Why has decreasing schooling inequality not led to decreasing earnings inequality in South Africa?

David Lam
University of Michigan
davidl@umich.edu

Murray Leibbrandt
University of Cape Town
Murray.Leibbrandt@uct.ac.za

Arden Finn
The World Bank
afinn1@worldbank.org

Nicola Branson
University of Cape Town
nicola.branson@uct.ac.za

Date of draft: 1 July 2025

David Lam is Professor of Economics Emeritus at the University of Michigan. Murray Leibbrandt is UCT Research Chair for Poverty and Inequality Research in the Southern Africa Labour and Development Unit (SALDRU) at the University of Cape Town. Arden Finn is Senior Economist at the World Bank. Nicola Branson is a Chief Research Officer at SALDRU.

Why has decreasing schooling inequality not led to decreasing earnings inequality in South Africa?

Abstract

Inequality in education has declined substantially in South Africa since the end of apartheid, with inequality in years of completed schooling declining by all standard measures of inequality. At the same time, inequality in earnings has not shown significant declines, and has increased by some measures. Given the strong positive relationship between earnings and years of schooling, why hasn't the decline in schooling inequality led to declines in earnings inequality? This paper explores this puzzle from both a theoretical and empirical perspective. We analyse how earnings inequality is affected by changes in returns to schooling when returns increase at some levels of schooling and decrease at other levels. We show that changes in the distribution of schooling over the 1994-2019 period would have significantly reduced earnings inequality in and of themselves. This was offset by disequalizing changes in the earnings-education gradient, including an increase in relative earnings for those with post-secondary education and a decrease in relative earnings for those with incomplete secondary education. The net result is a combination of decreasing schooling inequality and persistently high earnings inequality.

Introduction

Since the end of apartheid in 1994, South Africa has made progress in expanding education and decreasing schooling inequality (Branson and Lam (2021). At the same time, inequality in earnings, which has long been among the highest in the world, has remained persistently high, with some measures showing an increase since 1994 (Wittenberg 2017, Kerr 2025). Given that education is one of the strongest determinants of earnings, why hasn't the decline in schooling inequality led to a decline in earnings inequality?

The paradox of falling schooling inequality accompanied by rising earnings inequality is not unique to South Africa. Knight and Sabot (1983) showed theoretically and empirically that improvements in the education distribution can first increase and subsequently decrease earnings inequality, looking at data from East Africa. Almeida dos Reis and Paes de Barros (1991) and Lam and Levison (1991) documented a similar pattern in Brazil. Bourguignon, Ferreira, and Lustig (2005) found that educational expansion was associated with increasing income inequality in many of the countries included in their analysis of inequality in Latin America and Asia. Lam (2022) showed that a combination of falling schooling inequality and rising earnings inequality may be quite common as countries go through the process of economic development and educational expansion.

The goal of this paper is to further explore the issue both theoretically and empirically, focusing on South Africa's post-apartheid experience. We look at how earnings inequality is affected by changes in the distribution of education changes in the earnings-education gradient. We look at these issues in the case in which returns to schooling differ across the education distribution, and when the shape of the earnings-education gradient changes over time. A common feature of labour markets in developing countries in recent decades has been for returns to schooling to change at different rates (and even in different directions) at different levels of schooling. Returns to university may have increased, for example, at the same time that returns to primary schooling declined. In this context it can be misleading to generalize about whether the change in average returns to schooling has been equalizing or disequalizing. We develop a general framework for analysing these issues, and derive some simple analytical results about the impact of changes in relative earnings at different levels of education on earnings inequality.

Our theoretical results show that there will generally be a benchmark level of education that separates equalizing from disequalizing increases in relative earnings. As makes sense intuitively, increases in earnings at low grades will tend to be inequality reducing, while increases in earnings at high grades are inequality increasing. We show that the cutoff level of education separating equalizing from disequalizing effects of rising earnings will be different for different inequality measures, will change over time, and will typically be different for different race and gender groups.

Our results also help us understand the impact on earnings inequality of changes in the distribution of education. The basic result is that shifts in the distribution of education will be equalizing if people tend to move toward the mean level of education or toward some other cutoff level of education for a given measure of inequality. Moving people from grade 4 to grade 8, for example, will tend to be equalizing, while moving people from grade 12 to university will tend to be disequalizing. We show that whether a given shift in the education distribution is equalizing or disequalizing will change over time and will be different for different race and gender groups.

Our theoretical results provide a useful framework for empirical analysis. They call attention to an interesting summary statistic that has not previously been studied – the year of schooling which separates equalizing from disequalizing increases in grade-specific earnings. In the case of the variance of log earnings, for example, we show that this is the level of schooling at which mean log earnings is earned. Increases in earnings above this level will be disequalizing, while increases in earnings below this level will be equalizing. For the Generalized Entropy (0) measure, the cutoff is the level of schooling corresponding to mean earnings. These benchmark levels of schooling also provide benchmarks for understanding how changes in the distribution of schooling affect earnings inequality. Changes in the schooling distribution that shift the distribution toward the benchmark level of schooling will be equalizing, while shifts away from that schooling level (in either direction) will be disequalizing. We use this framework to guide empirical analysis of schooling inequality, returns to schooling, and earnings inequality in post-apartheid South Africa

In particular, we look at what happens when earnings relative to the mean increase at the post-secondary level, but decrease at the incomplete secondary level, the pattern that is observed in South Africa in the post-apartheid period. We show that increases in the relative earnings of those with grade 11 education would have been disequalizing in the early post-apartheid period, but became equalizing in later years. Since the relative earnings of those with grade 11 education declined over time, this contributed to rising earnings inequality by the end of the period. We show in our counterfactuals that the decline in the relative earnings of those with 8-11 years of education had a disequalizing effect on earnings inequality that was roughly similar to the inequality-increasing effect of the increase in returns to university education. Looking at the effect of changes in the distribution of education, our counterfactuals show that improvements in the distribution of education would have been equalizing if the earnings-education gradient had remained constant. The equalizing effects of the decline in schooling inequality were offset, however, by the disequalizing effects of the changes in the earnings-education gradient.

Decomposing earnings inequality and the use of counterfactual simulations

The link between education and the distribution of income has long been fundamental to research on inequality. Theoretical models and a vast body of empirical evidence point to a large explanatory role for education in the distribution of income, especially the distribution of labour earnings. Standard human capital models imply that both the distribution of education and the returns to education will affect earnings inequality. Decomposition of these two components, often referred to as the "quantity" (or "endowment") and "price" components, have played an important role in understanding changes in earnings inequality in both high-income countries and developing countries (for example Juhn, Murphy, Pierce 1993, for the United States, and Bourguignon et al., 2005, for Latin America and Asia).

Much of the research in this area builds on the decompositions developed by Blinder (1973) and Oaxaca (1973). Building on these decomposition approaches, research on income inequality has often used microsimulation approaches to construct counterfactual distributions based on hypothetical changes in the underlying distributions of characteristics ("endowment effects") or changes in the impact of those characteristics on income ("price effects"). Studies such as Bourguignon et al. (2005) and Bourguignon et al. (2008) construct household income using models that include individual earnings, individual labour supply, occupational structure of household members, and household composition (including fertility), using these models to decompose income inequality into components related to all of these factors.

Our approach is simpler, focusing only on labour market earnings at the individual level. We do not attempt to model labour supply or employment, and do not aggregate individuals into households. Our analysis is based on individual-level earnings regressions that are used for a number of counterfactual simulations. One set of these simulations will look at the impact of changing the schooling distribution while holding the earnings-education gradient constant. Another will hold the schooling distribution constant while changing the earnings-education gradient. In order to focus on the impact of changes in the earnings-education gradient at different levels of education, we will simulate the effect of holding relative earnings constant at particular grades while allowing relative earnings at other grades to change as they did. This will allow us to answer questions such as how the decline in the relative earnings of people with incomplete secondary education affected the evolution of earnings inequality.

We begin by presenting some of our key theoretical results. We then provide examples of empirical analysis that are guided by these results. Finally, we use counterfactual simulations to analyse how changes in education distributions and the earnings-education gradient can explain why South Africa has had significant decreases in schooling inequality without experiencing similar declines in earnings inequality.

Theoretical Links between Schooling Inequality and Earnings Inequality

We begin our theoretical analysis with a simplified version of the standard Mincer human capital earnings equation. Leaving experience and other determinants of earnings aside for now, the logarithm of the ith worker's earnings can be expressed as

$$\log Y_i = \alpha + \beta S_i + \mu_i \tag{1}$$

where Y_i is earnings, S_i is years of schooling, and μ_i is a residual uncorrelated with schooling. Given Equation (1), the variance of log earnings, a standard mean-invariant measure of wage inequality, is

$$V(\log Y) = \beta^2 V(S) + V(u) \tag{2}$$

where V denotes variance.[1] This simple result demonstrates an important point about the link between schooling inequality and earnings inequality. If the relationship between schooling and earnings is log-linear as in (1), then earnings inequality (as measured by the variance of log earnings) is a linear function of the variance in schooling. This has a number of implications about the link between the distribution of schooling and the distribution of earnings. Suppose, for example, that we could double the schooling of every worker, holding returns to schooling constant. This would quadruple the variance in years of schooling and thus quadruple the "explained" component of earnings inequality. If we measure inequality in schooling by some standard mean-invariant measure of inequality, this doubling of schooling would imply no change in schooling inequality, but would significantly increase earnings inequality. As another example, giving each worker one additional year of schooling would unambiguously reduce schooling inequality, but would have no effect on earnings inequality.

[1] The variance of log earnings is an attractive mean-invariant measure of inequality because of its natural relationship to the Mincer earnings equation and because it gives more weight to transfers at the bottom of the income distribution. It has the unattractive feature, however, that it does not always obey the Dalton Principle of Transfers. Foster and Ok (1999) show examples in which transfers from high-income individuals to even higher-income individuals cause inequality to decrease rather than increase by the variance of logs. We use it here along with other inequality measures because of its simple connection to schooling and the returns to schooling.

The potential for inequality in schooling to decline as inequality in earnings increases was pointed out by Knight and Sabot (1983). Bourguignon et al. (2005) provide general conditions under which this may occur, pointing out that when the earnings-education gradient is convex (the log-linear wage equation being just one simple example), raising every person's education by a constant percentage will keep schooling inequality constant, but will increase earnings inequality, since those with higher education (and, hence, higher earnings) will experience larger proportional increases in earnings. Bourguignon et al (2005) point out that there may often be a "paradox of progress," in which increases in education lead to increasing earnings inequality because of the convexity of the earnings-education gradient. They found this to be a factor influencing trends in income inequality in Argentina, Columbia, Indonesia, Malaysia, and Mexico.

Generalizing the relationship between schooling and earnings

The simple Mincer earnings equation above assumes that there is a single rate of return to schooling that applies to all levels of schooling. One of the important recent patterns in returns to schooling in developed and developing countries, however, is the emergence of convex returns to schooling, with returns increasing at higher levels of schooling (especially post-secondary) at the same time that they have fallen at intermediate levels of schooling.

This pattern complicates what we mean when we consider the relationship between returns to schooling and earnings inequality. What happens to inequality if, for example, we increase earnings at grade 8, holding earnings at other grades constant. What if we increase earnings at grade 4 or grade 12? This section provides an analytical way to answer these questions.

Consider a very general model of the relationship between schooling and earnings,

$$y_i \equiv logY_i = \alpha + \sum_j \beta_j S_{ji} + \mu_i \tag{3}$$

where Y_i is earnings, y_i is the log of earnings S_{ji} is a 0,1 indicator for whether person i is in the jth schooling category (which could be single years of schooling in the most general case, but could also be larger categories), and μ_i is a residual uncorrelated with schooling.[2] Denote mean log earnings as $\bar{y}$ and mean log earnings for schooling level j as $\bar{y}_j$. Note that β_j is a multiplicative shift in earnings for group j relative to the omitted category. For example, an increase in β_9 by .01 will increase the earnings of individuals with 9 years of schooling by 1 percent relative to all other groups, holding earnings of all other schooling groups constant. We assume for the results below that changes in β_j only affect earnings for group j.

The following proposition describes the relationship between returns to schooling and earnings inequality.

[2] Note that nothing about this analysis requires that these be schooling categories. The same results would apply to any other categories, such as age, region, or gender.

Proposition 1: If $\hat{S}$ is a level of schooling for which $\bar{y}_j > \bar{y}, \forall_j > \hat{S}$ and $\bar{y}_j < \bar{y}, \forall_j < \hat{S}$, then increases in β_j for $j > \hat{S}$ will increase the variance of log earnings, and increases in β_j for $j < \hat{S}$ will increase the variance of log earnings.

To prove Proposition 1, note that the variance of log earnings for this more general model can be written as:

$$V(\log Y) = \sum_j \beta_j^2 V(S_j) - 2\sum_j \sum_{k \neq j} \beta_j \beta_k p_j p_k + V(\mu), \tag{4}$$

where p_j is the proportion of earners in schooling category j. Since the S_j terms only take on values of 0 or 1, $V(S_j) = p_j(1 - p_j)$. What happens to earnings inequality if we increase one of the β terms? This is an increase in earnings for one category of schooling (relative to some arbitrary omitted category) and, unlike the result in Equation (2), no longer translates necessarily into an increase in earnings inequality. We take the derivative of Equation (4) with respect to β_1, which could arbitrarily be assigned to any schooling category and thus is completely general:

$$\begin{aligned} \frac{\partial V(\log Y)}{\partial \beta_1} &= 2\beta_1 p_1 - 2\beta_1 p_1^2 - 2p_1 \sum_{j \neq 1} \beta_j p_j \\ &= 2p_1 \left[\beta_1 - \beta_1 p_1 - \sum_{j \neq 1} \beta_j p_j \right] \end{aligned} \tag{5}$$

Note that:

$$\beta_1 p_1 + \sum_{j \neq 1} \beta_j p_j = \bar{y} - \alpha,$$

where $\bar{y} = E(\log y)$, and $\alpha + \beta_1 = \bar{y}_1$, where $\bar{y}_1 = E(\log y \,|S_1 = 1)$. Substituting into (12), the result simplifies to:

$$\frac{\partial V(\log Y)}{\partial \beta_1} = 2p_1[\bar{y}_1 - \bar{y}], \tag{6}$$

or

$$dV(\log Y) = d\beta_1 * 2p_1[\bar{y}_1 - \bar{y}]. \tag{7}$$

The result is intuitive. Increasing β_1 will increase the earnings of the first schooling category (arbitrarily defined) relative to the omitted category, and therefore relative to every other category as well. This will be equalizing if the first category had a mean of log earnings below the overall mean of log earnings and will be disequalizing if this mean was above the overall mean. The magnitude of the change will depend on how far the group's mean is above or below the overall mean, and on the relative size of the group. For example, if the group's mean of log earnings was 0.1 below the overall log mean (in other words, a difference of approximately 10%), and if the group was 10% of the income earning population, an increase in β_1 of 0.01 would reduce the variance of log earnings by 2*0.1*0.1*0.01=0.0002.

Equation (6) calls attention to a statistic that we do not ordinarily calculate – the year of schooling for which mean log earnings is equal to or closest to overall mean log earnings. Suppose there is a level of schooling $\hat{s}$ such that $\bar{y}_i > \bar{y}, \forall i > \hat{s}$ and $\bar{y}_i < \bar{y}, \forall i < \hat{s}$. Then increasing returns to schooling for all years below $\hat{s}$ is equalizing and increasing returns for years above $\hat{s}$ is disequalizing. This is the result in Proposition 2.

The result in Equation (6) has parallels to research looking at how measures of income inequality are affected by additions to income at different points in the income distribution. Lambert and Lanza (2006) show that for all standard measures of inequality, there will always be a "benchmark income" that separates equalizing additions to income from disequalizing additions to income. They show that this benchmark income will typically be different for different measures of income inequality. Hoffman (2001) derived similar results for several inequality measures, and estimated the benchmark value for each measure in Brazil. Roope (2019) provides additional theoretical results, and Roope (2021) estimates income benchmarks for various inequality measures in a number of countries. Assuming there is a monotonic relationship between years of schooling and earnings, it follows from the results of Lambert and Lanza (2006) and Roope (2019) that there will be an analogous "schooling benchmark" for which increases in earnings below that level of schooling are equalizing, while increases in earnings above that level of schooling are disequalizing. It also follows that the benchmark year of schooling will typically be different for different measures of inequality.

In the case of the variance of log income, Lambert and Lanza (2006) find that the benchmark income is the geometric mean. This is consistent with our result that the benchmark level of schooling is the schooling level corresponding to mean log earnings. As can easily be shown, the mean of log earnings is equal to the log of the geometric mean of earnings. Thus, the schooling level corresponding to the geometric mean of earnings is also the schooling level corresponding to the mean of log earnings. So there are two ways to describe our result for the variance of log earnings. The schooling benchmark $\hat{s}$ is the level of schooling corresponding to the geometric mean of earnings, or, equivalently, the level of schooling corresponding to the mean of log earnings (which is easier to calculate than the geometric mean).

It is interesting to consider whether the year of schooling at which mean log earnings is reached is less than or greater than mean years of schooling. That is, is $\hat{s} - \bar{s}$ positive, negative, or zero? In the simple linear Mincer earnings equation, $\hat{s} = \bar{s}$, since mean log earnings will be earned by someone with mean schooling, abstracting from other variables such as age and experience. More generally, however, $\hat{s}$ could be greater or less than $\bar{s}$, depending on whether returns to schooling are concave or convex in schooling. If returns are convex, as they have been in South Africa and in many other countries in recent years, then $\hat{s} > \bar{s}$. The year of schooling associated with mean log earnings is above mean schooling. This means that an increase in earnings will be equalizing even for some years above mean schooling. We look at this empirically below for South Africa.

Another interesting question is what happens when we change the distribution of schooling. One simple way to model this is to imagine shifting people from some arbitrary group 2 to some arbitrary group 1, so that $dp_2 = -dp_1$, or, equivalently, $\partial p_2/\partial p_1 = -1$.

$$\frac{\partial V(\log Y)}{\partial p_1} = \beta_1^2 - 2\beta_1^2 p_1 + (\beta_2^2 - 2\beta_2^2 p_2)\frac{\partial p_2}{\partial p_1} - 2\beta_1 \sum_{j \neq 1} \beta_j p_j - 2\beta_2 \sum_{j \neq 2} \beta_j p_j \frac{\partial p_2}{\partial p_1}$$

$$= (\bar{y}_1 - \bar{y})^2 - (\bar{y}_2 - \bar{y})^2 \,. \qquad (8)$$

The result is once again intuitive. Shifting the population from one group to another will be disequalizing if the second group has mean log earnings that are further from the mean (in absolute value) than the first group. For example, if mean log earnings for group 2 is 0.2 above overall mean log earnings, while mean log earnings for group 1 is 0.1 below the overall mean, then shifting 10% of the population from group 2 to group 1 will change the variance of log earnings by $(0.1^2 - 0.2^2)*0.1 = -0.003$. As above, an interesting point of reference is the level of schooling corresponding to mean log earnings. The generalization of Equation (8) is that changes in the schooling distribution that push the distribution toward the level of schooling with mean log earnings will be equalizing, while changes in the distribution that push the distribution away from the level of schooling with mean log earnings will be disequalizing. Recall that if returns to schooling are convex, the critical level of schooling will be higher than mean schooling.

The result in Equation (8) provides some insights into when we will observe the "paradox of progress" discussed by Bourguignon et al. (2005). Even if the earnings-education gradient is highly convex, simply moving people up the education distribution will not necessarily increase earnings inequality. In the case of the variance of log earnings, the critical question is whether we move them closer to the mean of log earnings or away from the mean of log earnings. Thus, shifting someone from grade 7 to grade 10 might be equalizing, while shifting them from grade 12 to university might be disequalizing. We explore this empirically below.

The result in Equation (8) can be applied to any variance. We have applied it to the variance of log earnings, which is a mean-adjusted measure of inequality. We could use it to talk about inequality in schooling by noting that we will reduce inequality if we reduce the variance while raising the mean. Using the result in (8), we will do this for the distribution of schooling if we shift people upward in the distribution so that we raise the mean, while on average moving people closer to the mean.

Other measures of inequality

The results above apply to the variance of log earnings, one measure of inequality. We can also consider what happens to other measures of inequality when earnings change at different levels of schooling, continuing to assume that the relationship between schooling and earnings is given by the flexible log earnings function in Equation (3). One measure that has a simple analytical result is the Generalized Entropy Measure GE(0), also known as the mean log deviation or Theil's L. The between group component of GE(0), where the groups are schooling levels, can be written as

$$GE(0) = \log \sum_{i=1}^{n} p_i Y_i - \sum_{i=1}^{n} p_i \log(\bar{Y}_i) \ , \tag{9}$$

where $\bar{Y}_i$ is the mean of earnings at schooling level i. We assume that the within-group component remains constant when we change mean earnings at a given schooling level (i.e., change some β). Taking the derivative with respect to some β_1:

$$\begin{aligned} \frac{\partial GE(0)}{\partial \beta_1} &= \frac{p_1}{\bar{Y}} - p_1 \frac{\partial log \bar{Y}_1}{\partial \beta_1} \\ &= p_1 \left(\frac{\bar{Y}_1}{\bar{Y}} - 1 \right) \end{aligned} \tag{10}$$

where p_1 is, as above, the proportion of the population in schooling group 1. Note that Equation (10) will be positive when $\bar{Y}_1 > \bar{Y}$ and will be negative when $\bar{Y}_1 < \bar{Y}$. That is, if group 1 has mean earnings above (below) the overall mean, then an increase in β_1 will increase (decrease) inequality as measured by GE(0). This mirrors the result shown in Hoffman (2001), Lambert and Lanza (2006), and Roope (2019) that the mean is the benchmark income for the GE(0) measure. The difference from the result for the variance of logs is that the sign now depends on the difference between the group's mean earnings relative to overall mean earnings, whereas the result for the variance of logs depends on the difference between the group's mean log earnings and the overall mean log earnings. As we show below, the schooling level corresponding to mean earnings will typically be higher than the schooling level corresponding to mean log earnings, given the convex relationship between schooling and earnings.

In general, the benchmark level of schooling will be different for every measure of inequality, since different measures of inequality are differentially sensitive to income in different parts of the distribution. Lambert and Lanza (2006) provide very general theoretical results about the benchmark income for different measures of inequality. As implied by their results for benchmark incomes, most other well-known measures of inequality do not lend themselves to the kind of simple analytical results shown above for the variance of logs and the GE(0) measure. As we demonstrate below, however, it is always possible to analyse the sensitivity of a given inequality measure to changes in earnings at a given schooling level through counterfactual simulations. For example, if we counterfactually raise the earnings of all workers with 4 years of schooling by 1%, leaving all other earnings constant, we can compare inequality measures before and after the adjustment. Doing this for all years of schooling, we can identify the level of schooling at which an increase in earnings at that schooling level switches from being equalizing to disequalizing.

Data and Sample

We use earnings and education data from household and labour market surveys between 1994 and 2019. We combine data from the Post-Apartheid Labour Market Series (PALMS) with General Household Survey (GHS) data for the period 2008-2019. PALMS is a series of stacked and harmonized household surveys comprised of the October Household Surveys, the Labour Force Surveys (LFS) and the Quarterly Labour Force Surveys (QLFS) (see Kerr et al. 2019). As documented in Kerr (2020), however, the earnings data in the QLFS results in unrealistic estimates and the GHS surveys are therefore a better source of earnings information from 2008 onwards. We therefore replace the post 2007 series in PALMS with information from the GHS (StatsSA, 2008-19). We also restrict the data between 2000 and 2007 to the September wave of the LFS, such that we have one survey per year. We do not extend the analysis beyond 2019 because of inconsistencies in the data beginning in 2020.

All earnings data are reported in real terms[3], and all data are weighted to be nationally representative[4]. Respondents provide earnings information either as point estimates or in brackets. Earnings provided in brackets are accounted for by proportionally increasing the weight of similar respondents who provide actual earnings (Kerr & Wittenberg, 2019b). Outliers are remove based on a regression-based method[5] (Kerr & Wittenberg, 2019). It is not possible to accurately identify the full labour force in the GHS (Kerr, 2020). This informs our choice of sample for the analysis. We restrict the sample to adults aged 25 to 59 and present estimates either for the population or for respondents with positive earnings.

The years of education variable is constructed from the survey question that asks the respondent's highest completed level of education. The categories in this variable have changed over time (Branson et al, 2020), with categories for lower and post-secondary qualifications expanding. To create a consistent measure over time, we classify certificates and diplomas with less than grade 12 as 11 years of education, NTC 1-3 as years 10 -12, diploma and certificates with grade 12 as 13 years of education[6] and all other higher qualifications as 15 years of education.

Table 1 presents summary information for our sample of positive earners by year. Average

---

[3] Adjusted to 2019-rand values based on CPI (https://www.statssa.gov.za/publications/P0141/CPIHistory.pdf?)

[4] PALMS includes weights calibrated to a consistent demographic model. We utilise these weights for the period until 2007. For the GHS, we use the survey provided person weight.

[5] Earning bracket weights and an identifier for outliers is provided in the PALMS data. We replicate the approach used in PALMS in the GHS series.

[6] Note that certificate with grade 12 is classified as NQF 5 while some diplomas with grade 12 are at NQF level 6. The earlier surveys group these two categories together and we therefore retain this grouping for consistency.

years of education increases steadily from 8.3 years in 1994 to 10.7 years in 2019. Mean real earnings increased substantially over the period, from R7,283 to R11,609. The racial and gender composition of earners has also changed. In 1994, 37% of earners were female; this share increased to 44% by 2019. The share of African earners has increased from 71% in 1994 to 83% in 2019, offset predominantly by a decline in the share of White earners. These changes in the racial and gender composition will be shown below to have a significant effect on overall earnings inequality.

Empirical Results

We use the data to analyse schooling inequality and earnings inequality during the post-apartheid period, using our theoretical results to guide the analysis. We first look at how the distribution of schooling evolved after 1994, documenting the substantial decrease in schooling inequality that has taken place over this period. We then look at trends in earnings inequality. Drawing on our theoretical results, we show how the schooling levels associated with mean log earnings and mean earnings changed over time, and how this links to changes in returns to schooling. We simulate the effect on aggregate earnings inequality of a 1% increase in earnings at each year of schooling, while keeping the distribution of schooling constant. Finally, we estimate counterfactual trends in earnings inequality that illustrate the effects of changes in the schooling distribution and changes in the relationship between schooling and earnings.

Changes in the distribution of education, 1994-2019

Figure 1 presents the cumulative distribution functions (CDFs) of schooling attainment for the population (top figure) and positive earners (bottom figure) aged 25 to 59 in South Africa over the 1995 to 2019 period. The vertical line represents completed secondary education -- 12 years. It is striking to see how quickly educational attainment increased over the analysis period. In 1995, approximately 70% of adult South Africans had completed grade 11 or less. By 2019, this share had decreased to 50%, with adults being more likely to complete secondary education and continue into post-secondary education. The bottom figure presents similar information for positive earners only and shows that the education levels of earners is, on average, higher than for the population as a whole.

An important feature of Figure 1 is that the CDFs imply first order stochastic dominance in comparisons of every earlier schooling distribution. Among other things, this implies Lorenz dominance, meaning that inequality in years of schooling unambiguously improved over time by any standard inequality measure. Beyond that, first order stochastic dominance implies Generalized Lorenz Curve dominance, meaning that there is a steady improvement in schooling inequality accompanied by a rising mean. Put another way, the schooling distribution unambiguously improved by any concave social welfare function, no matter what weights are put on different parts of the schooling distribution.

Figure 2 allows us to unpack these CDFs in more detail, as we can see where the gains and losses took place over the period. The proportion of earners with less than 7 years of schooling (completed primary) decreased rapidly, from 41% in 1994 to 13% in 2019. The share with completed secondary education (12 years and up) expanded the most, particularly from 2000 onwards. By 2019, the share has increased to close to 50%, more than double the share seen in 1994 (23%). A characteristic feature of the South African schooling distribution is that the largest share of earners fall within the incomplete secondary school (grade 8 through 11) group. This group increased modestly as a proportion of all earners, from 35% in 1994 to 39% in 2019. This group will play an important role in our analysis below.

Measures of schooling and earnings inequality

Figure 3 shows some key measures of the distribution of education for positive earners. The top panel shows the increase in mean years of education, which rose from 7.9 in 1994 to 10.6 in 2019. Since we are showing the population with positive earnings, rather than the total population, the volatility in the early years of the period is driven mainly by changes in labour force participation and employment rather than changes in the education of the overall population (when we look at the overall population aged 25-59 there is a steady increase in the mean over time). The standard deviation of schooling, a key determinant of earnings inequality in the standard human capital earnings equation, was relatively constant at about 4 years of education in the early years of the period, followed by a steady decline to 3 years in 2019. As shown in Equation (2), a decrease in the variance (or standard deviation) of schooling will lead to a decline in the variance of log earnings in a simple Mincer earnings equation if there is no change in returns to schooling. While the variance in years of schooling tends to increase as mean schooling increases at low levels of schooling, as shown in Lam and Levison (1992) and Lam (2020), by 1994 South Africa was at the point where the variance in years of schooling was levelling off, with steady declines after 2000. This suggests that improvements in the schooling distribution should have pushed in the direction of declines in earnings inequality. Below we show that this was in fact the case using counterfactual simulations that allow for a much more general relationship between schooling and earnings.

In the bottom panel of Figure 3 we see that the coefficient of variation, a simple mean-adjusted measure of education inequality (the standard deviation divided by the mean), shows a rapid decline over the period, from 0.55 to in 1994 to 0.29 in 2019. The bottom panel also shows the Gini coefficient for years of completed education for positive earners. The Gini also falls steadily over time, falling by almost 40 percent between 1994 and 2019. These declines in schooling inequality are consistent with the first-order stochastic dominance in the CDFs in Figure 1, which imply that inequality in education should have declined by any standard inequality measure. This reflects the fact that the schooling distribution has "compressed" over time, with big declines in the proportions with low levels of schooling and with a ceiling on the extent to which education can increase at the top of the distribution. This pattern is observed in virtually all countries at some point as education expands (Lam 2020), leading to a decrease in schooling inequality.

Turning to measures of inequality in earnings, the top panel of Figure 4 shows the variance of log earnings for the sample of all men and women aged 25-59 with positive earnings. The top panel also shows explained variance and residual variance from a log earnings regression estimated for each year that includes schooling dummies for each year of schooling along with age, age squared, racial group dummies and a dummy variable for female. The average R^2 in these earnings regressions is around 0.47, much higher than is found in similar earnings regressions in the United States. (Lam and Levison 1992). The explained variance fairly closely tracks the total variance. Overall the explained variance (the inequality predicted by the combination of the schooling distribution and the returns to schooling, along with race and gender) is relatively flat over most of the post-apartheid period. Residual variance, the component of variance that is uncorrelated with schooling, age, race, and gender, rises around 1999, falls, and then rises again toward the end of the period.

The bottom panel of Figure 4 shows two other earnings inequality measures – the Generalized Entropy (0) measure (GE(0), or Theil L) and the Gini coefficient. By these measures earnings inequality is slightly higher at the end of the period than it was at the start, with an overall upward trend. Taking Figures 3 and 4 together, an important puzzle is why earnings inequality did not decrease over the post-apartheid period despite the significant decreases in schooling inequality. As we will see below, changes in the structure of the earnings-education gradient are central in explaining this pattern.

The relationship between returns to schooling and earnings inequality

Figure 5 shows two statistics that we argue are key to understanding the relationship between the earnings-education gradient and earnings inequality. The first is the level of schooling at which mean log earnings is reached, the crossover point described in Equation (6) for the variance of log earnings. The second is the level of schooling at which mean earnings is reached, the crossover point described in Equation (10) for the GE(0) measure. The figure also shows the mean level of schooling for those with positive earnings, a useful point of comparison. These are estimated using the 30-39 age group in order to control for confounding effects of age. As seen in the figure, mean years of completed schooling over the period rose from 7.5 to 10.4 years, while the education level associated with mean log earnings rose from 9 years to 11.8 years. The education level associated with mean earnings rose from 11.1 years to 12.4 years. Note that the level of schooling corresponding to mean log earnings is above the mean level of schooling throughout the period. This implies that returns to schooling are convex in schooling in all periods. In fact, the gap between the two measures widened between 1994 and 2007, indicating increasing convexity in returns to schooling, although the gap has subsequently reduced. An important implication of Figure 5 is that an increase in earnings for earners with schooling in the grade 9 to grade 11 range, holding earnings at other grades constant, would have been disequalizing in the 1990s, based on the variance of log earnings, but would have been equalizing from 2007 onwards. This pattern suggests that it must have been the case that returns to education increased more in higher schooling categories (those with earnings above the mean and log mean) relative to lower categories.

This increase in returns to education at higher education levels is confirmed in Figure 6. The figure shows what has happened to returns to schooling at different points in the schooling distribution, using cutoffs for primary (grades 0-7), incomplete secondary (grades 8-11), completion of secondary (grade 12) and post-secondary (beyond grade 12). The figure shows the average returns per year of schooling in each of these schooling groups, based on a regression of log earnings on single year schooling dummies. For example, in 1995, each additional year of schooling between grade 8 and grade 11 was associated with increased earnings of almost 20%. The figure shows that South Africa has seen a dramatic increase in returns to grade 12 and above since 1994. The average increase in earning associated with each additional year of schooling beyond grade 12 rose from 27% in 1994 to 65% in 2019. As we will see below, our simulations indicate that this is a key factor explaining why improvements in schooling inequality have not led to decreases in earnings inequality. Another key feature of Figure 6 is the decline in returns to schooling in the incomplete secondary category, though these rebounded somewhat in the end of the period. Returns to schooling have also tended to decline for the primary grades.

Looking at the returns to education for each year of schooling provides a useful but somewhat limited view of the interaction between the distribution of schooling and the distribution of earnings. What we are also interested in – as motivated by our theoretical results – is how total earnings inequality would change if we increased earnings at a particular year of schooling while holding earnings at all other levels and the distribution of schooling itself constant. Recall from Equation (6) that for the variance of log earnings this is a combination of a) how big the gap is between mean log earnings for that year of schooling and overall mean log earnings, and b) the relative size of the group. We can see in Figure 5 where overall earnings inequality would go up and down if we increased earnings at a particular year of schooling. We can go further, however, and look at the magnitude of that change.

Figure 7 plots what happens to overall inequality for a 1% increase in earnings at a given level of schooling, *ceteris paribus*. The top panel shows the results for 1995, and the bottom panel shows the results for 2019. The horizontal line at 0 on the y-axis is the crossover point at which increasing earnings above that level of schooling begin to increase, rather than decrease, earnings inequality. For the variance of log earnings, this crossing point always corresponds with the education level at which mean log earnings is realized, as shown in Equation (6). Three other inequality measures are also presented in the figures. These are the Gini coefficient, the generalized entropy GE(0) measure, and the generalized entropy GE(1) measure. As shown in

Equation (10), the crossing point for the GE(0) measure corresponds to the education level associated with mean earnings. The convex relationship between education and earnings means that the crossover point for the variance of log earnings must be before that of the GE(0) measure, and this is confirmed in Figure 7. The crossover points for the Gini coefficient and the GE(1) measure are to the right of the other two measures.

The top panel of Figure 7 shows that increasing earnings for those with zero schooling has a large impact on reducing earnings inequality in 1995. This occurs for two reasons. First, earnings at this level are far below mean log earnings, mean earnings, and the cutoffs implied for other inequality measures, implying that additional income has a large equalizing effect. Second, this group constituted 10% of the adult population with earnings in 1995, giving extra weight to income additions at this level. As expected, we see in the figures that increased earnings at the highest levels of education increase inequality, as they must at some point, with the magnitudes differing for different inequality measures. The differences result from the fact that different measures of inequality have different degrees of sensitivity to increasing incomes at the top of the distribution. The variance of logs, for example, is known to be particularly sensitive to adding income at the bottom of the distribution and to be relatively insensitive to adding income at the top. Both of these features can be seen in Figure 7. The Gini coefficient is less sensitive than other inequality measures to income at both the top and the bottom of the distribution, and this is also evident in Figure 7.

Comparing the top and bottom panels of Figure 7, the crossover points defining the level of education at which increases in income switch from being equalizing to disequalizing shift to the right over time. By 2019 the crossover point is between grades 11 and 12 for the variance of logs, and above grade 12 for the other three measures. The equalizing effect of adding income at the lowest levels of schooling declines over time as the proportion of earners at those schooling levels declines. Note that by 2019 the impact of adding income at grades 9, 10, and 11 has become highly equalizing. This indicates that the levels of schooling corresponding to mean earnings and mean log earnings have increased. Whereas additions to the earnings of earners with 11 years of schooling would have either been disequalizing or roughly neutral in 1995, additions to earnings at grade 11 have become highly equalizing in 2019. The large magnitude of the effect at grades 10 and 11 also reflect the increasing share of earners at these grade levels in 2019.

Figure 8 provides another way of seeing how returns to schooling have changed over time and how these changes are related to the way in which changes in earnings at different points in the schooling distribution affect earnings inequality. The figure takes the predicted log earnings at each schooling level (predicted using separate regressions for each year, race, and gender group using a regression that includes single-year schooling dummies plus age and age squared) and subtracts the overall mean log earnings in that year. For grade levels with predicted log earnings below the zero line in a given year, increased earnings would be inequality-reducing for the variance of log earnings measure. For grade levels with predicted log earnings above the zero line, increased earnings would be inequality-increasing for the variance of logs.

The top panel of Figure 8 shows results for African males. In 1994, mean log earnings of African males with grade 9 education was roughly equal to the overall mean of log earnings for the full population. From Equation 6, this means that raising earnings of African males with schooling below grade 9 would have reduced overall variance of log earnings, while raising earnings of African males with schooling above grade 9 would have increased it. Over time this cutoff level of schooling increases. By 2019 earnings at grade 9 for African males was well below the overall mean. Increasing earnings of African males at any grade below grade 12 would have decreased inequality in 2019. Conversely, decreasing relative earnings at African males at grades 10 and 11, which is what actually happened, would have increased inequality in 2002 and later, even though the same change would have decreased inequality in 1994.

The second panel of Figure 8 shows results for African females. The benchmark mean log earnings for each year is the overall mean log earnings, the same as it is for all of the groups shown in Figure 8. We see that mean log earnings of African females are substantially lower than for African males at every grade in every year. For example, while mean log earnings of African males at grade 9 was roughly equal to the overall mean in 1994, African females had earnings 0.4 log points below the overall mean, implying that African male earnings were over 50% higher than those of African females. In 1994, earnings of African females did not reach the overall mean until they reached 12 years of schooling, compared to 9 years for African males. Thus, increasing earnings of African females for any level of education below 12 years would have been equalizing in 1994. Since less than 5% of African females had education over 12 years in 1994, increasing earnings of African females would have almost always decreased earnings inequality. As is the case for African males, for African females earnings in grades 8-11 fall increasingly below the overall mean over time. While African females with grade 9 were 0.4 log points below the mean in 1994, in 2019 they were 1.1 log points below the mean. Exponentiating, African females with grade 9 education went from 70% of mean earnings in 1994 to 35% of mean earnings in 2019. Given the large fraction of women with incomplete secondary schooling, these declines in relative earnings at these schooling levels would have had a significant negative effect on overall earnings inequality.

Returns to post-secondary education were very high for African males and females in all years, with the premium for university education increasing over time. Having a university education implied earnings of about 1.2 log point above the mean for African males in 1994, rising to 1.6 log points in 2019. This implies that African males with university education went roughly from having double mean earnings to having triple mean earnings. Given that grade 12 earnings declined relative to the mean over this period, the returns to university relative to grade 12 went up even more. Between 1994 and 2019, the increase in earnings from grade 12 to university went from 115% to 350% for African males, and from 65% to 320% for African females. As is clear from our analytical results, this large increase in relative earnings for university graduates would have significantly increased overall earnings inequality.

The third and fourth panels of Figure 8 show results for White males and females. Because there are very few Whites with schooling below 10 years in the data used here, we show the results beginning at 10 years of education. Several key features can be seen in these figures. First, earnings of Whites are consistently well above earnings of Africans. For example, mean earnings of White males with 12 years of education were 3.4 times the mean earnings of African males with 12 years of education in 2019. Second, log earnings of White males and females are above the overall mean log earnings at every grade shown, with the exception of White females with grade 10 in 2019. This means that increasing earnings of Whites at every level of education above grade 10 would have increased overall inequality in every year.

Figure 8 also illustrates the increased convexity of returns to schooling that has been observed over time, with the biggest changes taking place in roughly the first 10 years after 1994. In 1994 log earnings were almost linear in schooling for all groups. By 2006 the gradient is much more convex. This can be seen especially for African males and females, with the slope falling in the middle grades, especially grades 7 to 11, and the slope increasing at grade 12 and above. Note that while the return to grade 12 relative to grade 11 increases significantly over time, earnings at grade 12 relative to the mean actually decline over time. For African males, earnings at grade 12 were 0.4 log points above the mean in 1994, but fell to only 0.2 log points above the mean in 2019. For African females, earnings at grade 12 were 0.3 log points above the mean in 1994, but fell to 0.2 log points below the mean in 2019.

The effect of changes in the schooling distribution

Equation (8) provides a simple analytical result about how changes in the distribution of schooling will affect earnings inequality using the variance of log earnings. For any measure of inequality there will be some upward movements in the schooling distribution (shifting people from one schooling level to some higher schooling level) that will increase inequality, some that will decrease inequality, and some that will hold inequality constant. The basic intuition is that shifting people upward in the distribution (for example, moving some people from 9 years of schooling to 12 years of schooling) will reduce earnings inequality if we move people closer to the mean (or some benchmark measure of central tendency), and will increase inequality if we move people away from the mean. As shown in Equation (8), moving a person from schooling level S_1 to schooling level S_2 will have no effect on the variance of log earnings if log earnings at S_1 and log earnings at S_2 are the same distance from the overall log mean in absolute value. The details will be different for different measures of inequality.

As with our results regarding the impact of increasing earnings at a given level of schooling, the impact of a given shift in the schooling distribution is likely to change over time.

Shifting people from grade 9 to grade 12, for example, may be disequalizing when grade 9 earnings are closer to the mean than are grade 12 earnings, while the same shift may be equalizing if grade 12 earnings become closer to the overall mean. The impact of changing schooling distributions may also differ by population group. For example, shifting people in one race/gender group from grade 9 to grade 12 may be equalizing, but the same change may be disequalizing for another group, depending on how earnings at each grade for that race/gender group compare to the overall mean.

To see these effects, we simulate the effect of shifting some fraction of the population in a given race/gender group from one level of schooling to some higher level of schooling, holding earnings at each level of schooling constant. The upper panel of Figure 9 shows the results for African females from 1994 to 2019. Each line shows the impact on the overall variance of log earnings of moving 1% of the African female population from one level of schooling to some higher level of schooling, assigning them the mean log earnings at each grade. The top line shows the impact of moving 1% of African females from grade 12 to university. Not surprisingly, given the result in Equation (8), this would increase overall earnings inequality in every year, an indication that this shifts people away from the level of education associated with mean log earnings. To take one specific example, in 2019 mean log earnings for African females with 12 years of education was 0.1 log points above overall mean log earnings, while mean log earnings for African females with university was 1.6 log points above overall mean log earnings. Shifting African females from grade 12 to university must therefore increase the variance of log earnings. As shown in the bottom panel, it is also the case for African males that a shift of 1% from grade 12 to university is disequalizing in every year.

The bottom line in the upper panel shows the impact of moving 1% of African females from grade 7 to grade 12. This reduces overall earnings inequality in every year, indicating that this is a move *toward* the level of education associated with mean log earnings. Looking at the bottom line in the lower panel, the result for African males is quite different. In 1994 the impact of moving 1% of African males from grade 7 to grade 12 is actually disequalizing. This is because mean log earnings for African males with grade 7 in 1994 was 0.28 below overall mean log earnings, while mean log earnings for African males with grade 12 was 0.38 above the mean. The net effect of shifting an African male from grade 7 to grade 12, assigning them mean log earnings of African males at each grade, is to move the person farther above the mean log earnings than the distance they were initially below it. Over time the effect of moving African males from grade 7 to grade 12 becomes equalizing, although the magnitude of the effect is much smaller than it is for African females. The gender difference reflects the fact that for African females, grade 12 earnings are below overall mean log earnings for the entire period, while for African males grade 12 earnings are above mean log earnings in most years. For African females, moving from grade 7 to grade 12 is always highly equalizing (moving from well below mean log earnings to something not as far). For African males, moving from grade 7 to grade 12 means moving from below mean log earnings to above mean log earnings in most years. Whether this is equalizing or disequalizing depends on whether grade 7 earnings or grade 12 earnings are closer to mean log earnings. Over time the education level corresponding to mean log earnings rises, as shown above. For African males this causes mean log earnings to move further above grade 7 earnings and move closer to grade 12 earnings, with the result that a hypothetical shift from grade 7 to grade 12 becomes more equalizing.

The red line (second from top) in Figure 9 shows the impact of moving 1% of the population from grade 10 to what we call grade 13 (some post-secondary education short of university). For African females, this shift increases overall earnings inequality in 1994 and 1995, but reduces inequality in most later years. For African males, however, this shift is always inequality increasing. The reason for the gender difference is that African male earnings are higher than African female earnings at both grade 10 and grade 13. For African males, grade 10 earnings are only slightly below overall mean log earnings, while grade 13 earnings are well above mean log earnings. For African females, the amount by which grade 10 earnings fall below mean log earnings is larger than the amount by which grade 13 earnings exceed mean log earnings.

One lesson of Figure 9, then, is that the threshold for the "paradox of progress" identified by Bourguignon, Ferreira, and Lustig (2005), in which rising education can increase earnings inequality, will tend to vary over time and may not be the same for different sub-populations. We see in Figure 9 that there are points in the schooling distribution for which increasing education for African women would have decreased overall earnings inequality, while increasing education for African men would have increased inequality, even though the earnings-education gradient is highly convex for both groups. For the White population (not shown), increasing education in the range at which most White education is concentrated would have increased earnings inequality over most of the post-apartheid period. A shift of population from grade 10 to grade 12 for White males and females, for example, would have increased overall earnings inequality in almost every year, while the same change for African males and females would have decreased inequality in almost every year. Whether a general improvement in the education distribution increases or decreases earnings inequality will depend, then, on how big the changes are at different points in the distribution, and whether the net effect is to move the population closer to mean earnings or farther away from mean earnings (or some benchmark earnings level that will differ for different measures of inequality).

Changes in schooling distributions over time are not as simple as shifting some fraction of the population from one single year of schooling to some higher year of schooling. Changes occur over the entire distribution, with decreases in the proportion at some years of schooling and increases in the proportion at other years of schooling. As shown in Figures 1 and 2, for South Africa in the post-apartheid period there has been a large reduction in the proportion with the lowest levels of schooling (primary and below), large increases in the proportion with grade 12 education, and smaller increases in the proportion with incomplete secondary and university education. In order to see the combined impact of all these changes, we will use counterfactual simulations that look at what would have happened to earnings inequality if the distribution of schooling evolved as it did, but the earnings-schooling gradient had not changed.

Counterfactual simulations of earnings inequality

One interesting counterfactual simulation is to consider what would have happened to earnings inequality in the post-apartheid period if the earnings-education gradient had not changed, but the distribution of education changed as it actually did. A contrasting simulation is to look at what would have happened if the earnings-education gradient changed as it did, but the distribution of education had not changed. Figure 10 shows these two simulations, and compares them to the actual change in earnings inequality, using both the variance of log earnings and the Gini coefficient. We use 1997 as our baseline year in order to avoid some of the anomalies in the data for 1994 and 1995. We also show what would have happened if, in addition to the distribution of education remaining constant, the distribution of race and gender had remained constant in the population of South Africans aged 25-59 with positive earnings.

For our counterfactual simulations, log earnings is regressed on single year schooling dummies, age, and age squared for each race and gender group in each year. Predicted values from these regressions are used to assign earnings for each individual in a given year based on their race, gender, age, and education. A residual is added to each individual's earnings by multiplying that person's actual residual by the ratio of overall residual variance for that gender/race/year to the residual variance for that gender and race group in 1997. This keeps residual variance constant across years and removes the effect of the changes in residual variance shown in Figure 4.

The top panel shows simulations for the variance of log earnings. The black line shows the actual changes in the variance of log earnings, standardizing residual variance in each year to be the same as in 1997. The bottom line, in blue, shows a simulation in which mean log earnings at each schooling level is held at its 1997 level for every race and gender group, with residual variance rescaled to 1997. This simulation keeps the education-earnings gradient constant over time, while allowing the schooling distribution to change as it actually did. It thus provides an estimate of what would have happened to earnings inequality as a result of changes in the schooling distribution alone. The line shows that changes in the schooling distribution would have caused the variance of log earnings in 2019 to be almost 20% lower than in 1997, and over 30% lower than its actual 2019 level, if the earnings-education gradient had remained the same as in 1997. The bottom panel shows the same counterfactual for the Gini coefficient. The decline in inequality is smaller using the Gini, but we get the same key result that inequality would have been lower in every year if the earnings-education gradient had not changed. Changes in the schooling distribution would have caused the Gini coefficient in 2019 to be 15% lower than than its actual level if the earnings-education gradient had been the same as in 1997.

The top red line in both panels of Figure 10 shows the simulation in which the distribution of the population by education, race, and gender is normalized to be the same in every year as it was in 1997, while allowing the earnings-education gradient to change as it actually did. This is done by reweighting the sample in every year by multipliers that adjust the relative size of each education/race/gender cell to be the same as in 1997. Interestingly, the red line in both panels shows that if the education/race/gender composition of the population had remained constant, earnings inequality would have been significantly higher in the later years of the period. In other words, something about the change in the distribution of education, race, and gender caused overall earnings inequality to be more equal than it would have otherwise been.

The purple line shows what would have happened if we held the distribution of education constant but allowed the distribution of race and gender to change as they actually did. Under this scenario, inequality is lower than its actual level when education does not change but race and gender do. This suggests that the result in the red line is driven by the fact that changes in the distribution of race and gender tended to reduce earnings inequality. From Table 1 we know that the income-earning population experienced an increase in the proportion African and the proportion female over time. This is due to both changes in the racial composition of the overall population and increases in labour force participation of Africans, especially African women. While these changes could have been either equalizing or disequalizing, we know from our theoretical results that changes in the population will tend to be equalizing if they increase the share of the population in the middle of the income distribution. Since the White population was and continues to be highly concentrated at the top end of the income distribution, a decreasing share of Whites in the population will tend to reduce earnings inequality.

The purple line in the two panels of Figure 10, which shows what would have happened if the distribution of education had been constant while the earnings-education gradient changed, lies below the black line, which is the line showing the actual trend in earnings inequality. This implies that earnings inequality would have been lower if the distribution of education had remained constant, given the changes in the earnings-education gradient. At first glance this seems surprising. The bottom blue line indicates that earnings inequality would have declined significantly if the earnings-education gradient had been constant while the education distribution followed its actual history. We might therefore have expected that holding the education distribution constant and allowing the earnings-education gradient to change would have caused an even larger increase in inequality than actually occurred. In other words, we might have thought that the black line (actual inequality) would have fallen between the purple line (inequality if education were constant) and the blue line (inequality if the earnings-education gradient were constant). The pattern in Figure 10 suggests that there was an inequality-increasing interaction between the changes in the distribution of education and the changes in the earnings-education gradient.

Our theoretical results above help us understand this interaction. While the improvements in the distribution of education were inequality-reducing in the absence of changes in the earnings-education gradient, as shown in the blue line in Figure 10, they also tended to reinforce the inequality-increasing effects of the changes in the earnings-education gradient. There were two main features of the change in the earnings-education gradient. The first was an increase in returns to post-secondary education, especially university. This was combined with an increase in the proportion of the population going beyond secondary, as shown in Figure 2. The percentage of the income-earning population aged 25-59 with university education or higher rose from 2% in 1994 to 8% in 2019. This made the inequality-increasing impacts of the rising returns to university education even greater than they would have otherwise been. The second key feature of changes in the earnings-education gradient was a decrease in the relative earnings of those with incomplete secondary education. As seen in Figure 2, there was an increase over time in the share of the population that stopped school after incomplete secondary. The percentage of the income-earning population with completed schooling of grade 8 through grade 11 went from 35% in 1994 to 39% in 2019. This created another inequality-increasing interaction of the change in the distribution of education with the change in the earnings-education gradient.

We now turn to counterfactuals in which we take a narrower look at the impact of changes in the earnings-education gradient. In our simulations we will add or subtract earnings to earners at selected points in the schooling distribution in order to offset particular changes in the earnings-schooling gradient observed in Figure 8. Earnings of individuals with a given level of schooling are adjusted by an amount that would move the mean for that group to be the same distance from the overall mean as it was for people with the same race, gender, and education in 1997. For example, one simulation will adjust earnings at grade 12 in every year so that the gap between mean earnings at grade 12 and overall mean earnings is the same as it was in 1997 for each race and gender group. Earnings inequality will then be calculated for this new hypothetical earnings distribution, allowing the distribution of education to change as it actually did. This will provide evidence about whether changes in relative earnings at grade 12 have been equalizing or disequalizing over time. As above, we adjust residual earnings so that residual variance in every year is the same as in 1997.

Our counterfactual simulations will increase earnings for some education/race/gender groups and decrease earnings for others. Table 2 shows the gap between mean log earnings for selected education/race/gender cells and overall mean log earnings in 1997 and 2019. For example, Column 1 of Table 2 shows that mean log earnings for African males with grade 8 education was 0.10 less than overall mean log earnings in 1997. In 2019 this gap had increased to 0.59. In our counterfactual simulation, African males with grade 8 education in 2019 have their log earnings increased by 0.49 to offset the decline that occurred in the relative earnings between 1997 and 2019. Looking at the column for those with grade 15+ (university diploma or above), every race/gender group had mean log earnings above overall mean log earnings in both 1997 and 2019, with the gap increasing over time. For African males with university education, our counterfactual simulation lowers their log earnings by 0.7 in 2019, bringing the average university premium for this group back to its 1997 level.

The top panel of Figure 11 shows what happens to the variance of log earnings in these counterfactual simulations. As in Figure 10, the black line shows the actual variance of log earnings from 1997 to 2019, where residual variance has been rescaled to hold it constant at the 1997 level. The bottom line, in blue, is the same simulation shown in the bottom blue line in Figure 10, showing the simulation that keeps the education-earnings gradient constant over time, while allowing the schooling distribution to change as it actually did.

The purple line shows the simulated impact of changing the earnings of earners with university education (15+ in Table 2) so that they have the same earnings relative to the mean as this group had in 1997. As seen in Table 2, this counterfactual reduces the earnings for university education earnings for all groups in 2019, (and for all other years), a reflection of the increasing returns to university education shown in Figure 6. As seen in the top panel of Figure 11, if those with university education had kept the same earnings relative to the mean as they had in 1997, the variance of log earnings would have been significantly lower in every year. In 2019 the variance of log earnings would have been 12% lower than it was if the premium to university education had not increased. A similar pattern is seen for the Gini coefficient in the lower panel. In 2019 the variance of log earnings would have been 8% lower than it was if the university premium had not increased.

The green line in Figure 11 shows the counterfactual in which we hold the relative earnings constant for those with grade 8 to grade 11 education. In this case we are hypothetically increasing earnings for this group to offset the decline in the relative earnings of those with incomplete secondary education. A striking result is that this counterfactual produces earnings inequality that in many years is similar to the inequality in the simulation that holds relative earnings constant for those with university. This implies that the *decrease* in the relative earnings of those with incomplete secondary education had a similar effect of increasing inequality as the *increase* in the relative earnings of those with university. Results for the Gini coefficient are similar to those for the variance of log earnings.

The red line in Figure 11 shows the simulated impact of changing the earnings of earners with grade 12 so that they have the same earnings relative to the mean as this group had in 1997. Whether this counterfactual involves an increase or decrease in relative earnings differs across years and population groups. As seen in Table 2, this counterfactual increases the earnings for African males and females with grade 12 education in 2019, but decreases the earnings for White males and females with grade 12 education. The bottom line is that the effect of these adjustments in relative earnings at grade 12 is to modestly decrease earnings inquality in every year, although the changes are small. This implies that the actual changes in relative earnings at grade 12 worked in the direction of slightly increasing inequality. The fact that changes in grade 12 have relatively small effects on inequality is consistent with our theoretical results, given that the changes in the grade 12 premium are relatively small, as seen in Table 2. When we change the earnings on only one race/gender group at a time (not shown), we find that the decline in relative earnings at grade 12 for Africans tended to be slightly equalizing, while the increase in relative earnings at grade 12 for Whites tended to be slightly disequalizing.

Taken together, Figures 10 and 11 provide some important insights into how the changing schooling distribution and changing earnings-education gradient affected changes in earnings inequality in the post-apartheid period. Looking at the question of whether reductions in schooling inequality tended to reduce earnings inequality, we see clear evidence from our counterfactual simulations that earnings inequality would have been lower in every year and would have fallen significantly over time if the relationship between schooling and earnings had remained the same as it was in 1997. This effect was offset, however, by the significant changes in the earnings-education gradient. Two key features characterize the change in this gradient over the period. First, there were significant increases in the returns to post-secondary education. As seen in Table 2, the earnings advantage of those with university education rose by 0.3 (African females) to 0.7 (African males) log points from 1997 to 2019. For African males with university education this meant that their mean earnings went from being 2.8 times the overall mean to being 5.7 times the overall mean. If we hold university graduates to the relative earnings of 1997, we see in Figure 11 that earnings inequality would have been lower in every year and would have fallen substantially over time.

The second important change in the earnings-education gradient was the decline in relative earnings of workers with incomplete secondary school. For example, we see in Table 2 that mean log earnings for African males with grade 10 education fell from being 0.06 log points above the overall mean to being 0.4 log points below the overall mean from 1997 to 2019. Exponentiating, this is a decline from being roughly at the mean in 1997 to being 30% below the mean in 2019. The simulations in Figure 11 show that if earnings relative to the mean for those with grade 8-11 education had remained at their 1997 levels, earnings inequality would have been lower in every year and would have fallen steadily over time. Interestingly, the magnitude of the inequality-increasing effect of the decline in relative earnings at grade 8-11 is almost as large as the magnitude of the inequality-increasing effect of the increase in relative earnings for university graduates.

An interesting question is why South Africa didn't experience the "paradox of progress" that Bourguignon et al. (2005) found in a number of Asia and Latin American countries, in which improvements in the distribution of education had a disequalizing effect on earnings inequality. They argued that shifting people up in the education distribution tends to be disequalizing, *ceteris paribus*, given the convex relationship between earnings and education. We find in the South African case, however, that improvements in the distribution of education, in and of themselves, would have contributed to a decline in earnings inequality.

Our theoretical results help explain why improvements in the distribution of education were equalizing in South Africa in the post-apartheid period. While it was the case that the population shifted up the education distribution along a highly convex earnings-education gradient, that movement alone is not enough to create the paradox of progress and lead to rising earnings inequality. As our theoretical results demonstrate, the critical question is whether the population is, on average, moving closer to or further away from the benchmark level of income associated with rising or falling inequality. For the variance of log earnings, this benchmark is the mean of log earnings. For the GE(0) measure it is mean earnings. For other measures of inequality it will be some other level of income.

What happened in South Africa was a significant reduction in the proportion of the population with low levels of education (0-7), accompanied by increases in the proportions with incomplete secondary, complete secondary, and, post-secondary. While there was some increase at the top of the education distribution, including university, most of the increase was concentrated in the grade 8-12 range, as shown in Figure 2. This means that most of the increase in education was toward levels of education near the mean of log earnings and mean earnings, with a smaller movement toward levels of education with earnings well above the mean. If there had been bigger increases in the proportions going beyond secondary, especially to university, the increases in education would have been more likely to have been disequalizing. In this sense, the fact that the distribution of education evolved in a way that tended to reduce earnings inequality is not entirely good news. It reflects the fact that there was a compression of education away from low levels toward the middle, which was good, but with less success in moving people into post-secondary education. The reason South Africa didn't experience the "paradox of progress" is, in a sense, because the progress in expanding education was somewhat limited, with only modest progress in moving people beyond grade 12.

Conclusion

This paper sheds new light on the theoretical and empirical relationships between schooling inequality, returns to schooling, and earnings inequality. Our theoretical results demonstrate that the impact of changes in returns to schooling on earnings inequality will depend on the levels of schooling at which those changes occur. This is important, since results from many countries in recent decades show that changes in returns to schooling have been far from uniform across the schooling distribution. Increases in relative earnings at low grades will tend to be inequality reducing, while increases in relative earnings at high grades will be inequality increasing. This implies a cutoff level of schooling that divides inequality-reducing from inequality-increasing increases in relative earnings. We show analytically that for the variance of log earnings this cutoff is the level of schooling associated with mean log earnings. For the Generalized Entropy (0) measure it is the level of schooling associated with mean earnings. While other measures do not always have simple analytical solutions, we show that the cutoff can be identified empirically for any given distribution of schooling and returns to schooling.

Our results also shed light on how changes in the distribution of education will affect earnings inequality. The basic result is the shifts in the distribution of education will be equalizing if they move people closer to the cutoff level of education for a given measure of inequality. A shift from grade 7 to grade 10 will tend to be equalizing, while a shift from grade 12 to university will tend to be disequalizing. Which changes in the distribution are equalizing or disequalizing will tend to change over time, however, and will generally be different for different racial and gender groups.

Our empirical analysis applies our theoretical results to post-apartheid South Africa, a country with one of the highest levels of earnings inequality in the world. The trend in earnings inequality since 1994 has been disappointing, with many measures showing an increase and other measures showing a relatively flat but persistently high level of earnings inequality over the period. We show that the inequality in schooling has declined significantly over the period, however, with steady shifts upward in the distribution of schooling. Our theoretical analysis helps explain why these steady reductions in schooling inequality have not led to reductions in earnings inequality.

Our counterfactual simulations show that changes in the schooling distribution would, in and of themselves, have led to significant reductions in earnings inequality. If the relationship between earnings and education had remained as it was in 1997, improvements in the distribution of schooling would have caused the variance of log earnings to have been 30% lower than its actual level in 2019, and the Gini coefficient to have been 15% lower. This equalizing effect of declining schooling inequality, however, was offset by changes in the relationship between earnings and education. Our results show that returns to schooling have increased at the top of the schooling distribution, while returns declined in the middle parts of the distribution. Our theoretical analysis allows us to see that both the *increase* in relative earnings at the top of the distribution and the *decrease* in relative earnings in the middle of the distribution worked to increase earnings inequality in South Africa. While declines in the relative earnings of those with 9 or 10 years of schooling would have been equalizing in 1995, the increase in overall schooling means that declines in earnings of those with 9 or 10 years of schooling are now disequalizing. We show that if earners with 8-11 years of schooling had the same earnings relative to the mean that they had in 1997, earnings inequality would have been lower in every year and would have declined significantly over time. The magnitude of the disequalizing effect of declining relative earnings for those with grade 8-11 education is similar to the magnitude of the disequalizing effect of increasing relative earnings for those with university education.

We have not attempted to explain why the earnings-education gradient changed in the way it did in the post-apartheid period. Increases in returns to post-secondary schooling have been widely observed in countries at all income levels in recent decades. It is not surprising that these have also been experienced in South Africa. Declining returns to schooling at the incomplete secondary level are less widely studied. These could be related to increasing numbers of students reaching this schooling level, leading to downward pressure on earnings when they enter the labour market. But we show that the fraction of earners with incomplete secondary education has not increased dramatically over time. The same economic pressures that have increased returns to post-secondary education, such as increased technological change and pressures of global trade, may help explain the decline in earnings for those with incomplete secondary. Whatever the causes of these changes, our results show that they have played an important role in driving earnings inequality in South Africa in the post-apartheid period. The changes have for the most part worked in the direction of increasing earnings inequality, with the effects being large enough to offset what would otherwise have been steady declines in earnings inequality due to declining inequality in schooling.

References

Almeida dos Reis, Jose Guilherme and Paes de Barros, Ricardo, 1991. "Wage inequality and the distribution of education: A study of the evolution of regional differences in inequality in metropolitan Brazil," *Journal of Development Economics*, 36(1): 117-143. https://doi.org/10.1016/0304-3878(91)90007-I

Blinder, Alan S. 1973. "Wage discrimination: Reduced form and structural estimates." *Journal of Human Resources*. 8: 436–455.

Bourguignon, Francois, Francisco Ferreira, and Nora Lustig. 2005. *The Microeconomics of Income Distribution Dynamics in East Asia and Latin America*. Washington DC: Oxford University Press and The World Bank.

Bourguignon, Francois, Francisco Ferreira, and Phillippe Leite. 2008. Beyond Oaxaca–Blinder: Accounting for differences in household income distributions. Journal of Economic Inequality 6: 117–148. https://doi.org/10.1007/s10888-007-9063-y

Branson, Nicola, and David Lam. 2021. "The Economics of Education in South Africa," in Arkebe Oqubay, Fiona Tregenna, and Imraan Valodia, editors, *The Oxford Handbook of the South African Economy*, Oxford University Press, November 2021. https://doi.org/10.1093/oxfordhb/9780192894199.013.31

Branson, N., Culligan, S., Ingle, K. 2020. Developing Siyaphambili: A Stronger South African Nation Website. Moving towards a unified goal to combat inequality and unemployment. Cape Town: SALDRU, UCT. (SALDRU Report 20/01)

Foster, James, and Efe Ok. 1999. "Lorenz Dominance and the Variance of Logarithms," *Econometrica* 67(4): 901-908.

Juhn, Chinhui, Kevin M. Murphy and Brooks Pierce. 1993. "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy*, 101(3), pp. 410-42.

Hoffman, Rodolfo. 2001. "Effect of the rise of a person's income on inequality," *Brazilian Review of Econometrics* 21(2): 237-262.

Knight, J. B. and Sabot, R. H. 1983. "Educational Expansion and the Kuznets Effect". *American Economic Review* 73(5): 1132–36.

Kerr, Andrew. 2020. 'Earnings in the South African Revenue Service IRP5 data'. WIDER Working Paper 2020/62. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2020/819-1

Kerr, Andrew. 2025. "Earnings and Earnings Inequality in South Africa: Evidence from Household Survey and Administrative Tax Microdata from 1993 to 2020," *Review of Income and Wealth*, 71(1): e12695. https://doi.org/10.1111/roiw.12695

Kerr, Andrew, David Lam, and Martin Wittenberg. 2019. 'Post-Apartheid Labour Market Series [dataset]'. Version 3.3. Cape Town: DataFirst (producer and distributor).

Kerr, Andrew, and Martin Wittenberg. 2019a. "Earnings and Employment Microdata in South Africa". WIDER Working Paper 2019/47. Helsinki: UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2019/681-4

Kerr, Andrew, and Martin Wittenberg. 2019b. 'A Guide to PALMS version 3.3'. Available at: https://www.datafirst.uct.ac.za/dataportal/index.php/catalog/434/download/10286 (accessed 13 April 2021).

Lam, David. 2020. “Why Has Income Inequality Increased while Education Inequality Has Decreased in Many Developing Countries?” In *The Political Economy of Inequality: U.S. and Global Dimensions*, Sisay Asefa and Wei-Chiao Huang, eds. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 2020, pp. 79-116. https://doi.org/10.17848/9780880996730.Ch5

Lam, David, and Deborah Levison. 1991. “Declining Inequality in Schooling in Brazil and Its Effect on Inequality in Earnings,” *Journal of Development Economics*, 37: 199-225. https://doi.org/10.1016/0304-3878(91)90088-D

Lambert, Peter, and Giuseppe Lanza. 2006. “The effect on inequality of changing one or two incomes.” *Journal of Economic Inequality* 4: 253–277. https://doi.org/10.1007/s10888-006-9020-1

Oaxaca, Ronald. 1973. “Male–female wage differentials in urban labor markets.” *International Economic Review*. 14: 673–709.

Roope, Laurence. 2019. “Characterizing inequality benchmark incomes,” *Economic Theory Bulletin* 7: 131–145. https://doi.org/10.1007/s40505-018-0148-5

Roope, Laurence. 2021. “First estimates of inequality benchmark incomes for a range of countries,” PLoS ONE 16(3): e0248178. https://doi.org/10.1371/journal.pone.0248178

Statistics South Africa. 2008–19. ‘General Household Surveys’. [datasets]. Pretoria: Statistics SA [producer], 2019. Cape Town: DataFirst [distributor].

Wittenberg, Martin. 2017. Wages and Wage Inequality in South Africa 1994–2011: Part 2 – Inequality Measurement and Trends. *South African Journal of Economics*, 85: 298-318. https://doi.org/10.1111/saje.12147

Table 1: Sample means for selected variables by year, age 25-59 with positive earnings, 1994 – 2019

Year	Highest grade	Age	Female	African	Coloured	Indian	White	Log Earnings	Real Earnings	Sample size
1994	8.3	38.6	0.37	0.71	0.09	0.02	0.17	8.31	7,283	11,885
1995	8.6	38.6	0.39	0.67	0.11	0.03	0.18	8.48	9,082	20,596
1997	8.8	38.7	0.38	0.69	0.11	0.03	0.17	8.49	8,707	15,183
1998	8.4	38.7	0.39	0.75	0.09	0.03	0.13	8.42	8,863	8,458
1999	8.4	38.8	0.41	0.74	0.10	0.03	0.13	8.34	9,195	11,138
2000	8.7	39.1	0.44	0.70	0.11	0.04	0.15	8.30	7,945	18,850
2001	8.8	39.3	0.43	0.72	0.09	0.03	0.15	8.29	7,646	16,040
2002	8.8	39.2	0.44	0.74	0.10	0.04	0.13	8.29	7,864	14,192
2003	9.0	39.2	0.43	0.75	0.10	0.03	0.12	8.29	7,631	13,528
2004	9.1	39.4	0.42	0.76	0.09	0.03	0.12	8.36	7,909	14,630
2005	9.2	39.2	0.43	0.76	0.09	0.03	0.12	8.36	8,378	15,331
2006	9.4	39.3	0.43	0.75	0.09	0.03	0.12	8.44	8,738	15,822
2007	9.5	39.2	0.43	0.76	0.10	0.03	0.11	8.51	9,554	15,718
2008	9.6	39.0	0.41	0.75	0.10	0.03	0.12	8.49	9,694	12,801
2009	10.1	38.9	0.42	0.77	0.09	0.03	0.12	8.58	11,235	12,839
2010	10.1	38.8	0.43	0.77	0.09	0.03	0.11	8.50	10,274	12,351
2011	10.2	39.0	0.43	0.78	0.09	0.03	0.10	8.55	10,738	12,234
2012	10.3	38.9	0.43	0.80	0.09	0.02	0.09	8.57	11,835	12,117
2013	10.3	39.1	0.43	0.80	0.09	0.02	0.09	8.57	10,876	12,227
2014	10.5	38.7	0.43	0.82	0.08	0.02	0.08	8.62	12,019	11,935
2015	10.5	38.5	0.42	0.81	0.09	0.02	0.08	8.61	11,603	10,635
2016	10.6	38.7	0.42	0.82	0.09	0.02	0.08	8.60	11,035	10,165
2017	10.6	38.9	0.42	0.83	0.08	0.02	0.07	8.56	10,289	10,234
2018	10.7	39.2	0.42	0.83	0.09	0.01	0.07	8.58	10,768	10,110
2019	10.7	39.2	0.44	0.83	0.08	0.02	0.07	8.53	11,609	7,736

Notes: Sample includes adults aged 25-59 with positive earnings. Mean values are calculated using sample weights that include an adjustment for respondents who provided earnings in brackets. Real earnings are in 2019 rands.

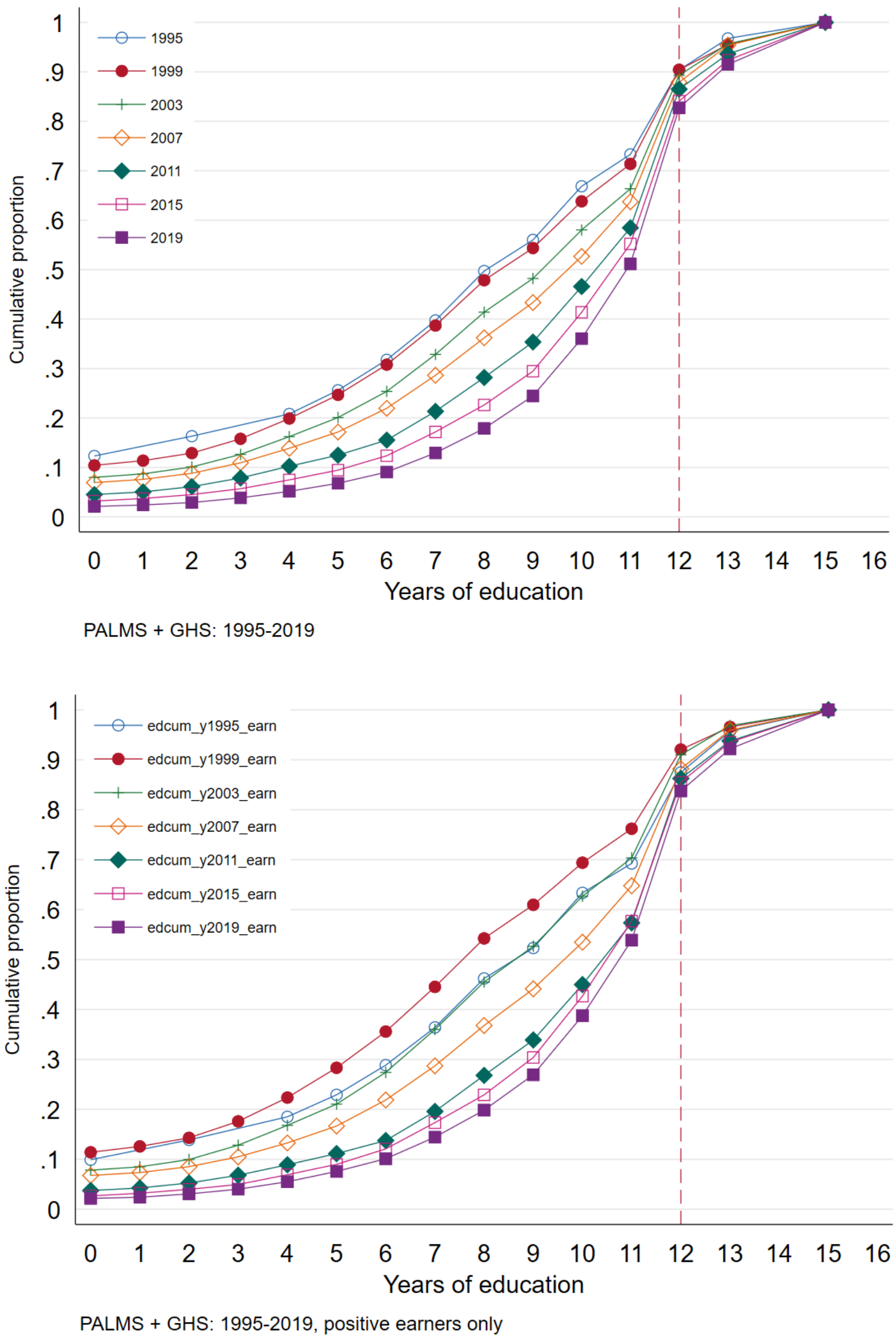


Figure 1. Cumulative Distribution Functions for years of education, South Africa, population aged 25-59 (top panel) and positive earners aged 25-59 (bottom panel).

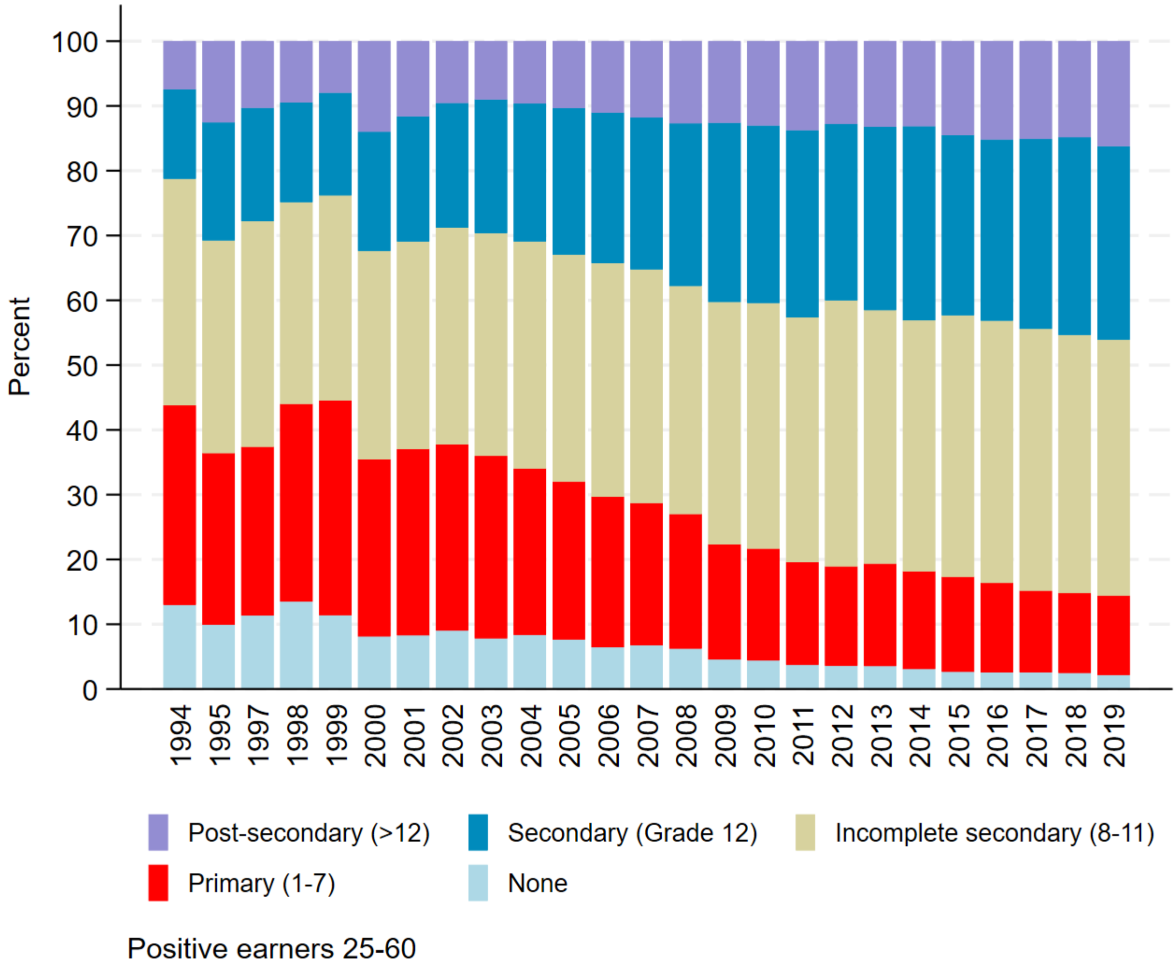


Figure 2. Distribution of years of education, positive earners aged 25-59, 1994-2019.

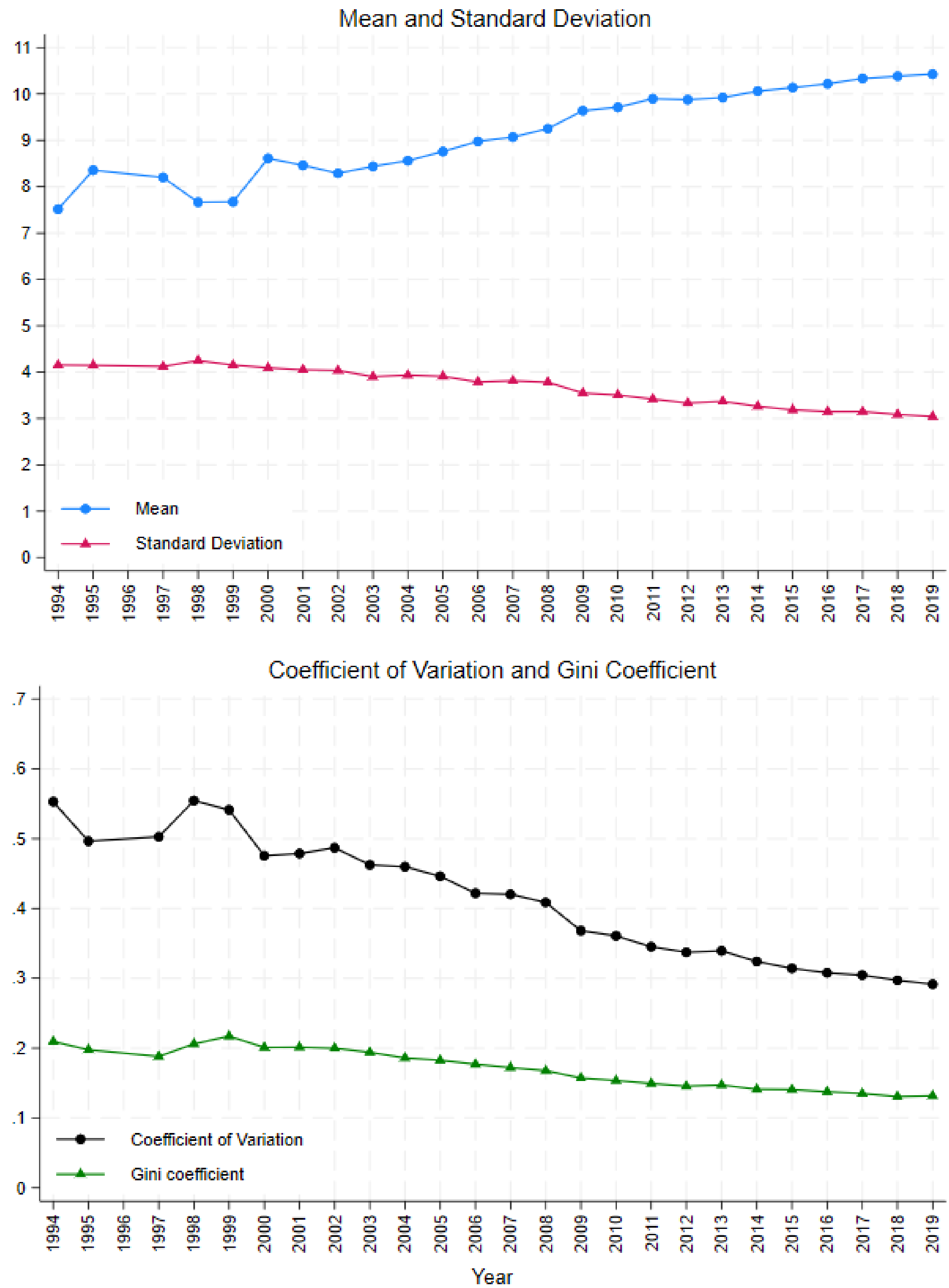


Figure 3. Mean, standard deviation, coefficient of variation, and Gini coefficient for years of completed schooling, income earners aged 25-59, 1994-2019.

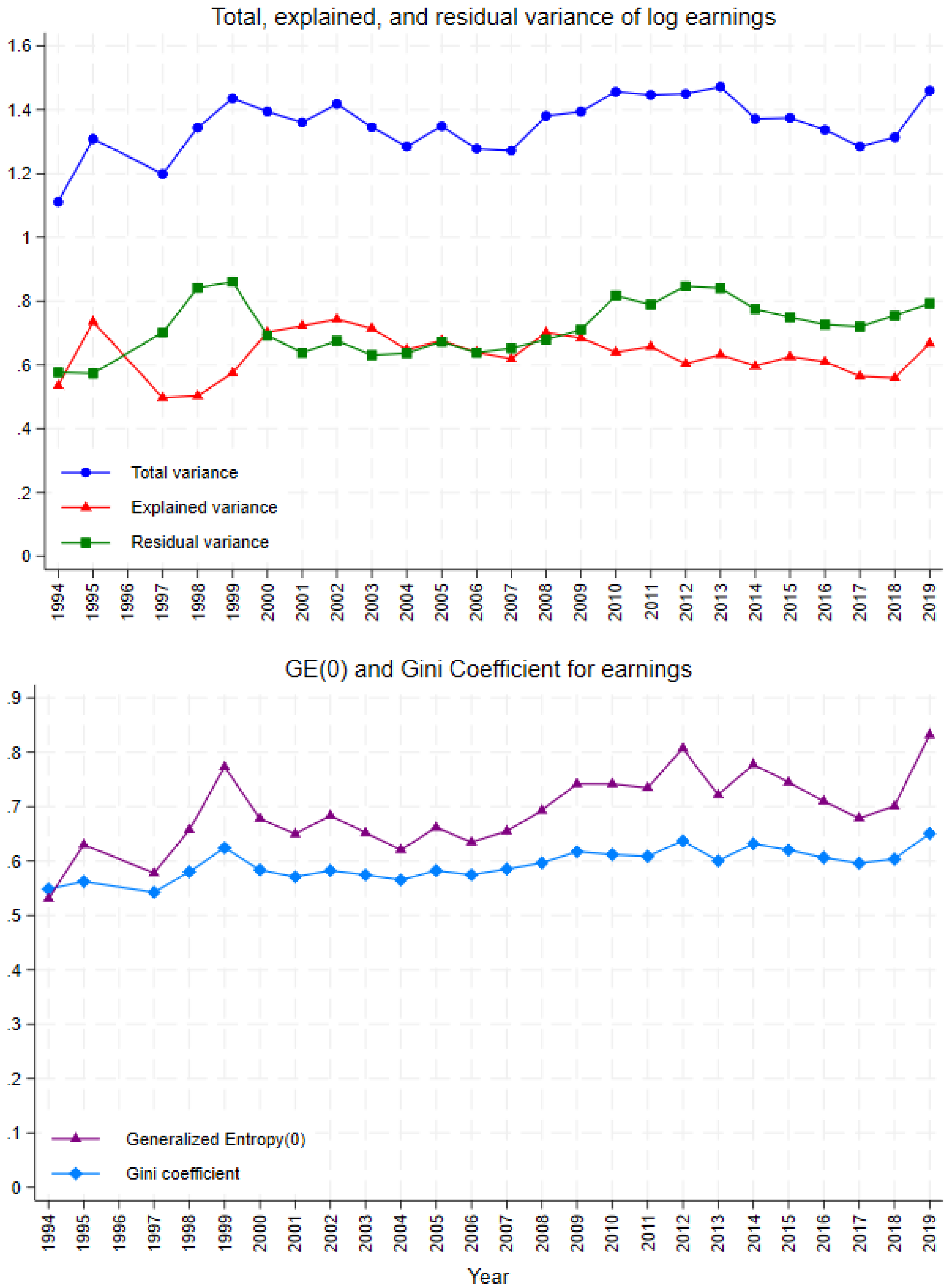


Figure 4. Measures of earnings inequality, earners aged 25-59, 1994-2019.
Note: Explained variance and residual variance of log earnings are based on regressions that include dummies for each year of schooling, age, age squared, dummies for racial groups, and a dummy for female.

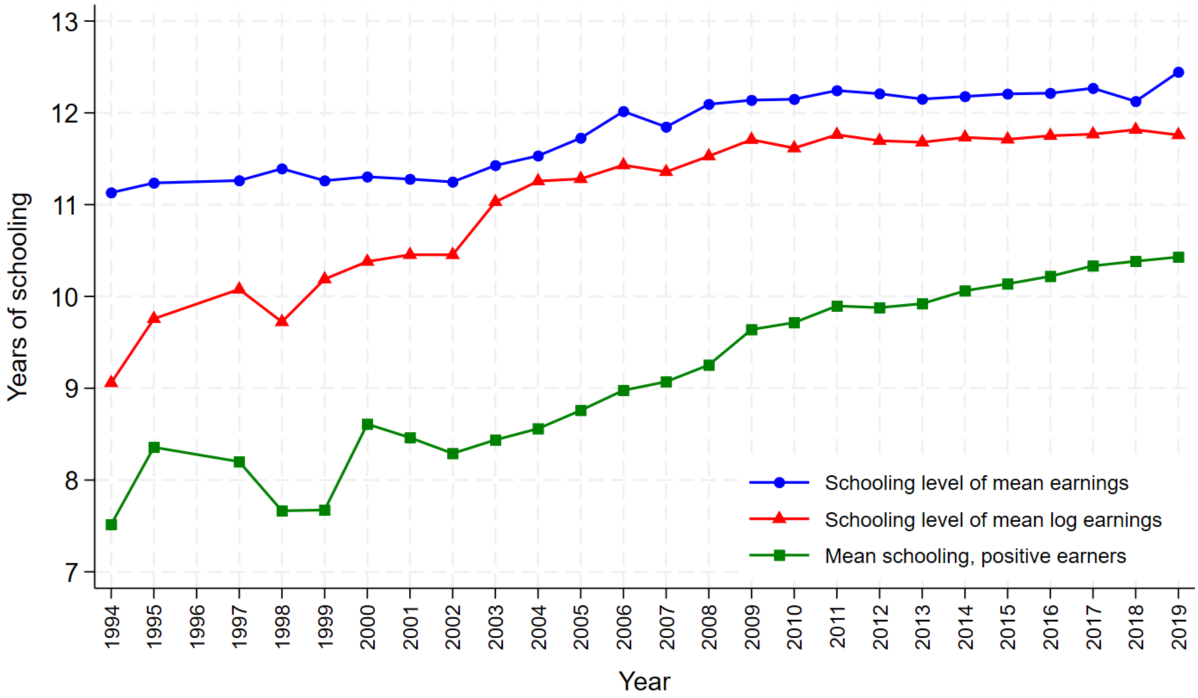


Figure 5. Mean years of completed schooling and schooling level of mean earnings and mean log earnings, South African earners aged 30-39, 1994-2019.

Note: All data are for positive earners, evaluated for 30-39 year-olds.

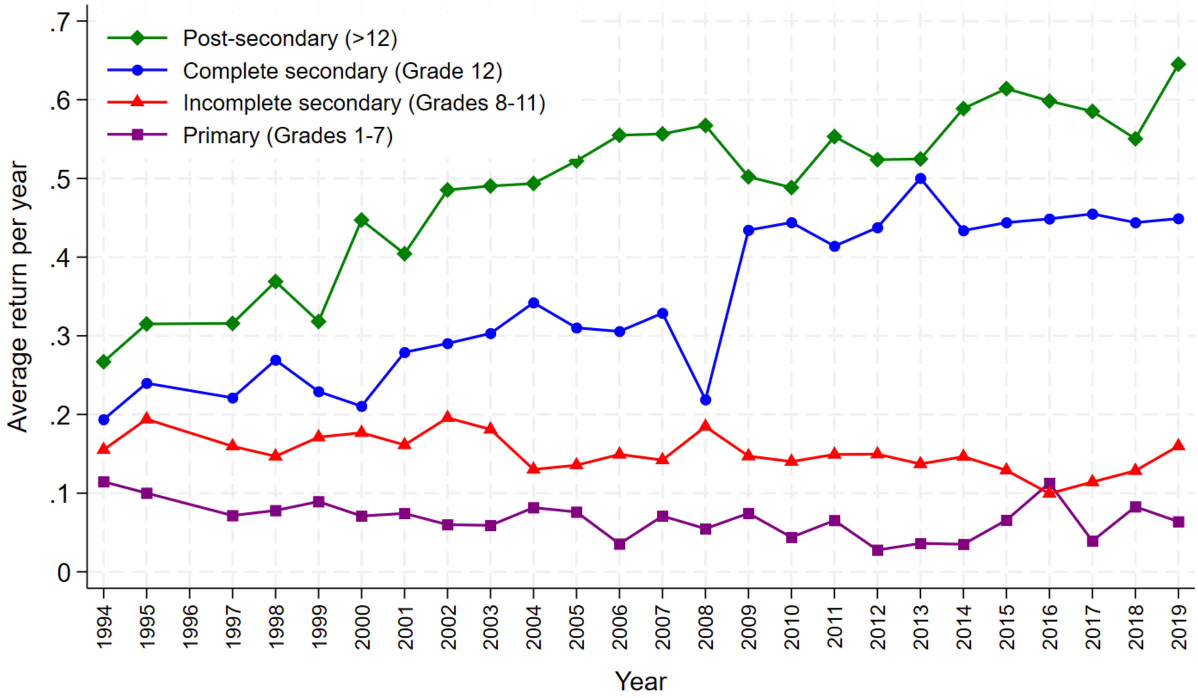


Figure 6. Average returns to schooling per year in schooling groups, South African earners aged 25-59, 1994-2019.

Note: Each line shows the weighted average of marginal returns to each year of schooling in the schooling group, based on log earnings regression on single years of schooling.

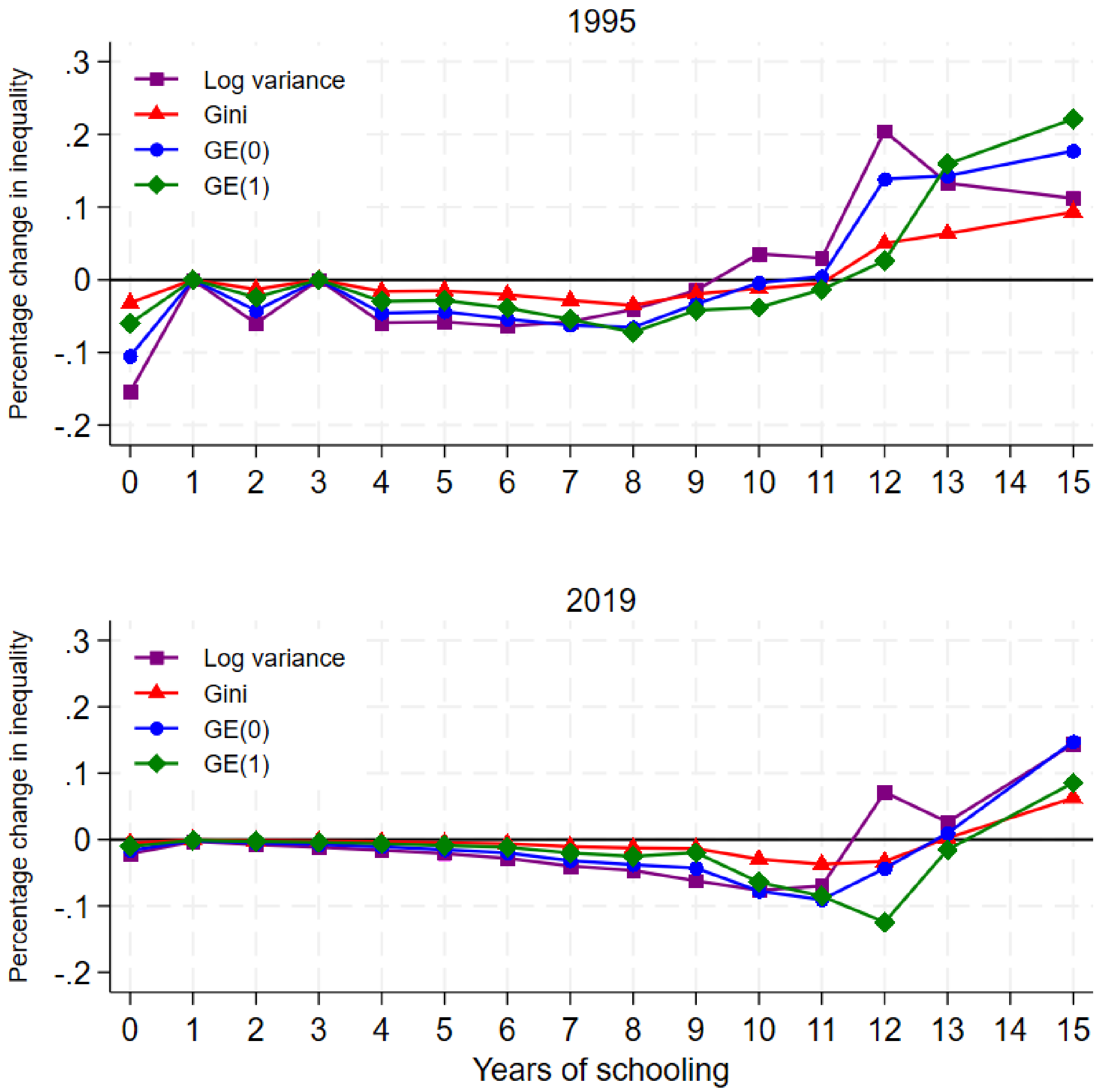


Figure 7.
Simulated impact of a 1% increase in earnings at each schooling level on measures of earnings inequality, 1995 and 2019

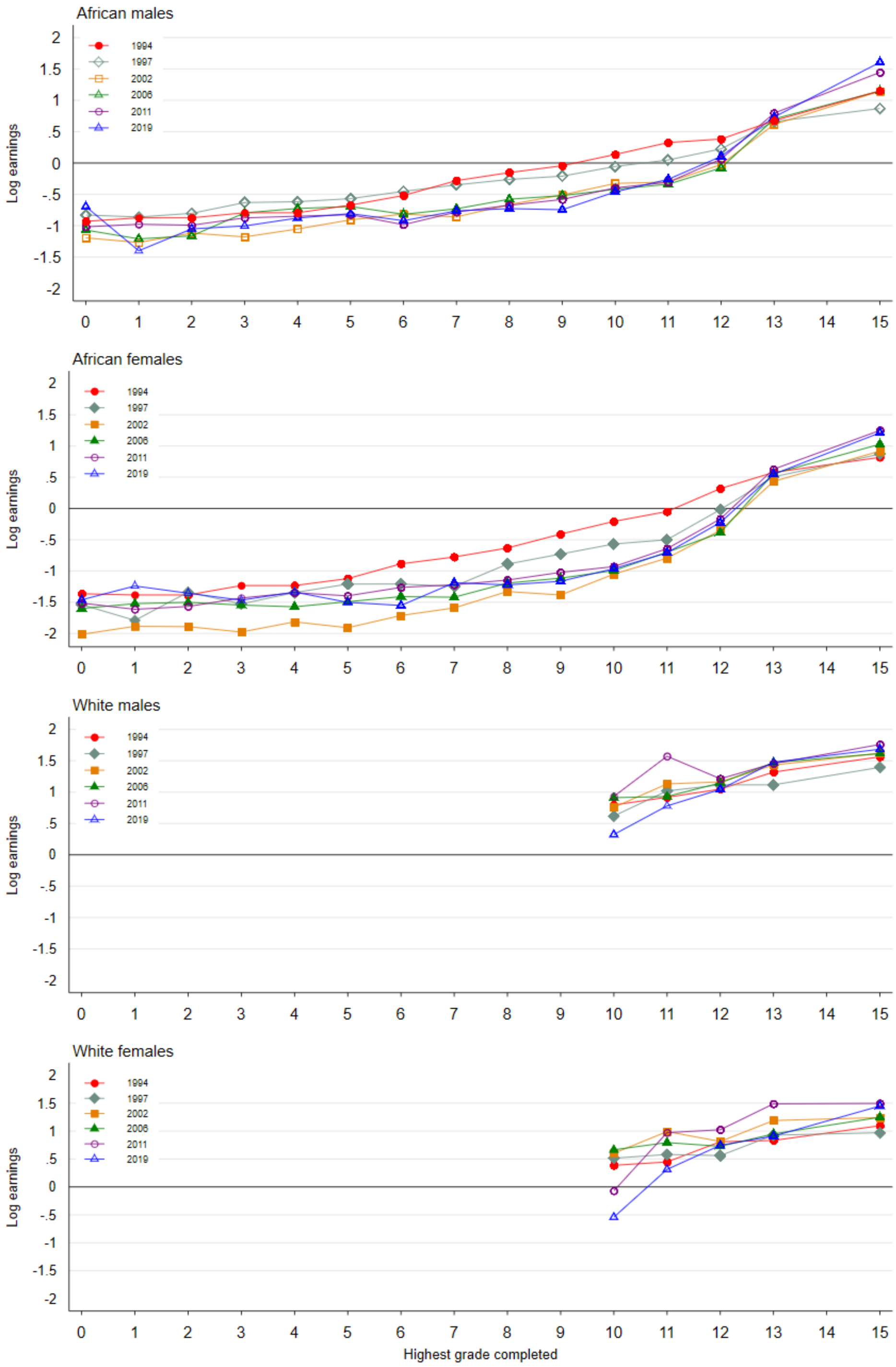


Figure 8. Predicted log earnings at each schooling level minus overall mean log earnings, 1994-2019

Note: Earnings at each grade predicted using separate regressions by year, race, and gender, including age and age squared, using positive earners aged 25-59. Mean log earnings in each year is calculated for full population of earners aged 25-59. Grade 13 indicates post-secondary short of university. Grade 15 indicates university degree or higher.

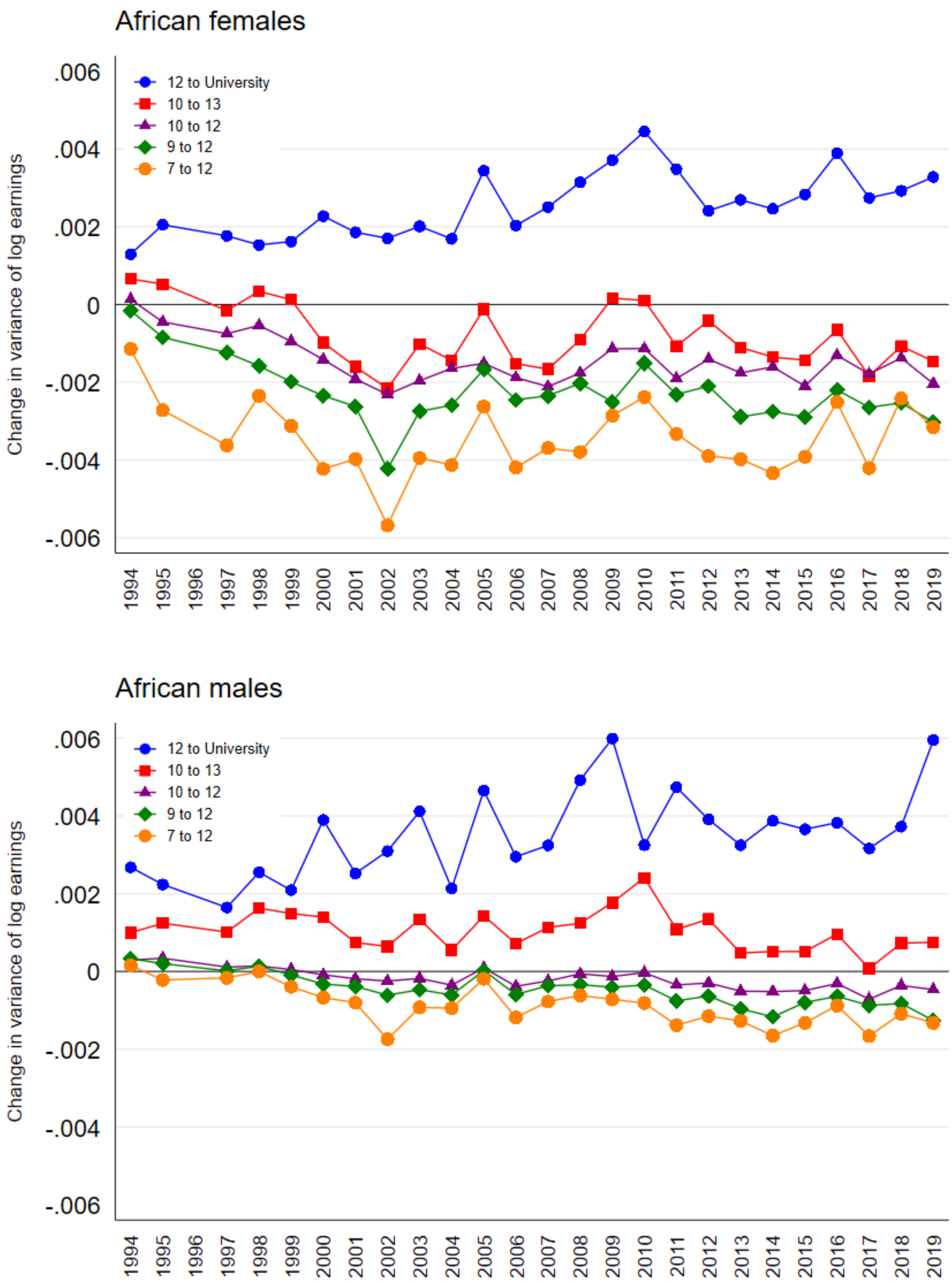


Figure 9

Increase in variance of log earnings from shifting 1% of population from lower grade to higher grade, African females and males, 1994-2019

Note: Earnings at each grade predicted using separate regressions by race and gender, with weights normalized to 1997 race-gender distribution

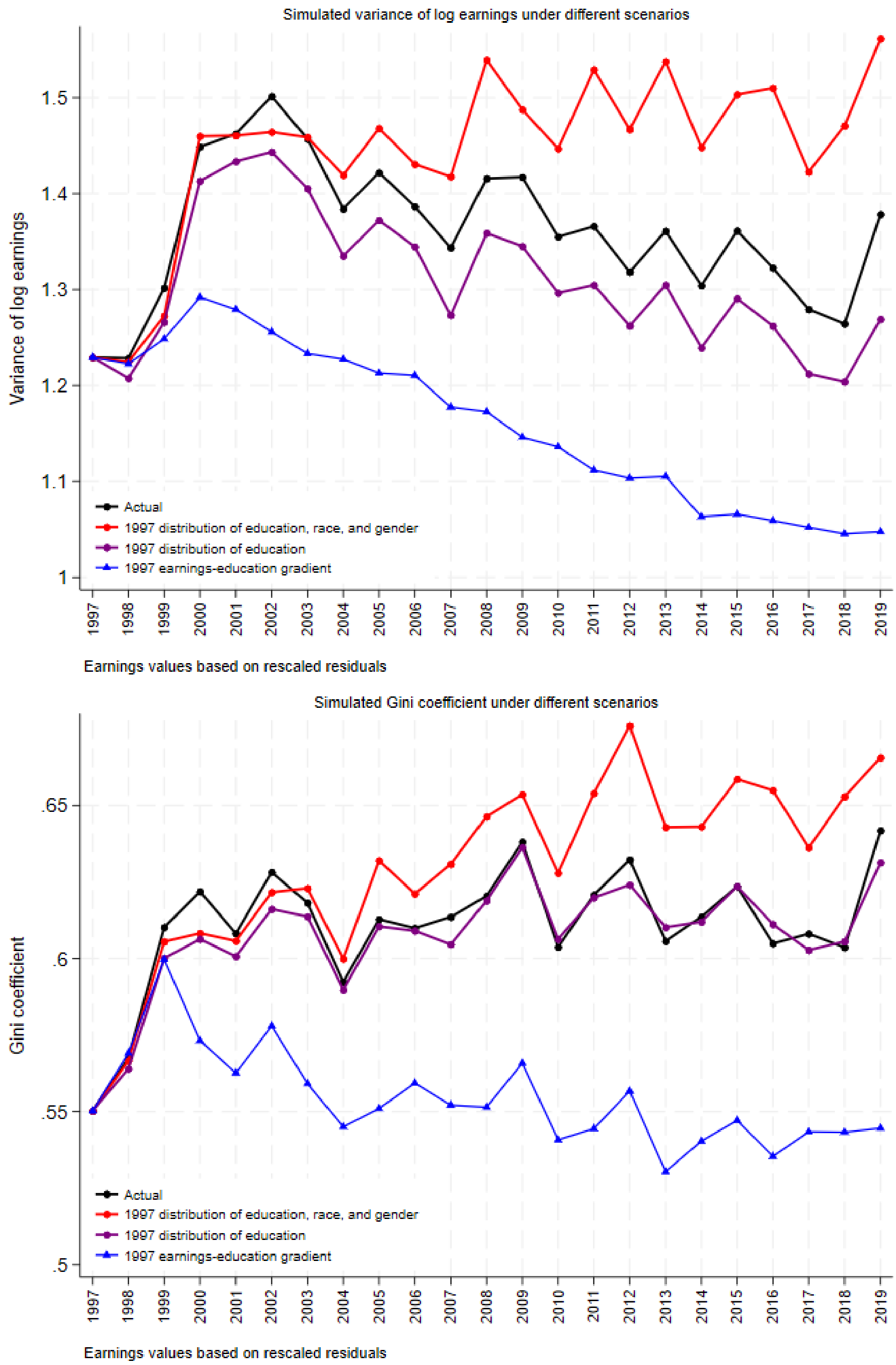
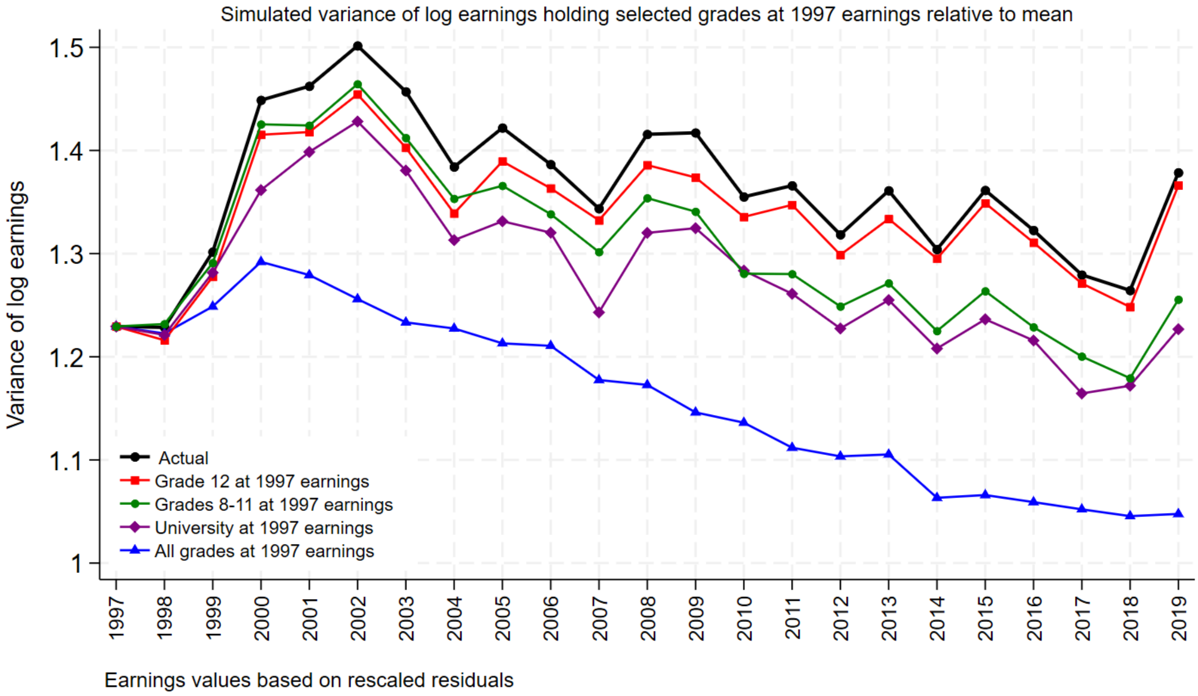


Figure 10.
Actual and counterfactual earnings inequality under alternative scenarios, positive earners aged 25-59, 1997-2019



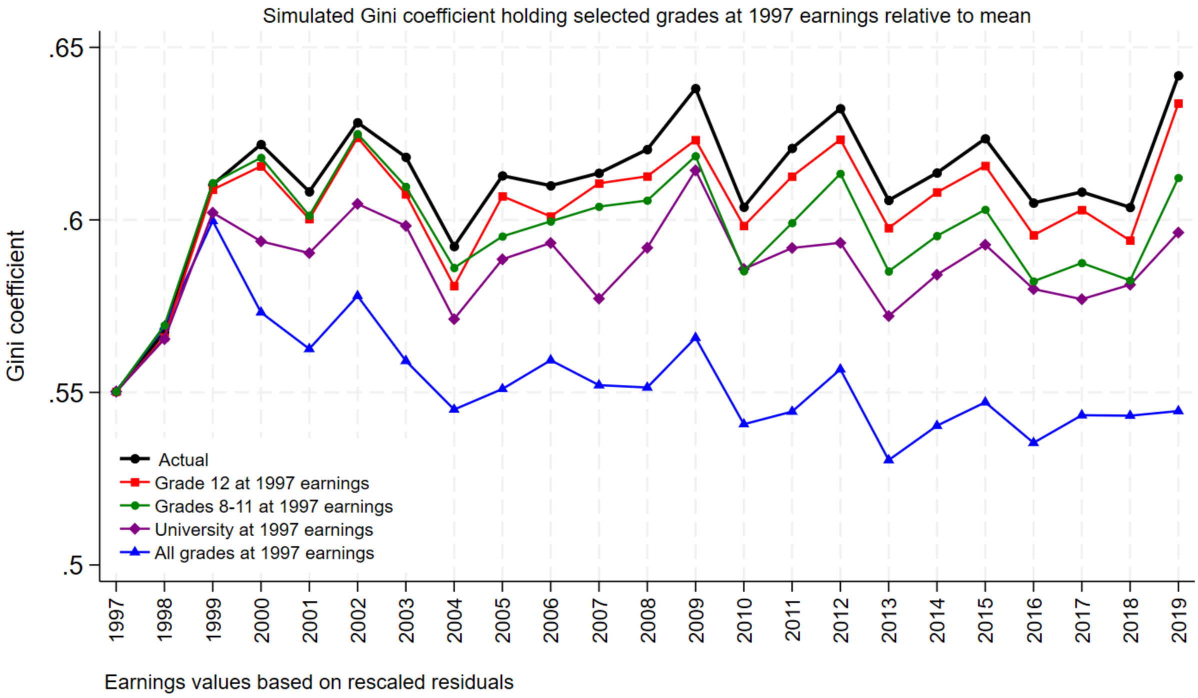


Figure 11.
Simulated earnings inequality holding selected grades at 1997 earnings relative to mean

Table 2. Change in log earnings relative to the overall mean by grade, race, and gender, 1997 to 2019

Group	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	Grade 15+	Grades 8-11	Grades 12-15+
Log earnings relative to mean, 1997								
African males	-0.10	-0.10	0.06	0.13	0.29	1.04	0.00	0.48
African females	-0.71	-0.62	-0.44	-0.42	0.04	0.99	-0.55	0.36
White males	--	--	0.73	1.15	1.24	1.54	0.91	1.31
White females	--	--	0.61	0.63	0.62	1.04	0.59	0.78
Log earnings relative to mean, 2018								
African males	-0.59	-0.64	-0.35	-0.18	0.20	1.74	-0.37	0.60
African females	-1.10	-1.06	-0.89	-0.62	-0.16	1.27	-0.82	0.35
White males	--	--	0.71	1.05	1.42	2.05	0.89	1.79
White females	--	--	-0.25	0.65	1.02	1.73	0.37	1.39
Relative log earnings 2018 minus 1997								
African males	-0.49	-0.55	-0.42	-0.31	-0.10	0.70	-0.37	0.11
African females	-0.39	-0.44	-0.45	-0.21	-0.20	0.29	-0.27	-0.01
White males	--	--	-0.02	-0.10	0.18	0.51	-0.02	0.47
White females	--	--	-0.87	0.01	0.41	0.69	-0.23	0.60

Note: "Grade 15+" indicates university diploma and beyond. "Grades 12-15+" indicate Grade 12 (end of secondary) up to and including university diploma and beyond. Grade 8 and 9 are omitted for white males and females due to small cell sizes.