South Africa is not a Latin American country: Prospects for wage redistribution

Debra Shepherd and Servaas van der Berg

The "Paradox of Progress"

Following Battistón et al (2014), the analysis of this paper aims to illustrate how educational expansion may be linked to rising inequality using a simple model that relates log earnings to education at period t as follows:

$$lnY_{it} = \alpha_t + \beta_t X_{it} + \varepsilon_{it} \quad [1]$$

where X_{it} is individual specific years of education and the zero-meaned error term, ε_{it}, summarises all unobservable determinants. Assuming independence between X_{it} and ε_{it}, β can be interpreted as a measure of the returns to education. Suppose further that all earners can be divided into two groups, H and L, with $X_H > X_L$ and $E(ln\,Y_H) > E(ln\,Y_L)$. The expected log earnings gap G can therefore be expressed as:

$$G = E(lnY_{Ht} - lnY_{Lt}) = \beta_t(X_{Ht} - X_{Lt}) \quad [2]$$

and the change in the earnings gap between t = 1 and t = 2 as:

$$\Delta G = E(lnY_{H2} - lnY_{L2}) - E(lnY_{H1} - lnY_{L1}) = (\beta_2 - \beta_1)(X_{H1} - X_{L1}) + \beta_2(dX_H - dX_L) \quad [3]$$

where dX_i is the change in education for individuals in $i = H, L$. From the above equation we can see that a change in inequality depends on (i) changes in the returns to education over time, (ii) the initial difference in education levels between the two groups, and (iii) the relative change in education. Therefore, if returns to education are constant over time and the growth in educational levels is similar across groups, $\Delta G = 0$. Adding convexity in the returns to education to the model:

$$lnY_{it} = \alpha_t + \beta_t X_{it} + \eta_t X_{it}^2 + \varepsilon_{it} \quad [4]$$

results in the expected change in the earnings gap over time becoming:

$$\Delta G = (\beta_2 - \beta_1)(X_{H1} - X_{L1}) + \beta_2(dX_H - dX_L) + (\eta_2 - \eta_1)(X_{H1}^2 - X_{L1}^2)$$

$$+\eta_2(dX_H^2 - dX_L^2) + 2\eta_2(X_{H1}dX_H - X_{L1}dX_L) \quad [5]$$

In this case, if $\beta_2 = \beta_1$, $\eta_2 = \eta_1$ and $dX_H = dX_L$, $\Delta G = 2\eta_2(X_{H1} - X_{L1})dX > 0$ under convex returns to education i.e. $\eta_2 = \eta_1 > 0$. Under convexity, Battistón et al (2014) show that even an expansion of education in favour of the less educated earners can lead to a rise in inequality. The following section discusses the determinants of earnings inequality, with emphasis paid to the evolution of the convexity in the returns to education in South Africa over the two decades since democratisation.

DETERMINANTS OF EARNINGS INEQUALITY

Earnings inequality is affected by two primary factors: first, the distribution of (observable and unobservable) characteristics of workers such as years of education, experience, gender and ability; and secondly the returns to those characteristics. Workers' characteristics, in turn, are affected by decisions such as whether or not to enrol in school (and at which stage to leave formal schooling) as well as policy such as expanding access to education. Returns to labour market characteristics depend on market forces; that is, the demand and supply of workers of different skills (education) and institutional/policy factors such as minimum wages. Both worker characteristics and returns to these characteristics have changed during the two decades after 1994.

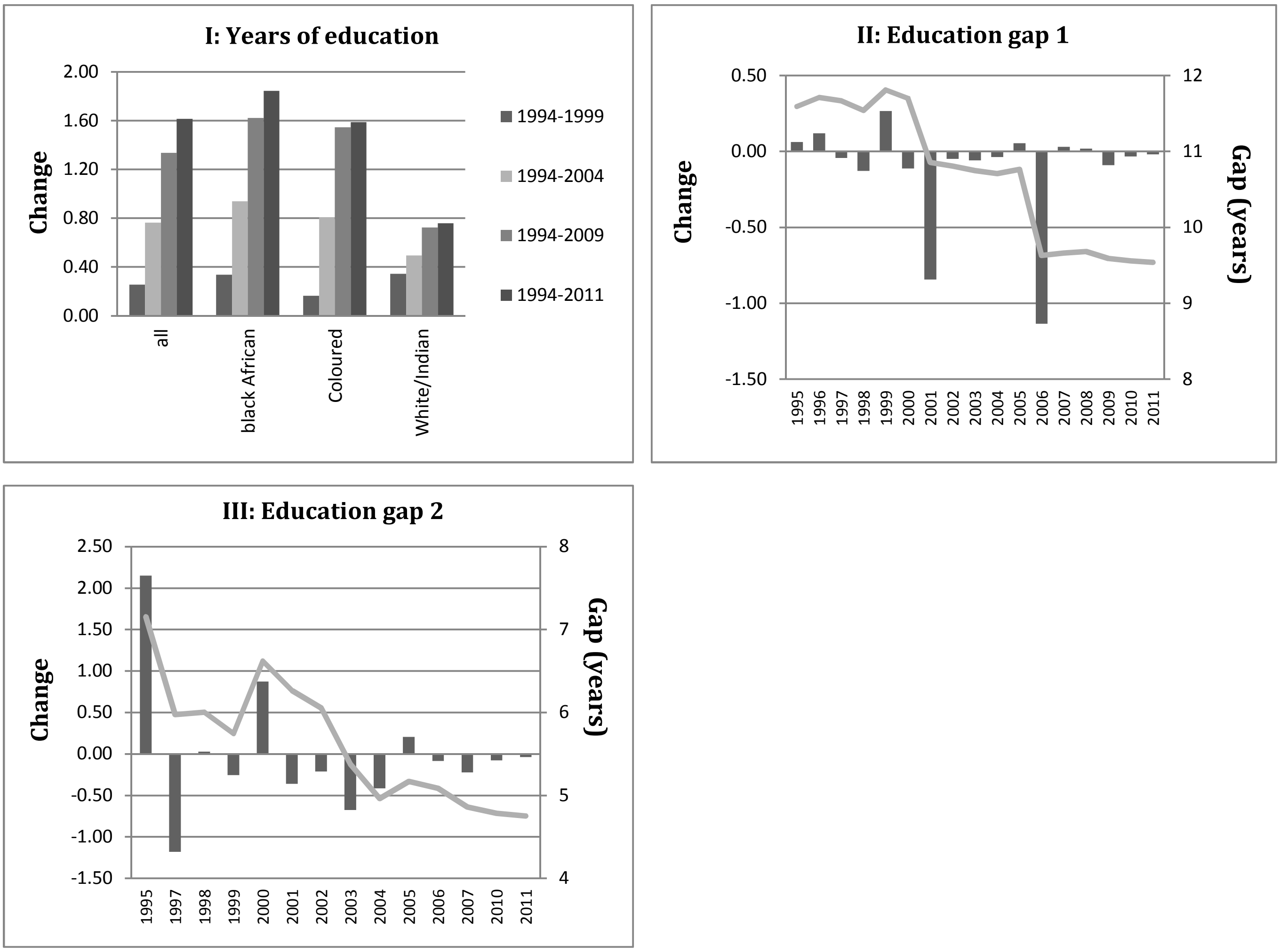
Changes in the distribution of education 1994-2011

South Africa has experienced a notable expansion in educational attainment since democratisation, with average years of education amongst the working age population increasing from 8.2 in 1994 to 9.6 to 2011. This educational expansion was not homogeneous across demographic groups, however, with most of the expansion being driven by the black African and Coloured population groups (see Figure 1 panel I).

Panels II and III of Figure 1 present the education gap between two groups of individuals over time: gap 1 presents an absolute measure of education inequality through the difference in average years of education between the top 20% and bottom 20% of the education distribution; and gap 2 presents inequality in education relative to earnings through comparing the average years of education between the top 20% and bottom 20% of the earnings distribution. Both education gaps reveal a decline in absolute education inequality over time, most notably since 2000. The education gap between extreme quintiles of the education distribution has dropped consistently over the 17 year period under consideration with the gap currently hovering around 9.5 years of education. Similarly, the education gap between extreme earnings quintiles declined by approximately 2 years over the decade 2001-2011.

If we make comparisons to similarly calculated gaps in thirteen Latin American (LA) countries by Battistón et al (2014) over a similar time period (1990-2009), the decline in educational inequality in South Africa has been significantly larger. Whilst the average expansion in the number of years of education was similar across the LA countries (1.5 years), only Chile realised a similarly sized decline in the education gap 1 measure of roughly 1.5 years. In fact, seven of the thirteen countries saw an *increase* in the education gap between the highest and lowest earners, although this was largely driven by changes in educational inequality during the 1990s. Both education gaps dropped during the period 2002 to 2009, suggesting that whilst the education growth path was biased towards the most educated and higher earners before 2002, this trend subsequently reversed (Battistón et al, 2014: 13). Therefore, as with South Africa, the Latin American region has seen a decline in both relative and absolute inequality in education since the early 2000s.

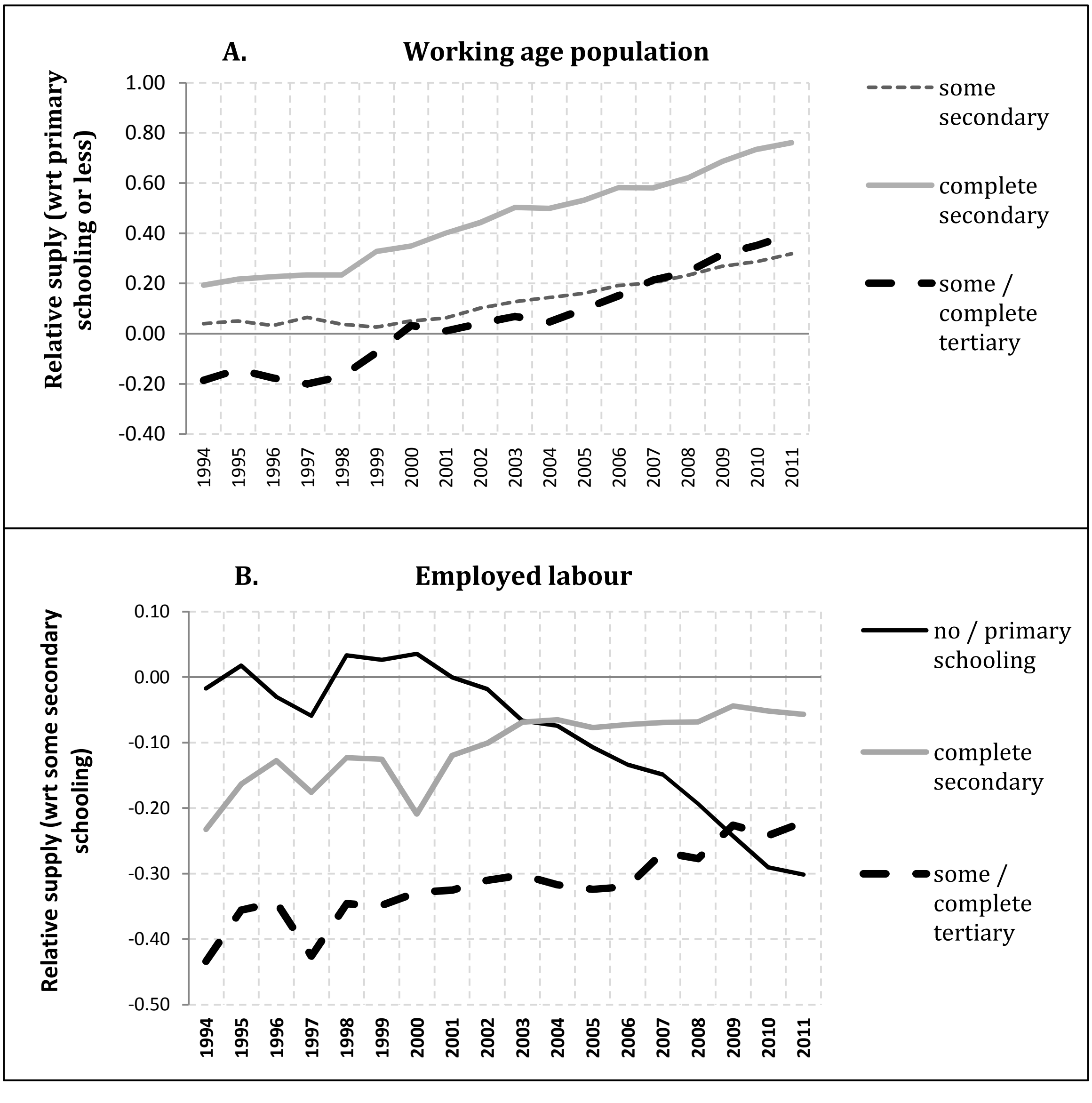
Figure 1: changes in years of education and educational inequality, 1994-2011



Source: own calculations using PALMS 1994-2012

Panel A of Figure 2 shows that the proportions of the working age population with some/complete secondary schooling and some/complete tertiary degrees (primary and no education) have steadily risen (declined) since 1994; the relative supply of tertiary educated individuals has risen faster since 2004. Panel B indicates the relative supplies of educated labour amongst the employed. The trends mirror that of the working age population, that is, a more educated workforce over time. Unlike Panel A, the relative labour supplies reflected in Panel B are measured with respect to the proportion of employed individuals with some (but not completed) secondary education. The South African workforce is predominantly comprised of individuals with this level of education, about one-third of all employed individuals, and this proportion was fairly unchanged over the time period. Panel B therefore illustrates that the employed are becoming consistently more skilled over time. Whereas roughly similar proportions (one-third) of individuals with no/primary schooling and incomplete secondary education were employed at the commencement of democratisation, the proportion of employed workers with no/primary schooling dropped by roughly 50 percent between 2000 and 2011. Considering the formally employed only, the changes are even more pronounced (not shown here), with an increase in the proportion of tertiary educated (some/complete) and complete secondary former sector workers of 100 percent and 35 percent over the period, respectively.

Figure 2: relative supply of educated labour, 1994-2011

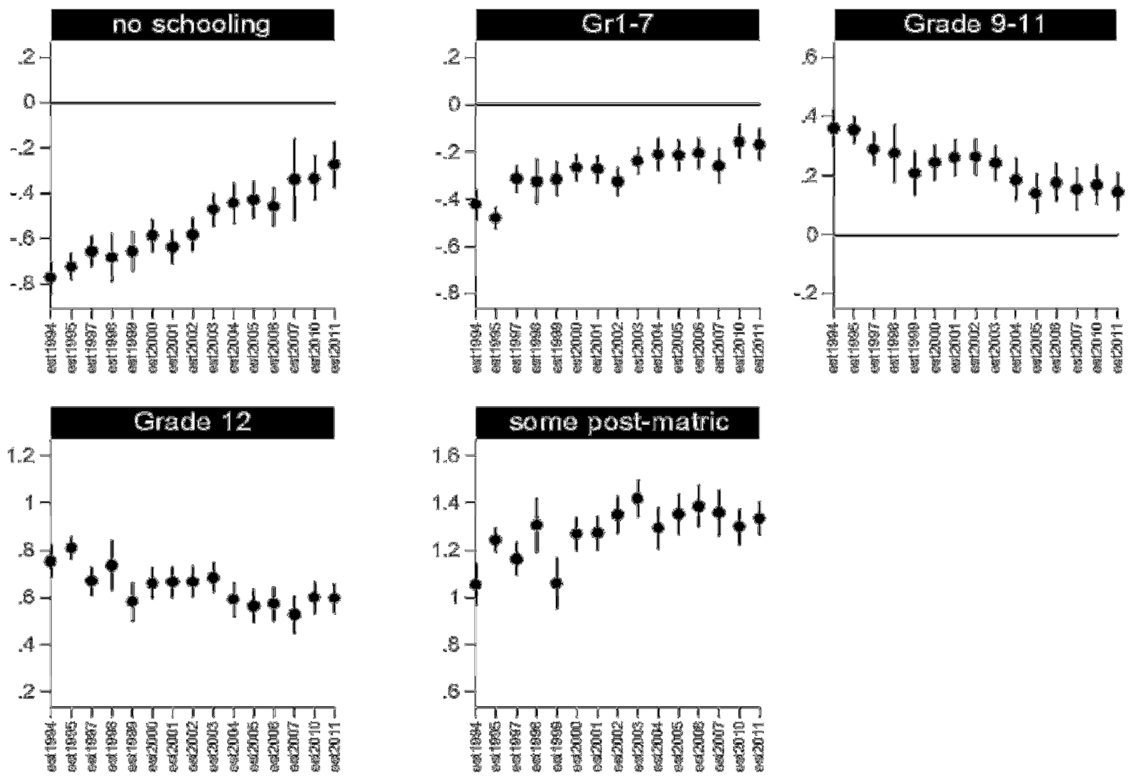


Source: own calculations using PALMS 1994-2012 (datafirst)

Note: the relative supplies of labour amongst the working age population are calculated as the log of the ratio of the proportion of individuals with a given level of education (some secondary, complete secondary or some/complete tertiary) to the proportion of individuals with complete primary education or less. The relative supplies of employed labour are calculated similarly, but relative to employed individuals with some secondary schooling.

Changes in the returns to education 1994-2011

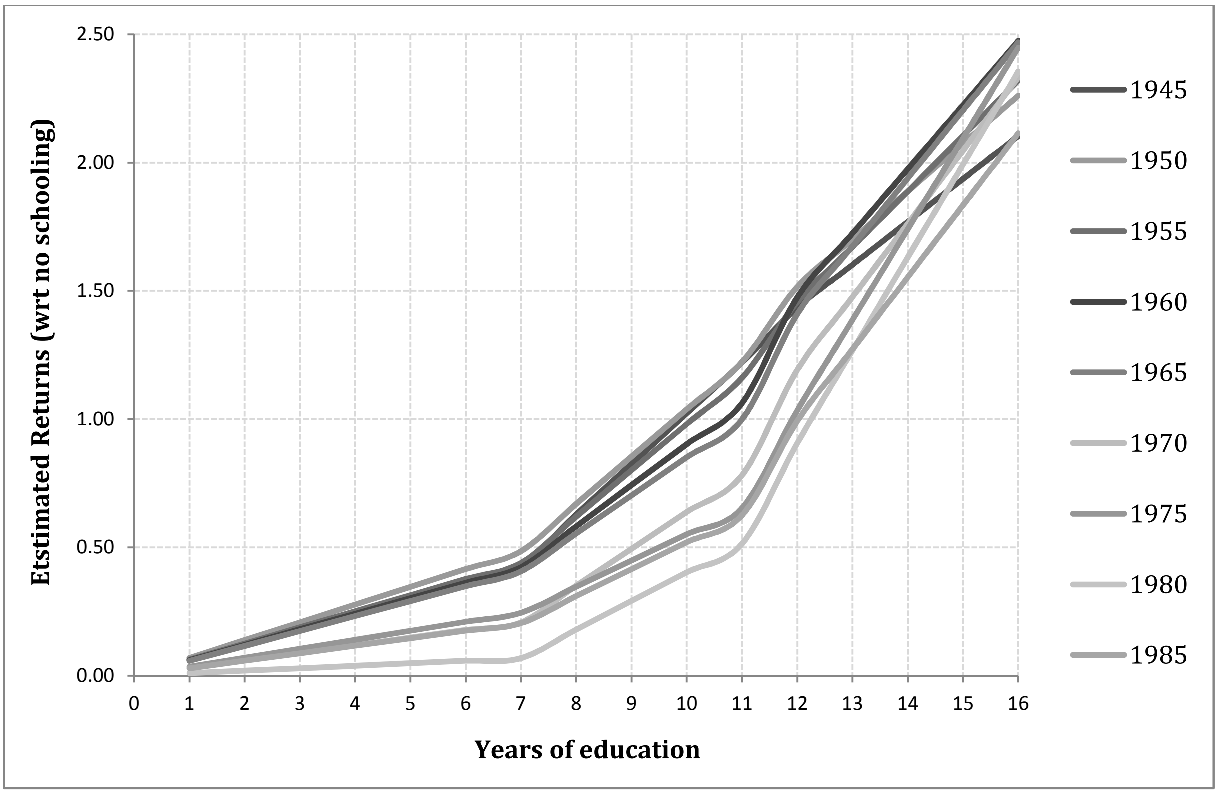
Figure 3: Returns to education (levels), 1994-2011



Source: own calculations using PALMS 1994-2012

Note: Returns to education are normalised to 0 for individuals with highest completed level of education equal to Grade 8. Aside from school level dummies, Mincerian wage regressions include controls for gender, race and age (quadratic).

Figure 4: Returns to education by birth cohort



Source: own calculations using PALMS 1994-2012

Note: Returns to education are normalised to 0 for individuals with no years of formal schooling. The wage regression is estimated on a pooled dataset (including years 1994-2011) with log(wage) as the dependent variable; time effects, age (quadratic) and birth cohort effects are included as controls. Returns are calculated as the coefficients on interactions between birth cohort effects and education splines.

THEORETICAL FRAMEWORK

In this section we present a model of labour demand disaggregated by level of education and age that allows us to estimate the impact of changes in the relative supplies of educated labour on the wage structure. The model follows the approaches of Katz and Murphy (1992), Card and Lemieux (2001) and Manacorda, Manning and Wadsworth (2012).

We assume the firms produce according to a typical neoclassical production function that combines labour and capital, and that capital is exogenous to the firms' decision of how much labour to employ. The aggregate labour input is modelled as a CES composite of S imperfect substitute skill (education) groups:

$$L_t = \left[\sum_{s=1}^{S} \theta_{et} L_{et}^{\rho}\right]^{\frac{1}{\rho}} \qquad [6]$$

The relative efficiency (represented by θ_{et}) of the different education groups are normalised at $\theta_{1t} = 1$. The elasticity of substitution between the different education groups is measured by

$\sigma_E = \frac{1}{1-\rho}$, which for simplicity sake we assume is time invariant and is constant over education groups; that is, $\sigma_E = \sigma$. The labour input for a given education group e is composed of different birth (age) groups as follows:

$$L_{et} = \sum_{a=1}^{A} \lambda_{ea} L_{eat} \qquad [7]$$

where the efficiency parameter λ_{ea} is assumed to be time invariant and normalised at $\lambda_{e1} = 1$. As with Battistón et al (22014), we assume perfect substitutability across age groups for a given level of education. In this setting, the marginal product of labour for a given age-education group will therefore depend on the aggregate supply of labour within its education category (Card and Lemieux, 2001: 710):

$$\frac{\delta y_t}{\delta L_{eat}} = \frac{\delta y_t}{\delta L_{et}} \times \frac{\delta L_{et}}{\delta L_{eat}} = \theta_{et} L_{et}^{\rho-1} (\theta_{et} L_{et}^{\rho})^{\frac{1}{\rho}-1} \times \lambda_{ea} \qquad [8]$$

Assuming competitive markets and normalising output prices to one, log wages are given by:

$$\log w_{eat} = \log\alpha + \log Y_t + \log\theta_{et} + \log\lambda_{ea} - \rho \log L_t + (\rho - 1)\log L_{et} \qquad [9]$$

However, equation (12) is not directly estimable. We can consider the following specification:

$$\log w_{eat} = constant + f_t + (f_t * f_e) + (f_e * f_a) + \epsilon_{eat} \qquad [10]$$

where the time fixed effects (f_t) absorb $\log Y_t - -\rho \log L_t$, the time-education interactions ($f_t * f_e$) absorb $\log\theta_{et} + (\rho - 1)\log L_{et}$. The education-age interactions ($f_e * f_a$) allow for identification of $\log\lambda_{ea}$, which in turn allows for the estimation of L_{et}.

Assuming that relative wages are equated to relative marginal products, the wage ratio relative to the lowest education group satisfies:

$$\log\left(\frac{w_{eat}}{w_{1at}}\right) = r_{eat} = \log\frac{\theta_{et}}{\theta_{1t}} + \log\frac{\lambda_{ea}}{\lambda_{1a}} + (\rho - 1)\log\left(\frac{L_{et}}{L_{1t}}\right) + \xi_{eat} \qquad [11]$$

where ξ_{eat} represents sampling or other sources of variation in r_{eat}. According to this model, education wage gaps for a given age group depends on the aggregate relative supply of divergently educated workers in period *t*, assuming age groups with the same level of education are perfect substitutes. If we were to relax this assumption of perfect substitutability – that is, assume the partial elasticity of substitution between age groups is σ_A - then the education wage gap would also depend on the age-specific supply of higher educated labour relative to uneducated labour; that is:

$$r_{eat} = \log\frac{\theta_{et}}{\theta_{1t}} + \log\frac{\lambda_{ea}}{\lambda_{1a}} + (\rho - 1)\log\left(\frac{L_{et}}{L_{1t}}\right) - \left(\frac{1}{\sigma_A}\right)\left[\log\left(\frac{L_{eat}}{L_{1at}}\right) - \log\left(\frac{L_{et}}{L_{1t}}\right)\right] + \xi_{eat} \qquad [12]$$

However, Card and Lemieux (2001) show that even if $1/\sigma_A > 0$, the age structure of the education wage gap can be constant over time if $\log\left(\frac{L_{eat}}{L_{1at}}\right) - \log\left(\frac{L_{et}}{L_{1t}}\right)$ is approximately constant over time, which will occur if the relative supplies of educated labour in each age group grow at a constant rate.

With estimates of L_{et} after implementation of equation [10], [11] can be estimated directly using:

$$r_{eat} = f_e + f_t + (f_e \times f_a) + (\rho - 1)\log\left(\frac{L_{et}}{L_{1t}}\right) + \xi_{eat} \qquad [13]$$

where $\log\frac{\theta_{et}}{\theta_{1t}}$ is allowed to vary additively in *e* and *t*. The returns to education are affected by changes in the education composition of the population in relation to the elasticities of substitution rooted in labour demand. The elasticity of $\log\frac{w_{eat}}{w_{1at}}$ with respect to the relative labour supply is given by:

$$\kappa_{eat} = \frac{\rho - 1}{\left(\log\frac{w_{eat}}{w_{1at}}\right)} \qquad [14]$$

The microsimulation methodology of this paper follows Gasparini et al (2005) and Bourgiugnon et al (2005). Specifically, the counterfactual earnings distribution that would arise in period *t* if education was distributed as in period *t** is estimated, holding all other earnings determinants at their values in period *t*; that is, the counterfactual log earnings is defined as:

$$lnY_{it}(X_{it}^{*}) = F(X_{it}^{*}, Z_{it}, \varepsilon_{it}, \beta_t, \gamma_t) \qquad [15]$$

where X_{it} is a vector of education-specific characteristics, Z_{it} is a vector of non-education labour market characteristics, ε_{it} is a vector of unobservable characteristics, and β_t and γ_t are the model parameters to be estimated. Using this representation, the difference between the observed earnings distribution and the counterfactual distribution through a measure of inequality such as the Gini coefficient provides an indication of the partial equilibrium, first-round impact of a change in the distribution of education.

In order to calculate [13], we need estimates of β_t, γ_t and ε_{it}. These are obtained from a standard Mincerian earnings function (Mincer, 1974) where log earnings are modelled as a linear function of observable labour market characteristics:

$$lnY_{it} = \alpha_t + X_{it}\beta_t + Z_{it}\alpha_t + \varepsilon_{it} \qquad [16]$$

X_{it} can be modelled either by the number of years of education and its square or a set of dummies for the highest educational level completed. Z_{it} may include characteristics such as age (and age squared), race and gender dummies, and dummies of area type and province. There are well documented issues related to the identification of β_t, in particular omitted variable bias that, according to Card (1999), typically lead to under-estimation of the returns to education. Selection biases (into paid employment) may be corrected for through a Heckman two-stage procedure. Correction for endogeneity bias usually requires access to panel data or suitable instruments for education (Angrist and Krueger, 1991), both of which are not available in our data. Therefore, we acknowledge that the simulations produced by this paper most likely offer lower bound estimates for the simulated change in earnings.

The distribution of education of year t* is replicated using the procedure of Legovini et al (2005). The adult population of year t are divided into homogenous birth-race groups. The following transformation is performed for each individual *i* within cell *j*:

$$X_{it}^{*} = \left(X_{ijt} - \mu_{jt}\right)\left(\frac{\sigma_{jt}^{*}}{\sigma_{jt}}\right) + \mu_{jt}^{*} \qquad [17]$$

where μ_{jt} and σ_{jt} are the sample mean and standard deviation within cell j in year t, and similarly for μ_{jt}^{*} and σ_{jt}^{*}. This adjustment leads to the distribution of education in each cell in year t having the same mean and variance of the corresponding cell in year t*. Again we emphasise that the results provide a partial equilibrium direct effect on the distribution of earnings through a change in the distribution of education, as it is highly implausible that a change in the educational levels are likely to keep the remaining determinants of earnings unchanged.

In order to identify the effect of a change in education levels on the returns to education, we focus on educational levels where $\hat{\beta}_{e-1,t} = \log\frac{w_{eat}}{w_{a1t}}$ are the estimated returns to having education level e relative to having the lowest level of education. The percentage change in the Mincerian returns to education in a response to counterfactual change in educational levels is estimated using equation [11], where $\frac{L_{et}}{L_{1t}}$ and its change are estimated using the structural parameters. Given that κ_{eat} is time-variant, the baseline elasticity at time t is used in order to simulate changes during time t*; that is, $\kappa_{eat=1994}$ is used to simulate the changes in the returns to education when the educational structure of, for example, t* = 2010 is replicated in t = 1994.

Preliminary Findings

Earnings inequality and education expansion

Table 1 reports the actual change in earnings inequality as well as the simulated change by altering the educational structure using equation [17]. As the results are path dependent, we report two alternative simulations: (1) the change in the Gini coefficient if the education structure of the earlier year (in this case 2000) is simulated on the population in the final year (2011); (2) the change in the Gini coefficient if the education structure of the later year is simulated on the population in the first year. According to table 1, the education expansion over the period 1994 to 2011 had direct, first-round unequalising effects on the earnings distribution. The time period has been split into two in order to match the change in educational expansion as seen from the education gaps of Figure 1. We see that although educational expansion was directly related to an increase in earnings inequality over both time periods, the impact was stronger over the earlier time period. The difference in the magnitude of the unequalising effects of educational expansion of the late 1990s and 2000s may be related to a bias in educational improvements towards the more educated (and wealthier) groups of individuals in the earlier period.

Figure 5: Effect of change in educational distribution on earnings inequality (Gini)

	Observed Gini			Education effect (ΔGini)		
Period	t_1	t_2	Change	Simulation 1	Simulation 2	Average
1994-2000	0.565	0.711	0.154	0.179	-0.004	0.088
2000-2011	0.711	0.749	0.038	0.041	0.108	0.067

Source: own calculations using PALMS 1994-2012

Note: simulation 1 simulates the education distribution of t_1 on the population in t_2, and vice versa for simulation 2.