Todd (2006) discuss how both convexity in the earnings profile and uncertainty about future schooling cost and wage offer error can create an option values for lower schooling years. These option values may be crucial in explaining why so many individuals choose to complete these levels of schooling despite their small effect on expected earnings. Although our model uses a different estimation strategy than that suggested in Heckman, Lochner and Todd (2006), it does allow us to determine the magnitudes of these effects using the decomposition technique outlined below.

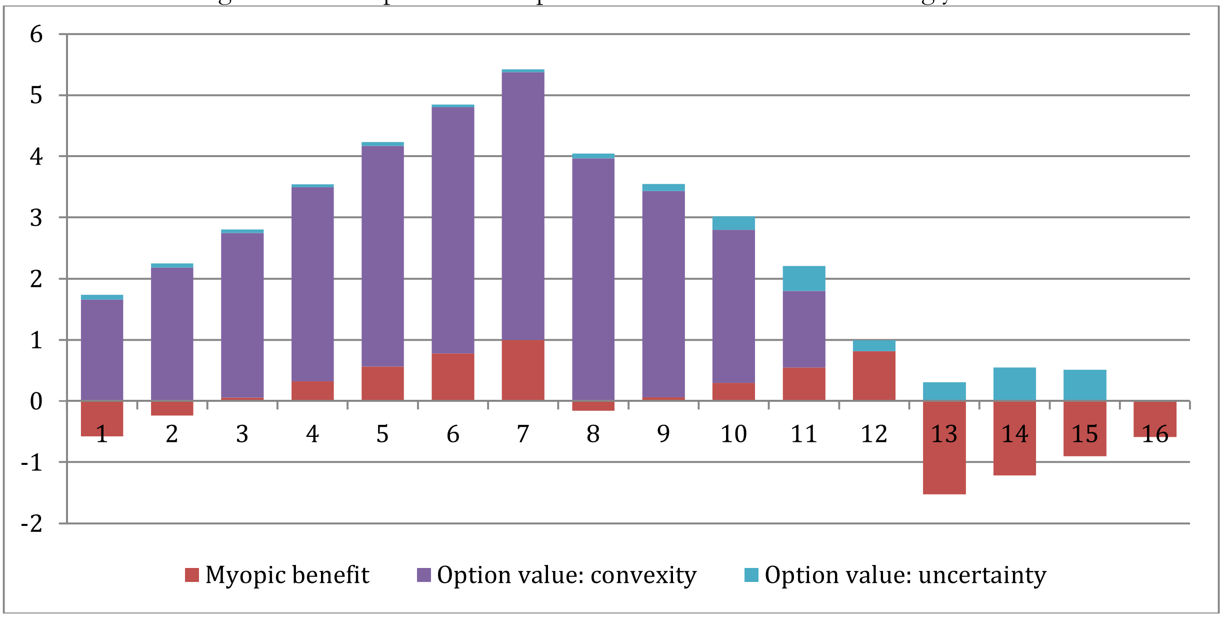
Agents in our model decide whether to invest in another schooling year by comparing the value of staying in school to the value of entering the labour market $V_{it}^{s}(s_{it}, z_i, \varepsilon_{it}^{s}) - V_{it}^{w}(s_{it}, \varepsilon_{it}^{m}, \varepsilon_{it}^{w})$. This term can be expanded into 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, and 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 a schooling models with a concave or linear schooling-earnings profile or certainty about future wage and school utility shocks, this term is the only deterministic explanation for the variation in schooling outcomes. 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 represents the option value associated with the non-linearity of the schooling profile 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 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 exceed the maximum of their expected values, this "sequential resolution of uncertainty" term will always be positive. 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 deterministic components of the net schooling benefit across the different schooling levels gives us the graph in Figure 5 (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 5 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.

Figure 5: Decomposition of expected value of additional schooling year



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

During primary and most of secondary school, the relatively small myopic net benefit components are dominated by the larger positive option value attached to access to the convexity of the schooling-earnings profile. 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.

During early schooling levels the option value attached to the sequential resolution of uncertainty 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 convexity 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 convexity 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 option value of the resolution of uncertainty 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.

6.2 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 1. 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 return for someone at the mean schooling level is still very high: approximately 16%. The schooling-earnings profile is only marginally less convex than was suggested by an exogenous schooling MLE model, but the addition of heterogeneity transforms the schooling-employment profile to something that is very nearly linear,.

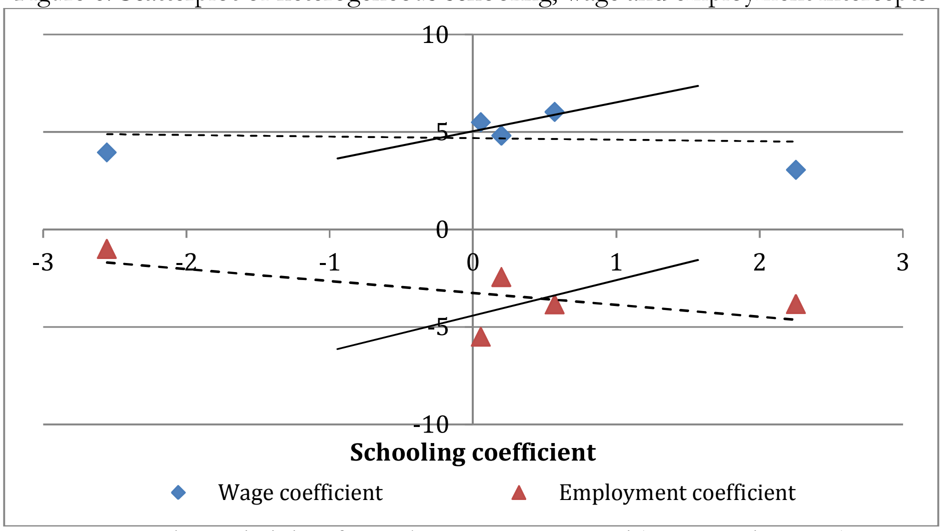
Table 2: Heterogeneous intercept estimates from DCDP model

	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 2, along with the population shares that each type represents, and plotted in Figure 6. 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 6) 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, only 13% of individuals belong to either of these two groups. 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 6.

Figure 6: Scatterplot of heterogeneous schooling, wage and employment intercepts



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

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. 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 found when estimating a static exogenous schooling model. 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.

8. Bibliography

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

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

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.

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). Education and the Distribution of Earnings. *The American Economic Review , 56* (1/2), 358-369.

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

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.

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.

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.

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.

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

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.

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

Heckman, J. J. (2010). Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature , 48* (2), 356-398.

Heckman, J. J., & Singer, B. (1984). A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica , 52* (2), 271-320.

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. 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.

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

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

Keane, M. P., & Wolpin, K. I. (1994). 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.

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

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

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

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.

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

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

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

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.

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

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

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

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

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

Willis, R. J., & Rosen, S. (1979). Education and Self-Selection. *Journal of Political Economy , 87* (5), S7-36.