REDI3x3 Working paper 53 April 2019

Convexity or heterogeneity? Estimates of the shape of the earnings profile

Rulof Burger and Francis Teal

Abstract

A number of empirical studies have found convex schooling-earnings profile is in various countries, especially in the developing world. This is usually interpreted as evidence of a very high demand for highly skilled workers, but can also reflect heterogeneous marginal returns to schooling that are positively correlated to number of years of completed schooling. This paper attempts to distinguish between convexity in the earnings profile and heterogeneity in the schooling returns by investigating a country with notoriously high inequality in the quality of schools and family background: South Africa. We use a natural experiment in education policy to estimate the non-linear earnings profile while allowing for endogenous schooling and heterogeneous returns. The results of our control function estimates suggest that, unlike what was found by a number of OLS studies, the South African schooling earnings profile is actually very close to linear and perhaps even concave. We also find evidence of substantial heterogeneity in the curvature of this profile, which may reflect the high levels of inequality in school quality and family background. The results suggest that individuals with low returns end up with fewer schooling years, while high-return individuals choose to complete more years of schooling..

KEYWORDS: Income mobility, inequality, longitudinal data analysis, measurement error
JEL CODES: J62, D63, C23

The **Research Project on Employment, Income Distribution and Inclusive Growth** is based at SALDRU at the University of Cape Town and supported by the National Treasury. Views expressed in REDI3x3 Working Papers are those of the authors and are not to be attributed to any of these institutions.

Convexity or heterogeneity? Estimates of the shape of the earnings profile

Rulof Burger (Stellenbosch University) & Francis Teal (University of Oxford)

1. Introduction

There is a substantial body of recent descriptive evidence, spanning a large number of countries, which suggests that the returns to schooling increases with schooling, i.e. the schooling-earnings profile is convex. The evidence for developing countries, and South Africa in particular, has been particularly conclusive in this regard. However, most of this literature uses estimators that exploit identifying assumptions deemed implausible by contemporary standards, specifically that schooling is exogenous in the earnings function and that schooling returns are homogenous across individuals. Although there are now a large number of studies that have estimated the return to endogenous schooling while allowing for individual heterogeneity in returns, such estimates are usually interpreted as local average treatment effects that are uninformative about the shape of the earnings profile.

Substantial convexity in the earnings profile appears to suggest that the labour market demands highly skilled workers, and that public policies aimed at improving the prospects of young labour market entrants should focus on improved access to higher education. However, this apparent convexity may also be a statistical artefact produced by heterogeneity in the returns to schooling due to variation in school quality, household or neighbourhood differences in preparing learners for school, or individual differences in inherent abilities or motivation. If individuals who receive high yielding schooling years also tend to stay in school for longer, then this will produce the same convex schooling-earnings correlation pattern that have now been found in many recent studies. However, the public policy implications from such a model are very different from those usually motivated by a convex earnings profile. Policies that improve access to schooling without simultaneously addressing the low quality of schooling or household inequality will be ineffective in addressing labour market inequality. It is instructive in this regard to note that many African countries, including South Africa, have made substantial progress in improving access to education in recent decades, but there is little evidence that successive generations of better educated labour market entrants have reaped the benefits of their greater investment in education.

The only way to distinguish between marginal schooling returns that truly increase in schooling years and heterogeneous returns that are positively correlated to schooling is to use a statistical technique that explicitly allow for endogenous schooling, heterogeneous returns and a non-linear schooling-earnings profile. The control function estimator can be used to allow for all of these features, but requires instrumental variables that affect schooling outcomes at various schooling years. It also requires identifying assumptions that are more restrictive than those necessary for identifying the local average treatment effect. These assumptions rule out using a linear first-stage regression of an endogenous variable with discrete characteristics, like schooling, and require accurately modelling the conditional expectation of schooling that fully reflect any potentially non-linear effects of the instrumental variables on schooling outcomes.

South Africa has notoriously high inequality in school quality and household income as well as a highly convex schooling-earnings profile, and therefore offers a promising context to explore the validity of the competing explanations for the observed correlation pattern between schooling and earnings. This paper will use the exogenous variation in schooling that derives from two education policies implemented by the Department of Education in the late 1990s. These policies aimed to reduce the number of over-aged learners in the school system, and had the effect of reducing grade repetition rates and decreasing enrolment amongst older learners. The time-, grade- and age-varying enrolment and promotion rates from a structural dynamic schooling attainment equation are estimated using a minimum distance estimator with a flexible specification. We argue that these estimates are more likely to accurately capture non-linearities in the conditional expectation of educational attainment that would invalidate the use of a first-stage schooling regression that linearly projects schooling onto exogenous variables.

The results of our control function estimates suggest that the South African schooling earnings profile is actually very close to linear and perhaps even concave. We also find evidence of substantial heterogeneity in the slope of this curve, which is consistent with other studies that have found very high levels of inequalities in household background characteristics and school quality. The results suggest that individuals with low returns end up with fewer schooling years, while high return individuals choose more years of schooling. This interaction produces what seems like a convex schooling-earnings profile when using estimators that do not allow for heterogeneous returns or endogenous schooling outcomes.

2. Literature review

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

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

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 represents the schooling-earnings profile.

Studies in the early empirical schooling returns literature typically used some variant of Mincer's (1974) human capital earnings function to estimate the returns to education using ordinary least squares (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 interpreted as the causal effect of an additional year of completed schooling 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 this approach are well documented (Denison (1964), Gronau (1974), Lemieux (2006), Heckman, Lochner, & Todd (2006), to mention a few). Three of the most important limitations of the conventional Mincerian approach include the possible correlation of schooling and unobserved ability, heterogeneity of returns across schooling years, and heterogeneity of returns across individuals.

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 (G. Becker, S., 1964; Mincer, 1958). This criticism has led to large 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 regressions. 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[1]. 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.

The early empirical literature essentially attempted to estimate *the* return to additional year of education, implicitly based on the notion that this rate is the same across different individuals and

[1] 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%.

levels of schooling. The 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[2] 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 (1973, 1985, 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 explicit 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 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.

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

[2] Mincer (1974) experimented with alternative functional forms, including one that included the years of schooling as a quadratic function. 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.

conditions are discussed in more detail in section 2.3, but are in many respects 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.

3. The Becker-Card model

Our literature survey highlighted three 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, and returns that vary across individuals. In order to aid the interpretation of the estimates it is useful to derive the estimable equations from an explicit theoretical model (Heckman, 2010). Towards this end we briefly review the Card (1999, 2001) model, which allows for endogenous schooling decisions and a non-linear schooling-earnings profile that varies across individuals.

The literature survey in section 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 be 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. 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.

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. 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 $\kappa_1 > 0$. 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. 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[3]

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

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

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

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 "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[4] 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

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

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

[3] 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.
[4] The second-order sufficient condition requires that the slope of the marginal cost curve is larger than that of the marginal benefit curve.

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.

4. Identification

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 κ_1 parameters from equation [5]. These coefficient estimates will allow us to construct the expected schooling-earnings profile for a randomly selected individual.

Suppose that the earnings function is given by equation [5] and that individuals' schooling decisions are represented by equation [4]. 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 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 $\hat{u}_{it} = s_i - \boldsymbol{z}_{it}\widehat{\boldsymbol{\pi}}$ 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. 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 $\mathbf{z}_{it}$.

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 $E(u_{it}|\mathbf{z}_{it}) = 0$, $E(u_{it}^2|\mathbf{z}_{it})$ 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 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.

However, the 2SLS estimator requires that there are at least as many IVs as endogenous regressors. In such cases 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 [4], 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 parameters $(\boldsymbol{\delta}, \bar{b}, \kappa_1)$.

1. (a_{it}, v_{it}, u_{it}) is independent of $\mathbf{z}_{it}$,
2. $\text{rank}\, E\big(\mathbf{z}_{it}'(\mathbf{z}_{it1}, s_i)\big) = K_1 + 1$,
3. $E(a_{it}|u_{it}) = \rho_1 u_{it}$,
4. $E(v_i|u_{it}) = \rho_2 u_{it}$

Assumption 1 imposes a "substantive restriction on the structural and reduced form error terms"(Wooldridge, 2007). Firstly, requiring $\mathbf{z}_{it}$ 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 $\mathbf{z}_{it}$ 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. Even though equation [4] is typically referred to as the reduced form equation, the assumption that u_{it} is independent of $\mathbf{z}_{it}$ comes close to assuming that we have correctly specified the structural equation for s_i. This assumption also rules out linear first-stage equations 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 $\boldsymbol{z}_{it}$ precludes any systematic variation between schooling and the returns that are not already captured by the quadratic schooling term.

The typical instrument informativeness assumption, which requires that our instrumental variables explain some variation in the schooling variable that is not already explained by the control variables, is stated as assumption 2. Assumption 3 specifies the precise functional form of the relationship between the structural and reduced form errors. When combined with assumption 1 and the form of equation [4] 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 4 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 assumptions 3 and 4 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 $\boldsymbol{z}_{it}$ rules out using the CF approach outlined above unless the endogenous variable is continuous. A similar issue arises when using 2SLS to estimate the ATE of endogenous schooling on earnings. Technically, this means that these approaches are not appropriate when we are interested in identifying the causal effect of schooling. However, there are examples of studies that have evidently considered the schooling variable to be sufficiently continuous to exploit this independence assumption, such as Heckman and Vytlacil (1998) and Söderbom et al. (2006). Even if we were comfortable ignoring the dependence between u_{it} and $\boldsymbol{z}_{it}$ that derives from the discrete nature of schooling outcomes, this approach clearly requires that our schooling equation is correctly specified and completely captures the complete effect of the instrumental variables on all the moments of the schooling variable. These more restrictive assumptions, which are not required for estimating local average treatment effects of schooling, are a crucial part of how we obtain answers to the more ambitious research question of estimating the causal effects of different schooling years on expected earnings. Persuasively answering this question requires scrutinising whether these stronger assumptions are valid, and perhaps using more appropriate techniques if necessary..

5. South African schooling policy

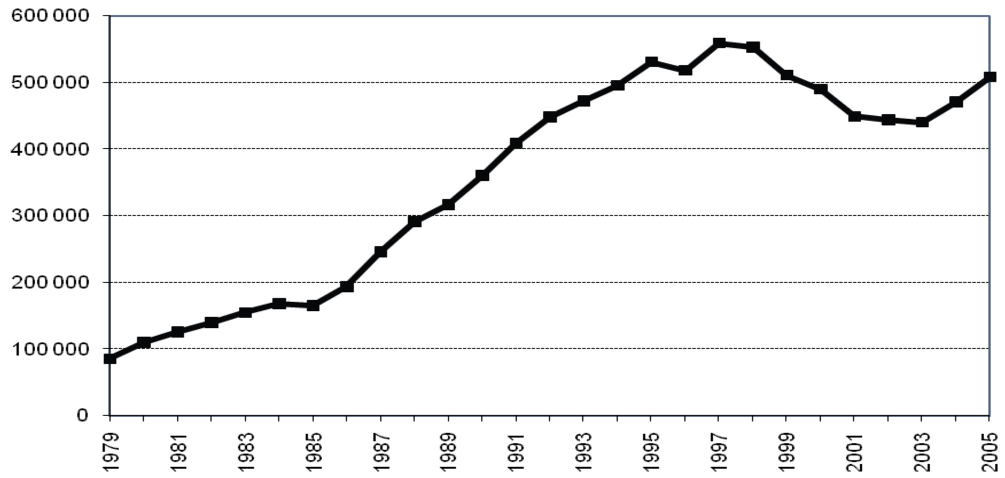
A high share of South African youths have historically chosen to stay in school for long beyond the normal school-going age due to late school entry, high repetition, drop-out and drop-in rates and the high unemployment rate amongst young workers with less than completed secondary education. In 1998 60% of Grade 12 learners were older than the correct grade-age (Guluza & Hoadley, 1998, p. 1). A high proportion of the over-age children remaining in school were black and lived in rural areas, perhaps the group that was also most likely to struggle to find employment otherwise. This presence of a large group of over-aged learners contributed to maintaining the already high class sizes in schools, which was perceived to impose a negative externality on pupils of the correct grade-age while tying up scarce resources in the education system. In an attempt to rectify this situation, the Department of Education phased in restrictions on over- and under-aged learners in the late 1990s, as well as limiting the number of times a student could be held back.

Shortly after political transition the 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) with such a policy to 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 suggests a plan to phase in such a policy progressively by geographical regions. In 1998 the Minister of Education published *Age Requirements for Admission to an Ordinary Public School*, which defined the appropriate age for admission to a certain grade as "the grade number plus 6" (Department of Education, 1998). However, the notice itself seemed to have been 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. One is therefore left to infer much about the implementation details from school data, the policy documents of provincial education departments and other research papers. One document of the Department of Education hints at what the policy intended: "Provisions for conditions of admission of learners to public schools as well as age grade norms are further elaborated on in the Admissions Policy for Ordinary Public Schools (Department of Education, 1998) that came into effect in January 2000. By Grade 9, which marks the end of compulsory basic education, the learner should be 15 years old. Recognising that the problem of over-age learners will not be eliminated immediately, the policy states that the onus will be on schools to place learners who are above the normal age for a grade in a 'fast-track facility' to help bring them in line with their peer group. Learners over the age of 16 wanting to attend school will be referred to adult education centres." (Department of Education: 2000, p.24). Although this policy was only formally promulgated in 2000 at national level, it appears to have been informally accepted as policy in some parts of the system before that. 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 had been uniformly enforced in all schools.

At about the same time, the Department of Education instituted new promotion and repetition policies. According to the OECD review of South African education, in 1998 "an admissions policy was issued that set norms for learners to proceed through school with their age cohort", "in order to improve the internal efficiency of the education system" (OECD 2009, p.51). These policies specified that any learner could only repeat once during any of the four education phases, i.e. the foundation phase (Grades 1 to 3, although Grade R has now been added to that), intermediate phase (Grades 4 to 6), the senior phase (Grade 7 to 9) or the further education and training phase (Grades 10 to 12) (OECD 2009). The policy is generally less strictly applied in Grades 1 and 2, where issues of under-age entry into schools and school readiness enter decisions on promotion and retention. The policy also in practice excludes Grade 12, as passing Grade 12 depends on success in the matriculation examination. On the other hand, it is not clear that the "weeding" that is observed in the system (holding back weak performing learners in Grade 11 to improve the matric pass rate) may imply some deviation from the formal promotion policies in the further education and training phase.

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: 304) ascribe to "the natural saturation of the system, and education department policy to limit under-age enrolment in Grade 1 and excessive repeating of all grades". Similarly, the number of matric candidates decreased by fully 20% between 1998 and 2003 (Figure 1) – this trend could not be affected by the policy not to admit under-age learners into Grade 1, and could therefore highlight the exit of over-age learners. Provincial policy documents indicate that by 2003 the directives on over-age children had already been adopted, along with the promotion of the FET system (Western Cape Education Department, 2003). According to documents from that province, schools in the Western Cape were to deny admission to learners who were more than two years older than the appropriate grade-age; similar policies seemed to have applied in other provinces. 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. Due to the higher repetition rates in historically black schools, this policy should by implication have affected black learners (both boys and girls) more directly than white learners.

Figure 1: Number of matric candidates, by year: 1979-2005



Source: Own calculations from information obtained from Department of Basic Education

A factor that may also have contributed to the implementation of the over-age policy, or to the choice of children to leave the school system, was the increased pressure applied by the education authorities on schools to improve their matric pass rates. Many schools responded to this by failing weaker learners in Grade 11 to reduce potentially high matric failures, despite the repetition policy that frowns on much repetition of grades in the school system. In addition, for similar reasons, schools became less accommodating towards children, particularly older ones, who had failed matric and wanted to repeat. One result may have been that many children may earlier have given up hope of achieving matric, thus also encouraging earlier entry into the labour market. The effect of the combination of over-age restrictions and weeding on matric data is evident from Figure 1, with the strong upward trend in the number of candidates due to improved progression through the school system more than cancelled by these phenomena in the late 1990s.

6. Dynamic model of schooling progression

The control function estimator discussed in section 4 provides a way of identifying the causal effect of different schooling years on expected earnings by estimating the schooling-earnings profile. However, we also discussed the more restrictive assumptions that are required for this approach to produce consistent estimates. There are legitimate concerns that simply using a linear first-stage schooling regression that ignores the discrete nature of schooling outcomes and the potentially nonlinear effect of the instrumental variables may violate the identifying assumptions and hence provide a questionable basis for identification.

The preceding section discussed a pair of schooling policies that should have induced differences in schooling decisions of young South Africans from different birth years. This is precisely the kind of exogenous variation in schooling outcomes – spread across different schooling years –

that may help us estimate the schooling-earnings profile consistently. Ideally, we would like to find plausible assumptions about the precise nature in which the schooling policies affected schooling outcomes that would help us choose the functional form of our first-stage estimator. In order to identify the effects of decreasing over-aged enrolment and higher promotion rates, we need to estimate both the enrolment and promotion rate for individuals of different ages, with different levels of completed schooling and across the various survey years. Although individual enrolment decisions and educational attainment are observable in the series of cross-sectional data sets, there exists no nationally representative panel that can tell us whether or not specific individuals successfully progressed to the next level of schooling during the period under consideration. Promotion rates are therefore inherently unobservable in the available data. However, our series of household surveys provides information on the distribution of educational attainment and enrolment rates for a specific generation of individuals at different points in time. Clearly the evolution of educational attainment for a birth cohort of individuals must be affected by promotion rates. We exploit the underlying relationship between grade promotion, enrolment and attainment in order to back out estimates of these promotion rates.

In order to formulate a dynamic equation for schooling promotion, we start by defining a few variables. In period t, $N_{c,t}$ South Africans were born in birth year (or cohort) c, and $N_{c,s,t}$ of them had completed exactly s years of schooling. The share of cohort c individuals with s years of schooling that chose to enrol in school during period t is $\theta_{c,s,t}$, and proportion $\lambda_{c,s+1,t}$ of these individuals successfully completed this next schooling year.

In the beginning of year t there are three types of individuals with s years of schooling: i) those who had $s-1$ years of completed schooling at the beginning of period $t-1$ and who successfully completed another grade during this year, ii) those who already had s years of schooling in period $t-1$ but who chose not enrol, and iii) those who had s years of schooling in the previous year and did enrol but failed to be promoted. These dynamics are reflected in the equation[5] for members of cohort c as

$$N_{c,s,t} = \{(1-\theta_{c,s,t-1}) + (1-\lambda_{c,s+1,t-1})\theta_{c,s,t-1}\}N_{c,s,t-1} + \lambda_{c,s,t-1}\theta_{c,s-1,t-1}N_{c,s-1,t-1} \quad [6]$$

The Stats SA household surveys provide us with a sample of individuals' schooling levels, birth years, enrolment decisions[6] and sampling weights in different surveys, which would allow us to

[5] Technically we also need to add restrictions to reflect that grade promotion cannot increase the number of individuals with no schooling or decrease the number with the maximum schooling years.

[6] In order to estimate enrolment and promotion rates for the period prior to 1999 we temporarily define school enrolment as being enrolled in any educational institution but having fewer than 12 years of completed schooling. Note that this definition differs from the more accurate definition used to estimate enrolment in Table 1, but produces a similarly sized decrease in enrolment between 1999 and 2003 of 497,297 (compared to 514,319 for the more accurate definition).

directly estimate $N_{c,s,t}$ and $\theta_{c,s,t}$. Furthermore, equation [6] can be used to express the promotion rate as a deterministic function of the values of $N_{c,s,t}$ and $\theta_{c,s,t}$:

$$\lambda_{c,s+1,t-1} = \frac{\Delta N_{c,s,t} - \lambda_{c,s,t-1}\theta_{c,s-1,t-1}N_{c,s-1,t-1}1(s>0)}{-\theta_{c,s,t-1}N_{c,s,t-1}} \text{ if } s \leq S_{max} \qquad [7]$$

In theory, this equation can be used to estimate promotion rates by replacing the values of $N_{c,s,t}$ and $\theta_{c,s,t}$ by their sample counterparts $\widehat{N}_{c,s,t}$ and $\hat{\theta}_{c,s,t}$. In practice, we found that sampling variation and measurement issues substantially distort the estimates of these promotion rates. Sampling error in $\widehat{N}_{c,s,t}$ simultaneously increases the numerator and decreases the denominator in equation [7], which is particularly problematic for uncommon combinations of c, s and t where it can even lead to estimates of $\hat{\lambda}_{c,s+1,t-1}$ that lie outside of the unit interval[7]. A more promising approach is to simultaneously look for values of $\hat{\theta}_{c,s,t}$ and $\hat{\lambda}_{c,s,t}$ which imply that cohorts progress through the school system in a way that is internally consistent and that closely resemble what we observe in the data. This can be implemented using a minimum distance estimator.

Our estimator requires moments for which we can compare estimable sample and predicted population values: we choose the educational attainment shares[8] $\frac{N_{c,s,t}}{N_{c,t}}$ and enrolment rates $\theta_{c,s,t}$. Our dataset allows us to obtain consistent estimates of the sample counterparts of these moments as $\frac{\widehat{N}_{c,s,t}}{\widehat{N}_{c,t}}$ and $\hat{\theta}_{c,s,t}$. For notational convenience, we denote the vector of population and sample moments by $\boldsymbol{M}$ and $\widehat{\boldsymbol{M}}$ respectively. Suppose the enrolment and promotion rates can be accurately specified as depending on an underlying set of parameters $\boldsymbol{\beta}$ that summarise the changing inclination to enrol in and complete various schooling levels across generations, time and age. Then, by equation [1], we can express the vector of population moments as a function of these parameters, $\boldsymbol{M}(\boldsymbol{\beta})$. This provides us with an identifying condition that must hold at the true value of this parameter vector, $\boldsymbol{\beta}_0$:$E_{\boldsymbol{\beta}_0}\left[\widehat{\boldsymbol{M}}\right] = \boldsymbol{M}(\boldsymbol{\beta}_0)$. The basic idea behind our minimum distance estimator is to choose parameters $\boldsymbol{\beta}$ to match the sample and population moments. More specifically, the minimum distance estimator can be defined as

$$\widehat{\boldsymbol{\beta}}_{MD} = \arg\min_{\boldsymbol{\beta}} \left(\boldsymbol{M}(\boldsymbol{\beta}) - \widehat{\boldsymbol{M}}\right)' \hat{V}_{\boldsymbol{\beta}}^{-1} \left(\boldsymbol{M}(\boldsymbol{\beta}) - \widehat{\boldsymbol{M}}\right) \qquad [8]$$

where $\hat{V}_{\boldsymbol{\beta}}$ is an estimate of the asymptotic covariance of $\widehat{\boldsymbol{M}}$. We can invoke the usual regularity conditions to show that $\widehat{\boldsymbol{\beta}}_{MD}$ is a consistent estimator of $\boldsymbol{\beta}_0$ from which we can construct

[7] For example, by the time those from more recent birth years reach their early twenties there is substantial variation in the proportion of sampled individuals with specific years of primary schooling between successive surveys. Naively applying equation [7] to explain this sampling variation would require large and sometimes nonsensical fluctuations in the promotion rates.

[8] We choose to express educational attainment relative to cohort size in order to reduce the effect of sampling variation in $N_{c,t}$.

consistent estimates of $\hat{\theta}_{c,s,t}$ and $\hat{\lambda}_{c,s,t}$. The most important requirement of the binding functions $\theta_{c,s,t}(\boldsymbol{\beta})$ and $\lambda_{c,s,t}(\boldsymbol{\beta})$ is that they should be flexible enough to retain the underlying features of the data. We specify these as:

$$\theta_{c,s,t} = \Phi\left(\sum_{j=0}^{6} \beta_{j,s,t}^{\theta} \sqrt{2}\cos(\pi j a) + \beta_{7,s}^{\theta} c\right) \quad [9a]$$

$$\lambda_{c,s,t} = \Phi\left(\sum_{j=0}^{1} \beta_{j,s,t}^{\lambda} \sqrt{2}\cos(\pi j a) + \beta_{2,s}^{\lambda} c\right) \quad [9b]$$

where $\Phi(.)$ is the cumulative density function for a standard normal distribution and $\sqrt{2}\cos(\pi j a)$ is an orthonormal basis function often used in non-parametric series estimation[9]. The time variation in the coefficients on these bases are restricted as

$$\beta_{j,s,t}^{\theta} = \beta_{j,s,0}^{\theta} + \beta_{j,s,1}^{\theta} \frac{\min\{t,2003\}-1997}{6} 1(t > 1997),$$

and similarly for $\beta_{j,s,t}^{\lambda}$.

Although flexible, this specification implies a number of important restrictions on the enrolment and promotion rates that are worth discussing. Although the parameters are allowed to vary across schooling stages[10], it is assumed that within schooling stages these rates are smooth functions of age, cohort and time. This property is important as it allows the estimator to use educational attainment and enrolment data for individuals from adjacent ages and surveys in order to provide estimates that are less sensitive to sampling variation and measurement error. The effect of age is represented by adding a series of cosine waves of different frequencies and amplitudes. We choose fewer bases for the promotion rate than for the enrolment rate, since the former was found to be more sensitive to the effects of sampling variation[11]. The $\beta_{j,s,t}$ coefficients are restricted to be constant until 1997 and after 2003, while changing linearly between these years. This change allows the data to reflect the gradual implementation of restrictions on over-aged learners and faster grade promotion. Our specification also includes a linear generational trend, so that long-term changes in enrolment are not opportunistically attributed to schooling policies. These functions are transformed using the normal c.d.f. so as to ensure that all enrolment and promotion rates are restricted to the unit interval.

[9]This type of non-parametric estimator has the benefit of being much quicker to estimate than kernel estimators, since it need not be evaluated separately at each age. This is a crucial advantage when combining a non-parametric specification with numerical optimsation techniques. The cosine function has the advantage of being smoother than a spline function and more stable at high values of a than a polynomial function. Age a is normalised to lie between 0 and 1 for ages from $s - 6$ to 30.

[10] After experimenting with different specifications that attempt to balance the objective of retaining the most important features in the data against smoothing away the effects of sampling variation, we defined eight schooling stages that are allowed to have different enrolment and promotion rates: grade 1, grades 2 and 3, grades 4 to 6, grades 7 to 9, grade 10, grade 11, matric and tertiary education. We also restrict the promotion rate coefficients to be the same for grade 1 to 3, and the enrolment rate coefficients to be the same for grades 2 to 6, and for grades 7 to 10.

[11] In a different specification we allowed the promotion rate to depend on four age-cosine bases instead of the two used in equation [9b], but this produced promotion rates that oscillated counter-intuitively with age while adding little to the explanatory power of the model. In order to avoid over-fitting the data we therefore opted for a more restrictive specification.

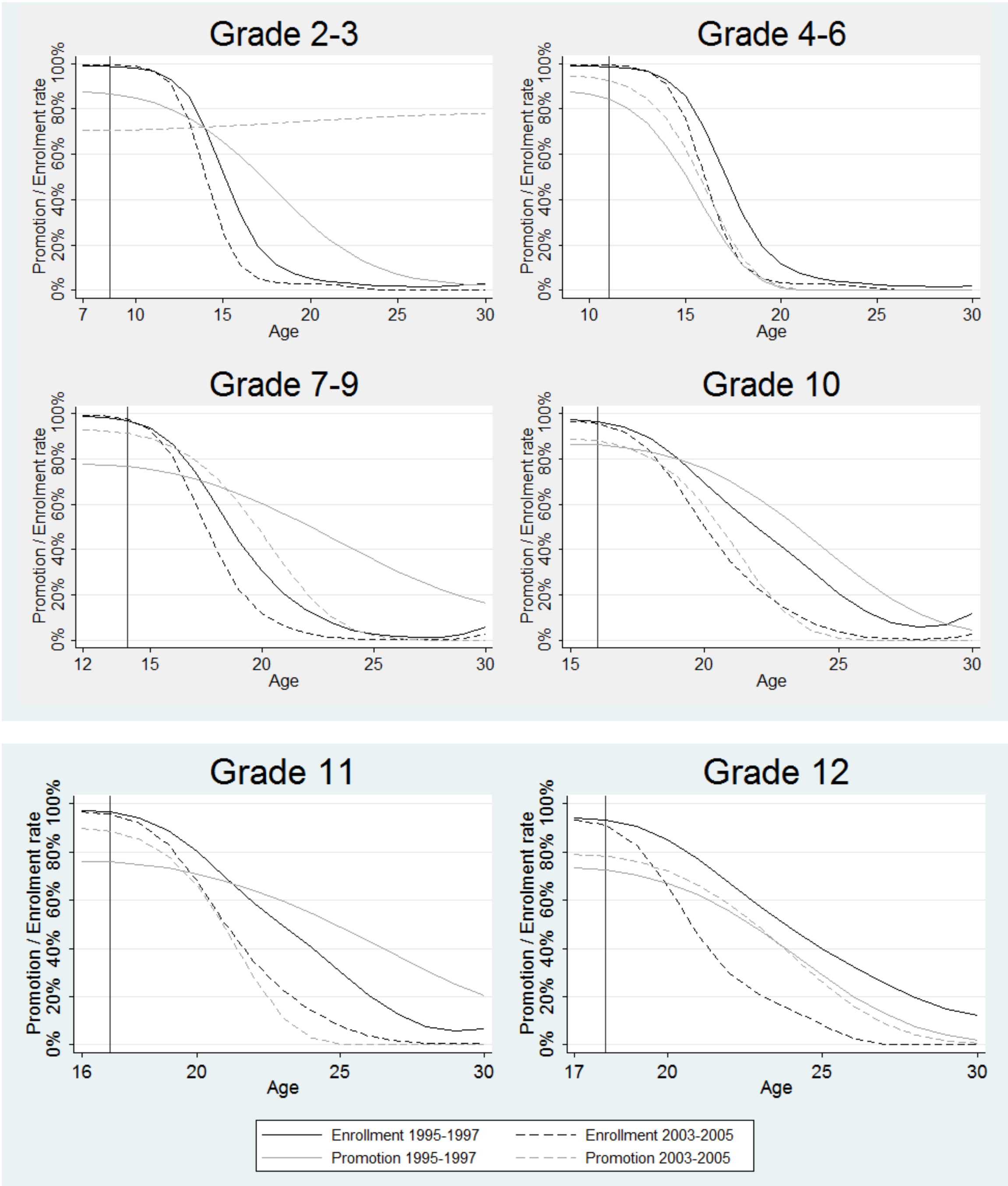
Given any value of $\boldsymbol{\beta}$, equations [9a] and [9b] can be used to construct enrolment and promotion rates for the different cohort-schooling year-period combinations. If we assume that at age 5 each cohort only consisted of individuals with no years of completed education, then these enrolment and promotion rates can be used along with equation [6] to produce iterative estimates of the evolving cohort education shares $\frac{N_{c,s,t}}{N_{c,t}}$. These predicted population moments are then compared to the corresponding sample moments to calculate the statistic[12] on the right-hand side of equation [8]. Our sample moments are restricted to individuals between the ages of 6 and 30, for whom there are fifteen potential schooling outcomes and thirteen survey periods. After dropping all cohort-schooling year-period combinations for which no individuals were sampled or for which the standard deviation of the sample mean was not strictly positive, we are left with 3,143 sample moments for educational attainment shares and enrolment rates. Numerical optimisation techniques are then used to minimise the weighted Euclidian distance between $\boldsymbol{M}(\boldsymbol{\beta}) - \widehat{\boldsymbol{M}}$. This was implemented using the iFit MATLAB toolbox (Farhi, 2011; Farhi, Debab, & Willendrup, 2013).

Figure 2 compares the predicted age-specific enrolment and promotion rates for different schooling stages obtained from the pre-implementation (1995-1997) and post-implementation (2003-2005) minimum distance coefficient estimates. The graphs compare the rates for a given cohort (those born in 1995), so that the observed differences are not affected by the gradual generational trend towards higher enrolment and faster promotion. We observe that promotion and enrolment rates are high and very high respectively at the correct grade ages (represented by vertical lines in Figure 2), but then decrease for older learners[13]. Furthermore, the effect of both the overage restrictions and faster grade promotion are clearly evident from these estimates. The enrolment of normal-aged students is very similar before and after the implementation of the policy, whereas a large decrease occurred for individuals who were a few years older than the correct grade age. Faster grade promotion appears to have been achieved by increasing the promotion rate for learners who are at or slightly above the normal grade age, whereas the promotion rates for those who were much older than the normal grade age actually decreased between 1997 and 2003.

[12] Given that the vector $\boldsymbol{\beta}$ contains 133 parameters, the curse of dimensionality makes it very costly to estimate the asymptotic covariance matrix $\widehat{V}_{\beta}$ using bootstrapping techniques. Instead, we approximate this matrix with a 133×133 matrix containing the variance of the sample moments on the main diagonal and zeros elsewhere.

[13] The one exception is the post-implementation promotion rates for the grade 2-3 schooling stage. Its counterintuitive age-pattern reflects the sensitivity of these estimates to sampling variation given the small sample of over-aged learners enrolled in grades 2-3 after the implementation of these policies, and is an example of why a more restrictive parameterisation is chosen for the promotion rate binding function [9b].

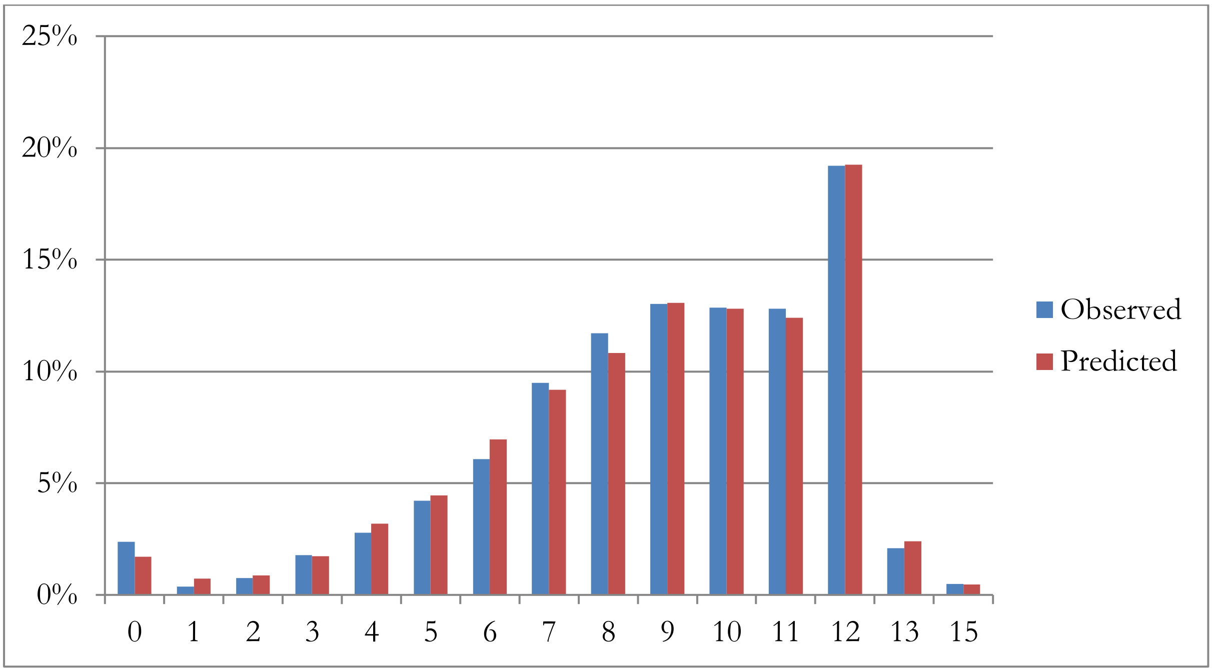
Figure 2: Promotion and enrolment rate estimates, by age, schooling phase and period



Source: Own calculations from OHS and LFS data (Statistics South Africa, various years)

Finally, Figure 3 compares the observed distribution of education attainment for black men (aged 15 to 30) to the distribution predicted by our model. We can see that our model does a reasonably good job of capturing the salient features of the whole distribution of schooling outcomes.

Figure 3: Observed and predicted distributions of educational attainment for black men aged 15-30



7. Estimating the schooling-earnings profile

We now proceed to estimate the earnings function while controlling for unobservable attributes that would otherwise confound the identification of the causal effect of schooling. We estimate the parameters of the Becker-Card model discussed in section 3, in order to infer the shape of the schooling-earning profile for the "average" wage earner. This 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 population average of the heterogeneous linear schooling parameter, $\bar{b}$, as well as the homogeneous schooling squared parameter, $\frac{1}{2}\kappa_1$, is analogous to the estimation of an ATE in a CRC model with exogenous conditioning variables. Section 4 discussed the conditions under which this can be achieved using the control function estimator.

Table 2 below reports the coefficient estimates from an OLS earnings regression with and without the first-stage schooling residual included as additional controls. Column 1 includes no residuals and therefore corresponds to the regular Mincerian OLS estimates. The estimates reveal the expected convex schooling-earnings profile that have been reported by previous studies for South Africa and elsewhere. The OLS return estimates can be seen to imply that schooling returns are negative for the first year of education, and then increase by 2.5% for each additional year of schooling. However, this estimator heroically assumes that the years of schooling completed is uncorrelated with the returns to schooling.

The specification in column 2 adds the schooling residual itself and interacted with schooling to obtain the CF estimates of the Becker-Card model. The schooling residual is calculated by subtracting the years of completed schooling predicted by the model in section 6 from actual completed schooling. This residual provides an estimate of the schooling error term, defined as variable u in section 3, which is potentially correlated with the heterogeneous returns to schooling, v, as well as to the wage error term, a. Under the assumptions stated in section 4, the correlation between these unobservable components can be controlled for by including the additional regressors in column 2.

On the other hand, the CF specification suggests a profile that is essentially linear and perhaps even marginally concave. The wage benefit of the first schooling year is 26%, but this benefit decreases by 1% 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 may have over-stated the degree of convexity in the schooling-earnings profile. The schooling-schooling residual interaction term is significantly positive 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. Apparently, those who invest in higher yielding schooling years also tend to stay in school for longer. Since individuals who leave school early tend to also have accumulated low returns schooling, this makes the returns to primary education appear very low. On the other hand, those who received high return primary and secondary schooling tend to only enter the labour market after completing many years of schooling, which artificially inflates the apparent return to advanced schooling years. This explains why OLS regression consistently find that the schooling-earnings profile is convex, whereas the CF estimates suggests that the average returns are not significantly non-linear. The schooling residual is significantly negative, which implies that the labour market productivity term is negatively correlated with the schooling error term.

Table 2: Log hourly wage regression estimates for black males aged 15-30: 1995-2005

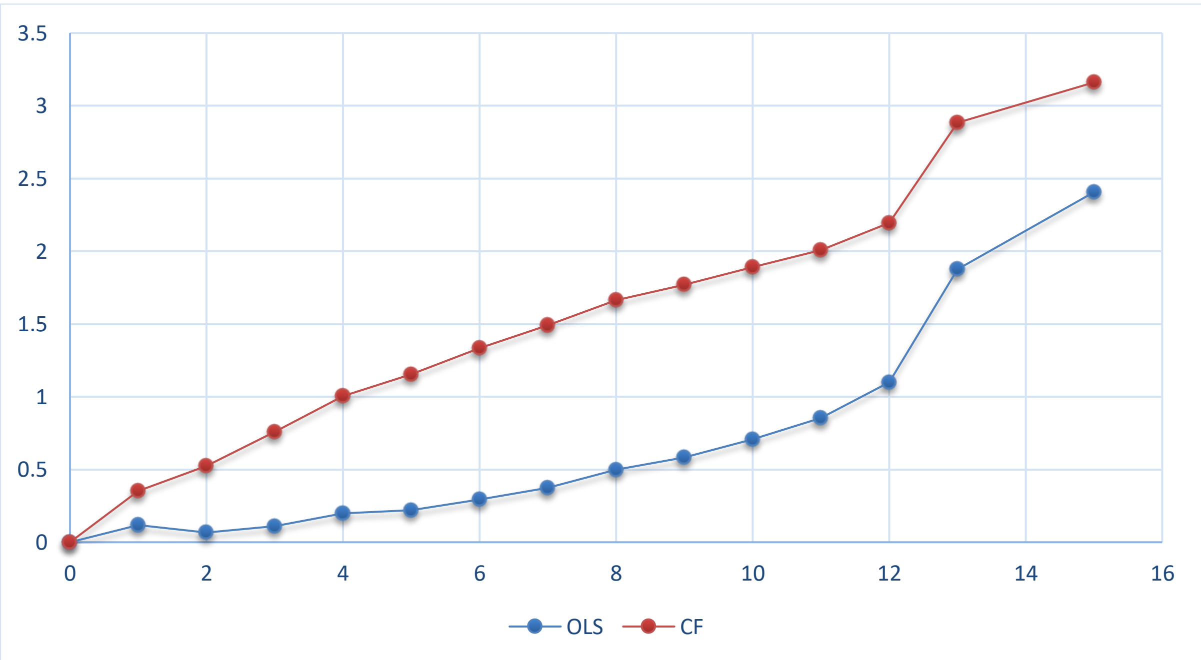
	(1)	(2)
VARIABLES	OLS	CF
Years of schooling	-0.0641***	0.257***
	(0.00736)	(0.0579)
Years of schooling^2	0.0126***	-0.00516
	(0.000416)	(0.00335)
Years of potential experience	0.0194***	0.0226***
	(0.00536)	(0.00606)
Years of potential experience^2	-0.000294	-0.000515**

	(0.000246)	(0.000254)
Birth year	-0.0432***	-0.0423***
	(0.00166)	(0.00212)
Schooling residual		-0.159***
		(0.0289)
Schooling residual* Years of schooling		0.0182***
		(0.00341)
Constant	86.03***	82.78***
	(3.305)	(4.229)
Observations	33,954	33,954
R-squared	0.226	0.227

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The more heroic assumptions required to estimate a quadratic earnings profile using the control function approach can also be exploited to estimate an even more flexible specification. In Figure 4 we plot the predicted earnings profiles estimated from OLS and CF wage regressions with an exhaustive set of schooling year dummy variables. This figure again demonstrates that the marginal returns to investing in schooling for black South African men are high on average: 15% estimates according to the OLS and 19% according to the CF estimates. These results also suggest that the endogeneity of schooling does not downwardly bias the OLS estimates of the marginal return to the average schooling year. Secondly, after controlling for schooling endogeneity and heterogeneity in the slope of the marginal returns, the earnings profile becomes very close to linear, except for a mild decrease of returns at higher schooling years and an additional 60% premium for completing any tertiary schooling years.

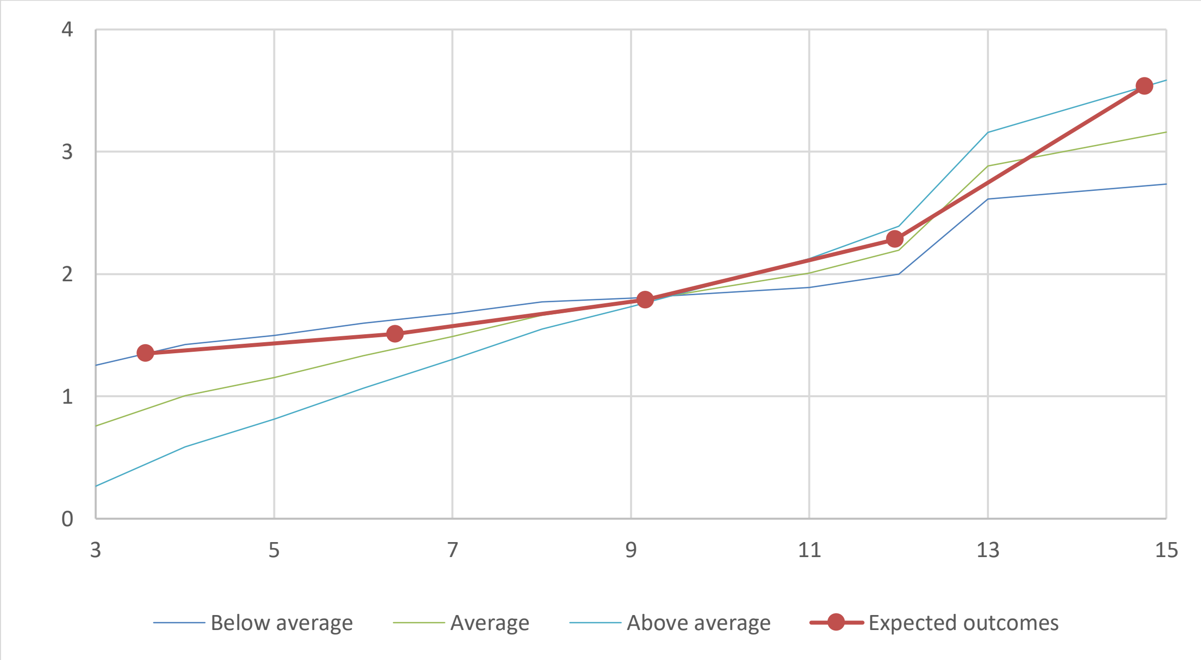
Figure 4: Estimated schooling-earnings profile (black males, aged 15-30, 1995-2005)



The predicted CF profile in Figure 4 represents the predicted wages evaluated at the expected value of the schooling residual (which is zero). Since the CF estimates suggest that heterogeneity in these profiles is an important explanation for the seemingly convex relationship between schooling and earnings, we now also construct these profiles for individuals with non-zero schooling residual values. Specifically, we use the distribution of schooling residuals to estimate the standard deviation of the schooling error term, and then calculate the expected profile faced by someone with a schooling error term two standard deviations below the mean, and for someone with two standard deviations above the mean. Since the schooling residual can be directly interpreted in terms of schooling years, we can also identify the expected years of schooling completed for individuals who faced each of these profiles. Figure 5 compares these three different earnings profiles.

We also add a line that connects the combinations of expected wages and expected schooling years for individuals with schooling errors that are one or two standard deviations above and below the mean. This line demonstrates the origins of the convex relationship between schooling and earnings, despite an earnings profile that is essentially linear. Individuals who face a steeper schooling-earnings profile – due to attending higher quality schools, coming from wealthier households or possessing more innate ability or motivation – will find it worthwhile to complete more years of schooling, whereas those who experience lower schooling returns can be expected to drop out of schooling earlier. The fact that those who complete fewer years of schooling also had low schooling returns makes it seem as if all primary education years are relatively low returns. By the time the high-returns individuals enter the labour market after accumulating many years of high-yielding schooling, they are observed to have many years of schooling and to earn very high wages. This creates the impression that more advanced schooling years are much more beneficial than primary schooling, and leads to researchers and policy makers to believe that greater access to advanced education years is all that is required to improve the prospects of labour market entrants. However, persuading learners that face low schooling returns to stay in school for longer will only push them further along the flat schooling profile, and hence produce disappointing labour market effects. A more appropriate response to the situation depicted in Figure 5 is to simultaneously improve access to schooling and to address the causes of heterogeneity in schooling returns, such as the low quality of education in some schools or the high levels of household income inequality.

Figure 5: CF estimates of heterogeneous schooling-earnings profiles (by individual schooling error)



8. Conclusions

This paper used two shifts in school policy to estimate the causal effect of schooling on the earnings of black South African men. Our model of earnings allows for endogenous schooling, individual heterogeneity in the returns to schooling and non-linearities in the schooling-earnings profile. This model is more general than has been estimated for South Africa (and any other country, to our knowledge) before. The results of our control function estimates suggest that the South African schooling earnings profile is actually very close to linear and perhaps even concave. We also find evidence of substantial heterogeneity in the slope of this curve, which is consistent with other studies that have found very high levels of inequalities in household background characteristics and school quality. The results suggest that individuals with low returns end up with fewer schooling years, while high return individuals choose more years of schooling. This interaction produces what seems like a convex schooling-earnings profile when using estimators that do not allow for heterogeneous returns or endogenous schooling outcomes.

9. Bibliography

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., 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.

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.

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.

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.

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

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

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.

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.

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.

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.

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.

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

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

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.

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., & 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.

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.

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.

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.

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

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. (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.

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.

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.

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

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

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.

Spence, M. (1973, Augustus). 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.

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

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

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#

The **Research Project on Employment, Income Distribution and Inclusive Growth (REDI3x3)** is a multi-year collaborative national research initiative. The project seeks to address South Africa's unemployment, inequality and poverty challenges.

It is aimed at deepening understanding of the dynamics of employment, incomes and economic growth trends, in particular by focusing on the interconnections between these three areas.

The project is designed to promote dialogue across disciplines and paradigms and to forge a stronger engagement between research and policy making. By generating an independent, rich and nuanced knowledge base and expert network, it intends to contribute to integrated and consistent policies and development strategies that will address these three critical problem areas effectively.

Collaboration with researchers at universities and research entities and fostering engagement between researchers and policymakers are key objectives of the initiative.

The project is based at SALDRU at the University of Cape Town and supported by the National Treasury.

Tel: (021) 650-5715

www.REDI3x3.org

REDI3X3

Research Project on Employment, Income Distribution & Inclusive Growth

Towards an integrated response to unemployment, inequality and poverty