A Paradox of Progress:

Rising Education and Unequal Labour Market Returns in Post-Apartheid South Africa

COVID-Generation Working Paper

By Timothy Köhler

December 2024

Stellenbosch University Iyunivesithi Universiteit

forward together
sonke siya phambili
saam vorentoe

RESEP Research on Socio-Economic Policy

A Paradox of Progress: Rising Education and Unequal Labour Market Returns in Post-Apartheid South Africa

Timothy Köhler[1]

December 2024

ABSTRACT

This paper analyses the evolution of the labour market returns to education and their drivers in post-Apartheid South Africa using over 20 years of harmonized household survey microdata from 2001 to 2023 including credible earnings data not available in the public domain. I document a significant increase in educational attainment, driven by the expansion of completed secondary and tertiary education, which resulted in a 40 percent reduction in educational attainment inequality. Despite this, the mean return to an additional year of education increased, suggesting that demand for higher-educated workers has outpaced supply. Increases in both educational attainment and returns drove wages upwards. While the former is dominant, the latter has grown in importance over time. Beyond the mean, the return to tertiary education has tripled, resulting in an increasingly convex returns structure. This return became particularly stronger for lower-wage workers and primarily explains the group's significant growth in real wages. The consequence was a reduction in overall wage inequality which, nevertheless, remains high. Finally, differential returns rather than differences in educational attainment have grown in importance in explaining inter-race wage inequality, which is suggestive of increasing discrimination but likely also reflects a growing importance of education quality differentials.

Keywords: Education; Returns; Employment; Wages; Inequality; South Africa

JEL codes: I24; I26; J31; O15

Acknowledgements: I am incredibly grateful to Dr. Gabrielle Wills for constructive feedback, thoughtful suggestions, and valuable guidance on an earlier version of this work. I also owe thanks to participants of the Research on Socio-Economic Policy's (RESEP) 2024 Quantitative Education Research conference. I gratefully acknowledge funding from Allan and Gill Gray Philanthropies as part of the Covid-Generation project. Any remaining errors or omissions are my own. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of Allan & Gill Gray Philanthropies.

---

[1]Development Policy Research Unit, School of Economics, University of Cape Town, South Africa. tim.kohler@uct.ac.za.

1 INTRODUCTION

Education, of course, has many individual and social benefits. Among others, these include poverty reduction, the promotion of rights, fertility decisions, job search efficiency, technological innovation, social capital, health status, and social cohesion (Becker, 1964; Colclough et al., 2010; Fasih et al., 2012; Montenegro and Patrinos, 2014; Deming, 2022). The relationship between an individual's education and their labour market earnings, loosely referred to as the private return to education, is one of the oldest, most often-studied topics yielding one of the most robust findings in social science (Colclough et al., 2010; Montenegro and Patrinos, 2014; Webber, 2014; Deming, 2022). In all countries of the world, the more education a worker has, the higher their earnings on average (Patrinos and Psacharopoulos, 2020). Since the development of human capital theory (Mincer, 1958; 1962; 1974); Becker, 1964; Becker and Chiswick, 1966), thousands of estimates of the private returns to education from a wide range of countries have been published. All highlight the importance of investing in education for development (Fasih et al., 2012).

The average returns to education do not, of course, apply to every individual, as considerable variation exists both within and across countries as well as over time. Recent international trends, driven by a combination of supply- and demand-side factors, have led to an increase in the returns to tertiary education, resulting in a convex earnings-education relationship (Psacharopoulos and Patrinos, 2018; Patrinos, 2019). This marks a clear departure from earlier periods when the relationship was concave, with higher returns typically associated with primary education. Concurrently, as predicted by human capital theory, educational attainment has risen substantially (Patrinos, 2024). The implications of rising educational attainment and increasing returns to higher education for labour market inequality are *ex ante* unclear, as they depend on the relative pace of these changes as well as the extent to which the labour market values different levels of education across different parts of the wage distribution. As a result, education can either result in greater equality by providing wider access to higher-wage jobs, or conversely, exacerbate existing inequalities if the returns to higher education disproportionately benefit those already positioned at the top of the wage distribution.

In this paper, I examine these dynamics within the context of post-Apartheid South Africa, a country that has witnessed a significant rise in educational attainment alongside persistent and extreme labour market inequalities. Following the country's democratisation in 1994, education was prioritised as an area for expansion and reform, with the implementation of free public schooling resulting in some of the highest enrolment rates among countries of a similar level of development (Bhorat et al., 2021). Concurrently, there was optimism that fairer labour market policies and practices would help mitigate inequalities in earnings. While enrolment rates have nearly reached parity and racial gaps have narrowed, extreme labour market inequalities persist. These inequalities stem from both a large share of the population lacking access to labour market incomes (unemployment) combined with a very unequal distribution of these incomes among the employed (Finn et al., 2016; Wittenberg, 2017; Bhorat et al., 2020; Diaz

Pabon et al., 2021; Kerr and Wittenberg, 2021; Leibbrandt et al., 2012, 2020; Bhorat et al., 2022; Leibbrandt and Diaz Pabon, 2022; Köhler and Bhorat, 2023). To effectively address these inequalities, it is essential to understand how the relationship between education and labour market outcomes has evolved over time.

I make use of over 20 years of harmonized household survey microdata to analyse the evolution of the labour market returns to education and their drivers in South Africa from 2001 to 2023. By using the longest uninterrupted series of reliable earnings data not available in the public domain, I shed new light on the complex interplay between rising educational attainment and a varying returns structure with greater precision than was previously possible. After providing a descriptive analysis of trends in educational attainment, employment, and wage variation across the education distribution, I estimate the private returns to education using conventional and selection-adjusted Mincerian earnings functions. I explore heterogeneity in returns over time, across South Africa's four racial population groups, varying levels of education, and across the earnings distribution using Recentered Influence Function (RIF) regressions. Further, using Oaxaca-Blinder (OB) and RIF decomposition, I decompose wage variation both (i) over time, to assess the individual contributions of changes in education distribution versus changes in returns to education, and (ii) across racial groups, to examine the contributions of differences in productivity characteristics (namely education) versus differential returns to education, and how these have evolved over time.

I document a substantial increase in educational attainment, driven by an expansion of completed secondary and tertiary education and a contraction of primary education, which resulted in a 40 percent reduction in educational attainment inequality. Despite this, I estimate an increase in the average return to education, indicating that the demand for these workers has outpaced supply. On average, increases in both educational attainment and returns drove real wages upwards. While the former was dominant, the latter has doubled in importance over time. This pattern holds for most racial groups apart from White individuals, which is likely explained by their already relatively high levels of education. Beyond the mean, I show that the education-earnings has become increasingly convex, reflecting returns to tertiary education which have tripled in size over time. This pattern is evident for all racial groups, but to varying degrees. Across the wage distribution, the rates of return, particularly for tertiary education, became considerably stronger for lower-wage workers. These higher returns primarily account for the change in wages experienced by these workers, which more than doubled in real terms. Consequently, overall wage inequality has decreased – the magnitude of which depends on the measure – but nevertheless remains high. Considering inter-race wage inequality, I estimate that while most is explained by differences in educational attainment for most groups, the contribution of this component has fallen from 67 – 73 percent in 2001 to 32 – 55 percent in 2023, likely related to the convergence of educational attainment across groups over time. Conversely, differences in returns have become increasingly important for all groups, increasing from explaining up to one-third (27 – 33 percent) in 2001 to 45 – 68 percent in 2023. While this is suggestive of increasing discrimination, it likely at least partially reflects a growing importance of differences in education quality.

The remainder of the paper is structured as follows. Section 2 reviews the international and South African literatures on the labour market returns to education. Thereafter, Sections 3 and 4 describe the data and methodologies. In Section 5 I present a range of descriptive statistics, while Section 5 presents the formal modelling and decomposition results. Following a discussion of the results and their implications in Section 6, Section 7 concludes.

2 THE RETURNS TO EDUCATION: A REVIEW OF THE INTERNATIONAL AND SOUTH AFRICAN LITERATURES

While the relationship between education and earnings has its roots in the writings of classical economists, formal modelling did not begin until the development of human capital theory starting with Mincer (1958; 1962), Becker (1964), Becker and Chiswick (1966), and culminating in Mincer's (1974) seminal work. Mincer (1974) provided a great service in estimating these returns through what is colloquially known as the 'Mincerian earnings function', a parsimonious equation which linearly relates earnings to levels of education and experience using the entire age stream of earnings as an input. By doing so, it provides an approximation of the private rate of return earned on the opportunity cost of the worker's time out of the labour market to attend school for an additional year.[2] Because of its convenience, this method has dominated how these returns have been and continue to be estimated today, and more broadly, it serves as the most widely used estimation model in economics (Patrinos and Psacharopoulos, 2020; Patrinos, 2024; Psacharopoulos, 2024). Using it, thousands of empirical estimates of the returns to education have been published over the past half-century for a large number of countries across the development spectrum, with most behaving as predicted by Mincer's (1974) original work. By additionally leading to numerous works which examine a range of issues such as discrimination, income distribution, and the demand for education, the model has clearly advanced the fields of labour and educational economics and is now also used in other fields including sociology and anthropology (Psacharopoulos and Patrinos, 2018; Patrinos, 2024).

Using data from hundreds of harmonised household surveys covering most countries and a consistent model specification, the average private rate of return globally has been

[2] Specifically, Mincerian estimates focus on the benefits and costs to the individual, and hence, are referred to as *private* rates of return. They capture the increase in an individual's earnings due to additional education, accounting for personal costs like tuition and foregone earnings during schooling. They differ from *social* rates of return which capture the broader social benefits of education relative to the costs borne by both individuals and society as a whole, such as government subsidies and public investments in education infrastructure.

estimated to be around 10 percent per additional year of education (Montenegro and Patrinos, 2014). This estimate, and Mincerian estimates in general, have not been without criticism. These primarily relate to two potential sources of bias: upward bias due to omitted variables, and attenuation bias due to errors in the measurement of schooling. The former, as a form of selection bias, suggests that variation in earnings by level of education may not be due to differences in education levels, but instead due to, or at least influenced by, other characteristic differences between lower and higher-educated individuals, such as 'inherent ability'. Researchers often incorporate additional explanatory variables into the model to address this issue. However, doing so, particularly for variables that are arguably endogenous or do not stem from human capital theory, has been shown to introduce additional sources of bias and, crucially, prevent the schooling coefficient from being interpreted as a rate of return (Barouni and Broecke, 2014; Montenegro and Patrinos, 2014; Psacharopoulos and Patrinos, 2018; Psacharopoulos, 2024). Becker (1964) noted this, but it is often ignored. In any case, evidence suggests that the magnitude of omitted variable bias is small, if not negative (Griliches, 1970; Schultz, 1988; Ashenfelter and Krueger, 1994). Moreover, this bias tends to be offset by the attenuation bias caused by individuals misreporting their schooling (Griliches, 1977; Card, 1999; Hertz, 2003). To address both, researchers have alternatively used quasi-experimental research designs to identify causal effects. These effect estimates are almost always positive and, importantly, tend to be of a very similar magnitude to 'naïve' Mincerian estimates (Webber, 2014; Gunderson and Oreopolous, 2020; Patrinos and Psacharopoulos, 2020; Deming, 2022).[3]

The 10 percent estimate, as mentioned above, represents the average return; however, evidence of substantial heterogeneity in returns exists, both across and within countries as well as over time. In the lead up to the 21st century, several stylised facts emerged. In most countries, the earnings-education relationship was concave, reflecting higher returns at the primary level and lower returns for subsequent levels (Colclough et al., 2010; Fasih et al., 2012; Montenegro and Patrinos, 2014; Psacharopoulos and Patrinos, 2018; Patrinos, 2019). This is consistent with the idea of diminishing returns to education, where earnings increase with education but at a decreasing rate. Average returns were higher in lower-income countries with lower education levels, where primary-level returns were the highest, in contrast to high-income countries where tertiary-level returns were highest (Fasih et al., 2012; Montenegro and Patrinos, 2014; Psacharopoulos and Patrinos, 2018). More recent evidence suggests these patterns have changed. As would be predicted by human capital theory, levels of education have risen substantially and consequently, average returns have declined, however only marginally (Patrinos, 2024). Across most countries, the earnings-education relationship is now convex, with tertiary as opposed to primary-level education yielding the highest return, followed by primary and then secondary (Colclough et al., 2010; Fasih et al., 2012; Montenegro and Patrinos, 2014; Psacharopoulos and Patrinos, 2018; Patrinos, 2019). This was a consequence of declining

[3] Studies which use an Instrumental Variables (IV) design tend to find that 'naïve' Mincerian estimates *under*- rather than over-estimate the returns to education (Card, 1999; Barouni and Broecke, 2014; Gunderson and Oreopolous, 2020).

primary-level returns and rising tertiary-level returns, plausibly due to both supply- and demand-side factors, such as increased coverage of lower levels of schooling, and technological change favouring higher skills, operating separately or in combination (Fasih et al., 2012; Patrinos, 2019).[4] The increase in tertiary-level returns, combined with the significant expansion of the tertiary-educated population, suggests that the growth in demand for tertiary education has outpaced supply.

Heterogeneity in the returns to education serves as a saliant feature of the post-Apartheid South African economy. Generally, higher levels of education are strongly rewarded in the labour market in terms of both employment and, conditional on employment, earnings. Hertz (2003) finds conventional returns of 11 – 13 percent for Black/African workers in 1993. Keswell and Poswell (2004) estimate higher returns of 15 – 26 percent for the same group between 1993 and 2000. As part of their meta-analysis of returns around the world, Montenegro and Patrinos (2014) estimate an average return for South Africa of 16 percent in 2000, which grew to 21 percent in 2011. Returns in the country are high by international standards. This latter estimate serves as the second highest in their dataset of 139 countries, only marginally behind Rwanda's 22.4 percent. Salisbury (2016) estimates an average return of 18.7 percent in 2008. Consistent with the contemporary international literature, there is ample evidence of increasing convexity in South Africa's returns structure (Mwabu and Schultz, 2000; Keswelll and Poswell, 2004; Lam et al., 2012; Branson and Leibbrandt, 2013; Barouni and Broecke, 2014; Montenegro and Patrinos, 2014; van der Berg, 2014; Salisbury, 2016; Branson and Lam, 2021). In particular, the returns to primary have reduced, those to tertiary have grown, and those to secondary have remained approximately constant. This holds both on average and for each racial population group.[5] Explanations for such convexity are similar to those for the international context referenced above (Branson and Leibbrandt, 2013). Notably, while exhibiting lower levels of education, the return to tertiary education for Black/African workers has increased the most over time (Branson and Leibbrandt, 2013). This is broadly consistent with human capital theory, which relates higher returns with lower education levels. Despite this improvement, White workers have continued to receive a far greater return for an additional year of education, a differential which has only marginally improved over time (Sherer, 2000; Allanson et al., 2002; Salisbury, 2016). The reason behind this discrimination – the labour market's lower valuation of Black/African individuals' education – however, may no longer be primarily rooted in racial prejudice. Instead, it likely at least partially reflects differences in the quality of education, which have improved during the post-Apartheid period but remain a significant challenge at the time of writing (Spaull, 2013; Salisbury, 2016; Moses et al., 2017; Branson and Lam, 2021).

[4] Colclough et al. (2010) and Patrinos (2019) argue that the main causes are likely supply- rather than demand-side pressures.

[5] Race as a form of classification in South Africa is still widely used in the literature with the four largest race groups being African/Black, Indian/Asian, Coloured (mixed-race), and White. It is important to note that this serves a functional rather than normative purpose.

3 DATA

3.1 The Labour Force Surveys and Quarterly Labour Force Surveys

The analysis makes use of two distinct individual-level datasets covering most of the post-Apartheid period, both conducted by South Africa's national statistics office, Statistics South Africa (StatsSA). First, the Labour Force Survey (LFS) is a sample-based, cross-sectional household survey of approximately 30,000 households conducted bi-annually from 2000 to 2007 in February or March and September of each year. The survey uses a two-stage stratified sampling design, and the sample covers the non-institutional population with the exception of residents in workers' hostels. The LFS replaced the earlier October Household Surveys (OHS) which ran from 1993 to 1999, which I avoid additionally using here due to several representivity concerns experienced throughout its implementation (Kerr and Wittenberg, 2019). Additionally, the first three rounds of the LFS – March 2000 (2000:1), September 2000 (2000:2) and March 2001 (2001:1) – all had substantial problems that resulted in employment being significantly overestimated (Kerr and Wittenberg, 2019). I thus only make use of LFS data from September 2001 (2001:2) – September 2007 (2007:2).

The second dataset I use is StatsSA's Quarterly Labour Force Survey (QLFS). The QLFS replaced the LFS in the first quarter of 2008 (2008Q1) and remained in place at the time of writing. Like the LFS, the QLFS is also a sample-based, cross-sectional household survey of also approximately 30,000 households. It similarly uses a two-stage stratified sampling design, covering the non-institutional population with the exception of residents in workers' hostels. I use data from 2008Q1 – 2023Q4, the latest available wave at the time of writing. Both the LFS and QLFS serve as South Africa's official source of labour market statistics during their respective periods and thus include detailed demographic and labour market data for individuals over 15 years of age. After appending the QLFS and LFS waves together, I arrive at a sample of nearly 4 million working-age (15 – 64 years) individuals over the 23-year period. Throughout the analysis, I ensure all variables of interest – namely, educational attainment, labour market definitions, and earnings – are consistently defined both across and within the surveys.[6]

3.2 Earnings data

It is important to briefly discuss the nature of the earnings data. The LFS earnings data I use is publicly available while that of the QLFS is not but was privately provided by StatsSA.[7] Importantly, while earnings data from the QLFS is publicly available for most waves, the public data includes poor-quality earnings imputations for workers who did not

[6] Regarding educational attainment, researchers should be aware that the education codes in the LFS changed once in 2004:2 and in the QLFS twice in 2013Q3 and 2016Q3. Not accounting for these changes in any analysis of educational attainment using this data may result in significant measurement error.

[7] The earnings data collected in the QLFS is usually not released as part of the QLFS, but instead is released in a separate, annual dataset and publication called the Labour Market Dynamics of South Africa.

report them which, unfortunately, cannot be distinguished from the reported data. The LFS data does not include any such imputations. Those in the QLFS have been shown to produce implausible and volatile estimates, however reliable results can be obtained when the underlying unimputed data is used and adjusting for outliers and non-response (Wittenberg, 2017; Kerr and Wittenberg, 2021; Köhler et al., 2023; Köhler and Bhorat, 2023; Kerr, 2024). I therefore merge in the raw, unimputed QLFS earnings data privately provided to us by StatsSA. While earnings data is not available for all QLFS waves during 2008 and 2009,[8] it is for all other waves in both surveys. Consequently, this data represents the longest uninterrupted series of reliable wage data during the post-Apartheid period.[9]

Both surveys consistently ask all employed respondents to report their gross earnings in one of two ways. First, they are asked to report a value in South African Rands for what they earn over a select time period (for example, annually, monthly, or weekly). If they do not, they are then asked to select the bracket within which their earnings falls. Consequently, there are three categories of earnings response in the data: (i) those who report Rand values; (ii) those who do not report Rand values but do report their earnings bracket; and (iii) those who reported neither. Figure A1 in the appendix presents the share of employed respondents across these categories over the period.[10] In the early 2000s, about 75 percent reported a value in Rands, 20 percent reported a bracket, and the remaining 5 percent reported neither. The rate of non-response has, however, grown substantially over the period. In 2023Q4, only approximately 45 percent reported a value in Rands, while the share who reported a bracket has remained relatively constant. Concerningly, about 37% did not report either.[11]

I follow Köhler and Bhorat (2023) and adjust the raw data for both outliers and non-response using two parametric statistical techniques with consistent specifications to ensure comparability within and across the surveys. For outliers, I adopt the studentized regression residual approach, which identifies outliers by comparing their reported and predicted earnings.[12] Only 1 percent of reported earnings were identified and recoded as missing. For non-response, I impute earnings for workers who (i) neither reported their exact earnings nor a wage bracket, (ii) only reported a bracket, or (iii) were identified as

---

[8] The QLFS only started collecting wage data in 2009Q3, while the data for 2009Q3 and 2009Q4 have not been made available.

[9] While other datasets, such as the Post Apartheid Labour Market Series (PALMS), exist which similarly stacks cross-sectional household surveys to produce such a series, at the time of writing they include the problematic public QLFS earnings data.

[10] I observe abruptly more non-response in 2010Q3 and 2010Q4 relative to the remainder of the series, primarily due to respondents neither reporting a Rand value nor a bracket. An examination of the data shows that this non-response is evident from when respondents are asked to report their salary interval (that is, the frequency they receive their earnings) which occurs *before* they are asked to report their earnings. While the reason behind this non-response is unclear, the estimates throughout the analysis do not appear to be significantly affected by it.

[11] The reason underlying the deterioration of earnings response is unclear.

[12] Specifically, the studentized regression residual approach entails, first, estimating an expanded Mincerian wage regression of the logarithm of monthly earnings on a vector of observable covariates (here, the vector of observable covariates includes the usual Mincerian covariates – years of education and experience (and its squared term) – as well as age, sex, racial population group, province, marital status, main industry and occupation, a public sector indicator, a formal sector indicator, and survey wave fixed effects using ordinary least squares (OLS). Other variables which may be relevant, such as trade union membership and an urban versus rural indicator, are omitted because they are not available throughout the series. Second, the residuals are predicted and standardized. Finally, observations with large residuals are flagged as outliers. Following the literature, outliers are defined as those with absolute studentized residuals in excess of three.

outliers. I use multiple imputation (MI), which is similar to stochastic imputation but is advantageous in that it repeats the imputation process multiple times to produce multiple values of what the true data might be. Because the missing earnings data in the surveys exhibit a monotone pattern,[13] for each wave (apart from those in 2008 and 2009 for reasons outlined above) I first multiply impute a bracket for those in group (i) or (iii) by estimating an ordered logit model on a vector of observable covariates,[14] and thereafter multiply impute log monthly earnings based on the imputed bracket and the same vector using predictive mean matching with ten nearest neighbours. For observations in group (ii), the imputation process skips the first step. This process is repeated iteratively to arrive at ten imputations for each non-responder. The data are then combined using the standard rules for estimation and inference (Rubin, 1987).

Following the outlier and MI processes, I obtain wage data for almost all workers in the sample.[15] Figure A2 in the appendix presents the distributions of real hourly wages for the reported data, each iteration of the imputed data, and their combination in both the LFS and QLFS. Differences between the reported and imputed distributions are expected given the assumption that earnings data is missing not at random – that is, the probability of response is non-random, and in particular, tends to exhibit an inverse relationship with earnings itself (Wittenberg, 2017). As such, it is not surprising that, for every imputation iteration, the imputed data distributions are located to the right of the reported data distributions. Moreover, the former all exhibit a similar shape to one another, and are not indicative of unreasonable values. Finally, while the interested reader is referred to Köhler and Bhorat (2023) for a comprehensive outline of both the outlier and MI techniques, it should be noted that the resultant earnings estimates are strongly robust to varying imputation algorithms and number of imputations.

Throughout the analysis, the employed sub-sample includes all workers regardless of employment type, and hence speaks to the entire employed population. Their earnings are adjusted for inflation (benchmarked to January 2024) and are expressed as real hourly wages using data on usual working hours. All estimates are weighted using sampling weights and standard errors are adjusted for the complex survey designs.

[13] The monotone missingness pattern is due to the questionnaire's skip logic. If bracket wage data is missing, then exact wage data will be missing. As such, imputations are generated by specifying a sequence of independent univariate conditional imputation models.

[14] The selection of covariates is based on those which are required in the complete data model of interest, those which appear to determine missingness, and those which explain a considerable amount of the variance of log monthly earnings. These are included in both imputation models, and include age, sex, racial population group, years of education, experience (and its squared term), province, marital status, main industry and occupation, a public sector indicator, a formal sector indicator, and salary frequency.

[15] The share of successful imputations is relatively constant across survey waves. Earnings could not be imputed for just 2.7 percent of workers in the pooled sample due to missing data on variables used in the imputation model, such as sectoral formality and marital status.

4 METHODOLOGY

4.1 Average and heterogenous returns at the mean

The analysis begins with a descriptive examination of the following: changes to the distribution of educational attainment over time; variation in the probability of employment and, conditional on employment, real hourly wages by level of education over time; and finally, the evolving relationship between educational attainment inequality and wage inequality. Thereafter, I estimate the private returns to education through the use of Mincerian earnings functions. As described in Section 2, these models linearly relate wages to years of education, years of potential experience, and a quadratic term for years of potential experience which accounts for non-linearities in the relationship between potential experience and wages. Formally, I estimate the following parsimonious, conventional specification using Ordinary Least Squares (OLS):

$$log\ (wage)_{it} = \beta_0 + \beta_1 educ_{it} + \beta_2 exper_{it} + \beta_3 {exper_{it}}^2 + \varepsilon_{it} \qquad (1)$$

where $log\ (wage)_{it}$ represents the natural logarithm of worker i's real hourly wage in period t; $educ_{it}$ is a continuous years of education variable derived from respondents' self-reported highest levels of education; $exper_{it}$ is a continuous years of potential education variable, measuring the maximum amount of time individual i could have spent in the labour force, derived as their age minus $educ_{it}$ minus six; and ε_{it} is the error term. β_1 represents the estimated average private return to education for a given period, across all levels of education. As discussed in Section 2, I refrain from including additional explanatory variables which may introduce additional sources of bias and, crucially, will prevent β_1 from being appropriately interpreted as a rate of return.

In addition to estimating average returns, I follow the literature and examine four key sources of heterogeneity in returns. First, I analyse how β_1 has varied during the post-Apartheid period by estimating specification (1) separately for each period t. Second, I examine how β_1 varies across South Africa's four racial population groups by estimating specification (1) using stratified samples. This speaks to variation in the labour market value of an additional year of education for a given group, which would be suggestive of discrimination. I examine this source of variation both within a given period as well as over time. Third, I adjust specification (1) to what is known as the 'extended' Mincerian earnings function to estimate differential returns across levels of education. Formally, I estimate the following specification:

$$\begin{aligned} log\ (wage)_{it} = \beta_0 &+ \beta_1 educ_{it} + \beta_2 educ\ level_{it} \\ &+ \beta_3 educ_{it} \times educ\ level_{it} + \beta_4 exper_{it} \\ &+ \beta_5 {exper_{it}}^2 + \varepsilon_{it} \end{aligned} \qquad (2)$$

where $educ\, level_{it}$ is a categorical variable indicating individual i's highest level of schooling: primary; secondary; or tertiary. Education level-specific returns are estimated by taking the relevant derivative of $log\ (wage)_{it}$ with respect to $educ_{it}$. Finally, the fourth extension considers the fact that both specifications (1) and (2) produce rates of return at the mean of the wage distribution. Hence, the fourth extension considers return heterogeneity across the wage distribution. I do so through the use of Recentered Influence Function (RIF) regression, which I discuss in the next section.

Estimates obtained from specifications (1) and (2) are, of course, limited to the sub-sample of the population that are employed given that earnings data is not observable for the unemployed. Hence, using these estimates to draw inference over the entire population would be flawed given non-random selection into both labour force participation and, conditional on participation, employment. This is particularly relevant in the case of South Africa, which has been consistently characterized by one of the highest rates of unemployment globally throughout the post-Apartheid period. To address this, in addition to conventional estimates I present adjusted estimates to account for non-random selection into both labour force participation and employment. I do so through the use of an extension of Heckman's (1979) two-step procedure. For each wave, I fit a probit model to predict the probability of (narrow) labour force participation on a vector of demographic covariates[16] and use the predicted values $\hat{p}$ to calculate the inverse Mill's ratio (IMR) = $\frac{e^{-\frac{1}{2\hat{p}^2}}}{\sqrt{2\pi}\phi(\hat{p})}$. The IMR is then included as an additional covariate in a second selection equation which models the probability of employment, again using a probit and the same vector of demographic covariates. The predicted values from this model are then used to calculate a second IMR, which is included as an additional covariate in the conventional Mincer specification (1). The resulting estimates then reflect the true returns to education adjusted for sample selection bias. I conduct this procedure for the estimation of both average and heterogenous returns. These are my preferred estimates; however, I additionally present the conventional estimates for completeness.

4.2 Average and heterogenous returns across the wage distribution

The fourth extension considers differential returns across the wage distribution through the use of RIF regression. Also known as unconditional quantile regression, RIF regression was developed by Firpo et al. (2009) and expanded by Fortin et al. (2011) as a means of extending regression analysis beyond the mean to any point of the outcome distribution. In contrast to quantile regression which models the *conditional* relationship between a select covariate and a given point of the outcome distribution, RIF regression is more versatile in that it allows for the analysis of the *unconditional* relationship between said covariate and any distributional statistic of the outcome, easier implementation with

[16] This vector includes age, sex, race, province, marital status, and years of education, selected through the application of LASSO regression.

standard regression tools, computational efficiency, and smooth integration into decomposition frameworks. In these models, the outcome is the RIF of any functional of the outcome instead of the outcome itself. In my analysis, I focus on specific quantiles of the wage distribution. For quantile q_τ for $\tau \in (0,1)$ – for instance, $\tau = 0.1$ represents the 10th percentile – the RIF is formally defined as:

$$RIF(log(wage)_{it};\ q_\tau) = q_\tau + \frac{\tau - 1(log\ (wage)_{it} \leq q_\tau)}{f_Y(q_\tau)} \tag{3}$$

where $1(log(wage) \leq q_\tau)$ is an indicator function equal to one if $log(wage) \leq q_\tau$ and zero otherwise, and $f_Y(q_\tau)$ is the probability density function of log real hourly wages evaluated at q_τ. Once the RIF is calculated, it serves as the outcome in the standard Mincerian regression model:

$$RIF(log(wage)_{it};\ q_\tau) = \beta_\tau + \beta_{1\tau} educ_{it} + \beta_{2\tau} exper_{it} + \beta_{3\tau} exper_{it}{}^2 + \varepsilon_{it,\tau} \tag{4}$$

where $\beta_{1\tau}$ is the estimated quantile-specific return to education. I estimate this specification using the selection-adjusted models to examine returns from the 10th to the 90th percentile of the wage distribution, in 10-point increments, for each period. As before, I further examine such variation across levels of education and racial population groups.

4.3 Decomposition analysis

Finally, I seek to decompose variation in wages both (i) over time – to examine the individual contributions of changes to the distribution of productive characteristics (namely, education) versus changes to the returns to these characteristics – and (ii) across racial population groups within a given period – to examine the individual contributions of differences in these characteristics (again, namely education) versus differences in returns. The latter speaks to how the labour market differentially values different group's productive characteristics, and hence would be suggestive of discrimination. On the latter, I further examine how these components have varied over time. I conduct these decompositions both at the mean using Oaxaca-Blinder (OB) decomposition and across the wage distribution using RIF decomposition.

First, I use OB decomposition to decompose variation in mean wages between two time periods, and variation in mean wages between two given racial population groups (either African/Black, Coloured, or Indian/Asian relative to White) within a given period. Following Oaxaca (1973) and Blinder (1973), I assume wages can be expressed as a linear function of observable and unobservable covariates, as specified in equation (1), for either two time periods or groups denoted $t \in (1,2)$. A model which pools data for both periods or groups can simply be expressed as equation (1) without the period subscripts. If an indicator variable $T = 0$ for $t = 1$ and $T = 1$ for $t = 2$, then the following represents the difference in mean wages across periods or groups, denoting $\boldsymbol{X_i}$ as the righthand-side of the conventional Mincerian specification:

$$\begin{aligned} E[\,log(wage)_i \mid T = 1\,] - E[\,log(wage)_i \mid T = 0\,] \\ = E[\,\boldsymbol{X_i} \mid T = 1\,]'(\beta_2 - \beta) + E[\,\boldsymbol{X_i} \mid T = 0\,]'(\beta - \beta_1) \\ + (E[\,\boldsymbol{X_i} \mid T = 1\,] - E[\,\boldsymbol{X_i} \mid T = 0\,])'\beta \end{aligned} \tag{5}$$

Equation (1) can then be estimated as follows, where horizontal bar accents represent sample means:

$$\begin{aligned} \overline{log(wage)}_{i2} - \overline{log(wage)}_{i1} \\ = \left[\overline{X}'_{i2}(\hat{\beta}_2 - \hat{\beta}) + \overline{X}'_{i1}(\hat{\beta} - \hat{\beta}_1)\right] + \left(\overline{X}'_{i2} - \overline{X}'_{i1\beta}\right)\hat{\beta} \\ = \widehat{\Delta^{\mu}_{S}} + \widehat{\Delta^{\mu}_{X}} \end{aligned} \tag{6}$$

The first term in equation (6), $\widehat{\Delta^{\mu}_{S}}$, is referred to as the wage structure effect – the relative contribution of changes to the returns to productive characteristics to temporal average wage changes. The second term, $\widehat{\Delta^{\mu}_{X}}$, is referred to as the composition effect – the relative contribution of changes in productive characteristics to temporal average wage changes.[17]

Second, RIF decomposition is also used to decompose variation in wages either across periods or groups, but across the wage distribution instead of only at the mean. It operates in a similar way to OB decomposition. As described above in the case of RIF regression, the key difference to OB decomposition is that the outcome is the RIF of any functional of the outcome instead of the outcome itself – in this case, specific quantiles of the wage distribution. If f is the functional, then $\widehat{\Delta^{\mu}_{S}}$ and $\widehat{\Delta^{\mu}_{X}}$ in the case of the mean μ can be expressed here as:

$$\widehat{\Delta^{f}_{S}} = \sum_{j=1}^{k} \widehat{\Delta^{f}_{S,J}} = \sum_{j=1}^{k} \overline{X'_{2,J}}\left(\hat{\beta}^{f}_{2,j} - \hat{\beta}^{f}_{j}\right) + \overline{X'_{1,J}}\left(\hat{\beta}^{f}_{j} - \hat{\beta}^{f}_{1,j}\right) \tag{7}$$

$$\widehat{\Delta^{f}_{X}} = \sum_{j=1}^{k} \widehat{\Delta^{f}_{X,J}} = \sum_{j=1}^{k} \left(\overline{X'_{2,J}} - \overline{X'_{1,J}}\right)\beta^{f}_{j} \tag{8}$$

[17] These components are sometimes referred to as the 'price' and 'quantity' components, respectively.

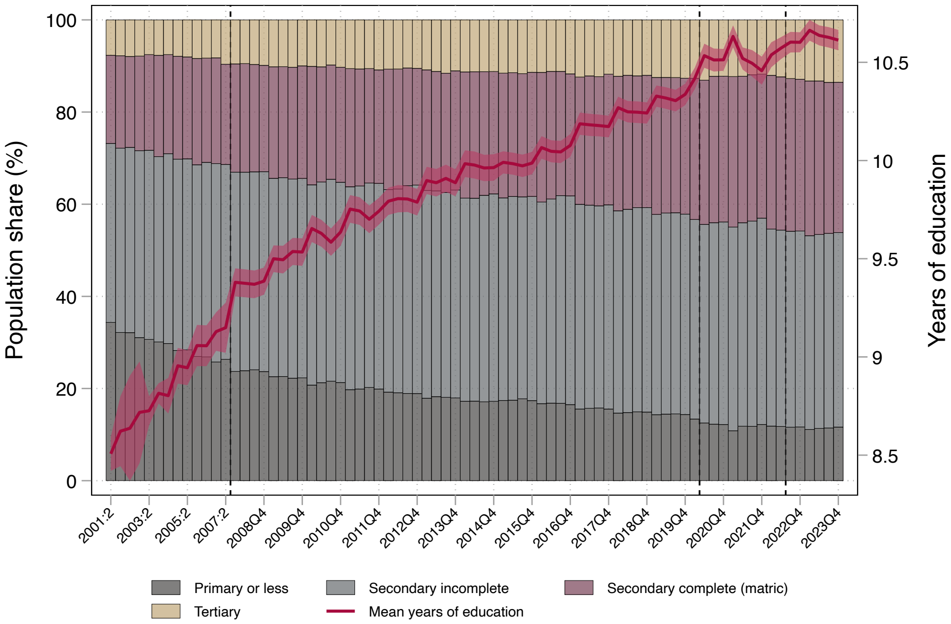
Both the OB and RIF decompositions are estimated using the selection-adjusted Mincerian specifications. For the latter, I again examine differences from the 10^{th} to the 90^{th} percentile of the wage distribution in 10-point increments.

5 DESCRIPTIVE STATISTICS

5.1 A rise in educational attainment and fall in educational attainment inequality

South Africa has experienced a substantial increase in educational attainment during the post-Apartheid period. Figure 1 and Table 1 present trends in the absolute and relative distributions of educational attainment from 2001 to 2023. The average working-age individual had 8.5 years of education in 2001, growing by over 2 years or 25 percent to 10.6 years by the end of 2023. This was driven by an expansion of both completed secondary and tertiary education alongside a contraction of primary or less. In 2001, just under 9.5 million working-age individuals or one-third (34 percent) of the population had a primary level education or less, which shrunk by 50 percent by the end of 2023 (4.7 million or just under 12 percent). Concurrently, the share of the population with a tertiary education nearly doubled from 8 to 13.5 percent. The share with a complete secondary education also grew considerably but by a marginally smaller rate, accounting for 19 percent of the population in 2001 and 33 percent in 2023. In other words, the share with at least a complete secondary grew from 27 to nearly half (46 percent) over the period. While the number of those with an incomplete secondary education grew in absolute terms, their share remained relatively unchanged at approximately 40 percent. These shifts are also clear when examining cumulative distribution functions of years of education, as shown in Figure A3 in the appendix. These plots also demonstrate first-order stochastic dominance, indicating that education levels have increased for all levels of education, reflecting in an unambiguous improvement in the education distribution.

Figure 1: Distribution of educational attainment, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2 – 2007:2, QLFS 2008Q1 – 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Vertical lines represent, in order, the change of the survey instrument, the onset of the COVID-19 pandemic in March 2020, and the repeal of all remaining COVID-19 pandemic restrictions in June 2022.

The pattern and magnitude of these shifts in the distribution of education are evident not just for those of working-age, but also for all labour market groups. As shown in panel (b) of Table 1, in both 2001 and 2023 the employed population exhibit higher levels of education than the broader population. This is not unexpected given the well-established role that education plays in determining both extensive- and intensive-margin labour market outcomes. This is evident when examining the distributions of educational attainment across labour market groups over time, presented in Figure A4 in the appendix, which highlights higher rates of employment for those with higher levels of education throughout the period. While I interrogate this further later, improvements in the education distributions for the employed, searching unemployed, and the economically inactive are clear. For the employed, those with a primary-level education shrunk from representing nearly one-third (31 percent) to just over 8 percent of the population, while those with a complete secondary or tertiary education grew from 24 to 37 percent and 15 to 24 percent, respectively. Similar to the working-age population as a whole, the respective incomplete secondary share remained constant at approximately 30 percent.

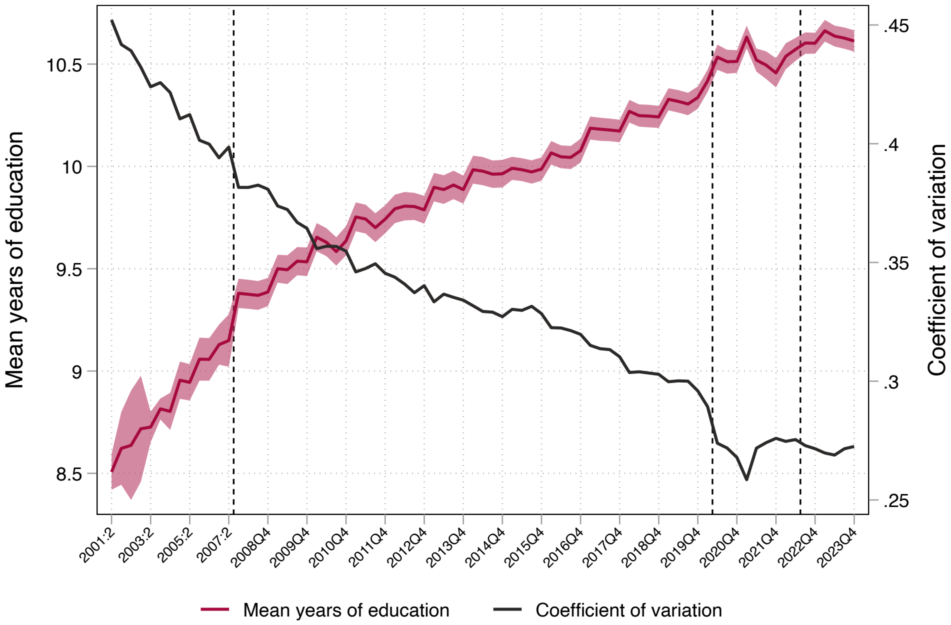
Table 1: Composition of working-aged and employed population by highest level of education, 2001 – 2023

	2001:2		2023Q4		Change		Share of change (%)
	Level	Share (%)	Level	Share (%)	Level	%	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel (a): Working-age population							
Primary or less	9 492 957 (159 532)	34.34	4 701 643 (102 743)	11.62	-4 791 315*** (189 754)	-50.47	-37.38
Secondary incomplete	10 735 797 (152 249)	38.84	17 084 322 (224 995)	42.22	6 348 525*** (271 666)	59.13	49.53
Secondary complete	5 291 465 (101 132)	19.14	13 197 440 (191 484)	32.62	7 905 975*** (216 549)	149.41	61.68
Tertiary	2 123 482 (75 828)	7.68	5 478 359 (124 463)	13.54	3 354 877*** (145 743)	157.99	26.17
Panel (b): Employed population							
Primary or less	3 415 905 (74 859)	31.04	1 352 896 (44 433)	8.23	-2 063 009*** (87 052)	-60.39	-37.95
Secondary incomplete	3 310 308 (60 939)	30.08	5 049 017 (97 702)	30.71	1 738 709*** (115 149)	52.52	31.98
Secondary complete	2 653 557 (70 881)	24.11	6 074 619 (110 252)	36.95	3 421 062*** (131 071)	128.92	62.92
Tertiary	1 624 935 (64 659)	14.77	3 964 978 (99 239)	24.12	2 340 043*** (118 444)	144.01	43.04

Source: Authors' own calculations using LFS 2001:2 and QLFS 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors are presented in parentheses and adjusted for the complex survey design. * p<0.10; ** p<0.050; *** p<0.001.

This substantial improvement in the distribution of educational attainment has resulted in a significant reduction in educational attainment inequality. Figure 2 presents trends in the means years of education alongside the coefficient of variation (CV), a popular relative measure of inequality calculated as the ratio of the standard deviation of years of education to its mean. Higher levels of education cause the mean level to increase and simultaneously the deviation of individual education levels from the mean to decrease. Consequently, the CV decreases. This pattern is clear here. From 2001 to 2023, while mean years of education grew by 25 percent from 8.5 to 10.6 years, the standard deviation fell by a similar magnitude from 3.8 to 2.9. Consequently, educational attainment inequality as per the CV fell by 40 percent.

Figure 2: Mean years of education and educational attainment inequality, 2001 – 2023

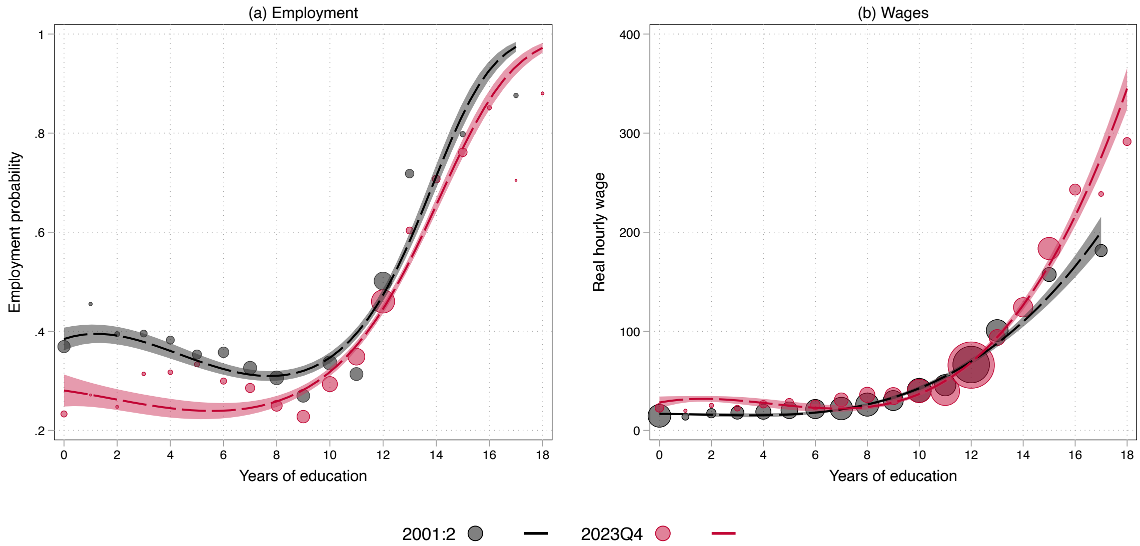


Source: Authors' own calculations using LFS 2001:2 – 2007:2, QLFS 2008Q1 – 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Vertical lines represent, in order, the change of the survey instrument, the onset of the COVID-19 pandemic in March 2020, and the repeal of all remaining COVID-19 pandemic restrictions in June 2022.

5.2 Temporal variation in labour market outcomes by level of education

As is the case in most countries globally, there is a strong, positive association between education and both employment and, conditional on employment, earnings in South Africa which has persisted throughout the post-Apartheid period. What has also persisted, and increased with respect to earnings, is the non-linear, convex nature of both relationships. As discussed in Section 2, while most countries only in recent years have come to exhibit such an earnings-education relationship, with tertiary as opposed to primary-level education yielding the highest return, such convexity has long characterised South Africa. Figure 3 presents binned scatterplots of employment probability in panel (a) and mean real hourly wages in panel (b) by years of education in both 2001 and 2023. Within either period, both the probability of employment and average wages are relatively constant until and inclusive of approximately 11 years of education. Thereafter, one additional year of education to 12 years, equivalent to a complete secondary qualification ("matric" in South Africa), yields a high return with respect to both outcomes, as indicated by the steep slopes of the fitted lines. Concerningly, employment rates have fallen across the entire education distribution, particularly for those at the bottom end. On the other hand, real wages appear to have risen for those with either low and high levels of education, while remaining constant for those in the middle, resulting in an even more convex relationship. Both the employment and wage dynamics likely partially reflect marginally worse aggregate labour market conditions over the period,[18] but also a combination of a change in the supply of and demand for workers across the education distribution.[19] I explore these dynamics in more detail in the modelling analysis to follow.

Figure 3: Binned scatterplot of employment probability and mean real hourly wages by years of education, 2001 versus 2023



Source: Authors' own calculations using LFS 2001:2:2 and QLFS 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors are adjusted for the complex survey design. Cubic functional form of the fitted line overlaid. Shaded area represents 95 percent confidence intervals. Wages adjusted for inflation and expressed in January 2024 Rands.

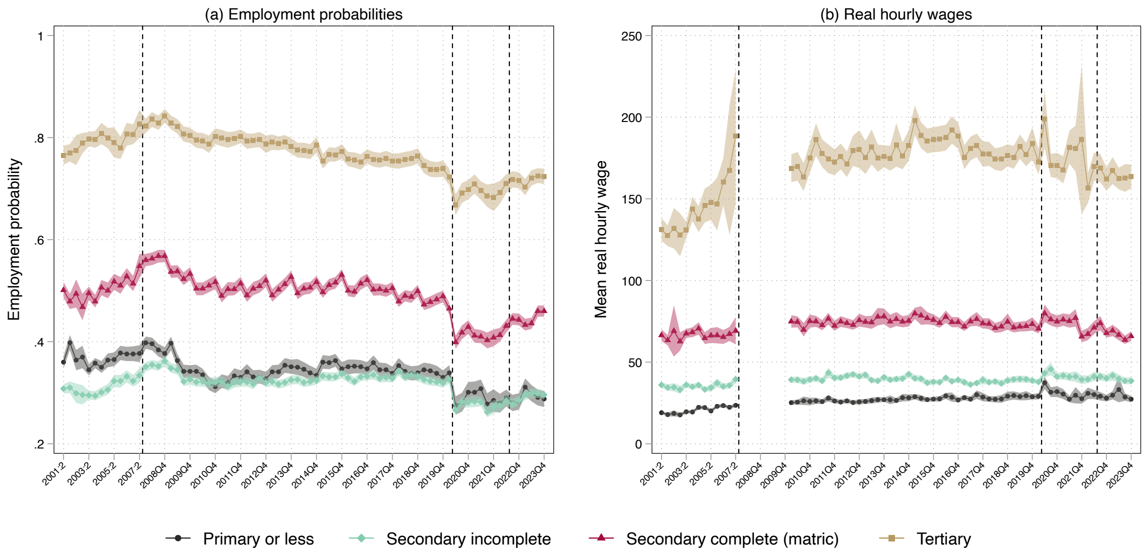
The changes to the earnings-education and earnings-employment relationships above are not due to a discrete change between 2001 and 2023, but instead reflect a gradual change over the period. Figure 4 presents trends in the probability of employment and mean real hour wages by broad level of education over the period. As implied above, in any period, it remains the case that more education is associated with both a better chance of employment and, conditional on employment, higher earnings. The one exception is those with an incomplete secondary, who generally face no better chance of employment compared to those with a primary or less, again highlighting the labour market's valuation of at least a complete secondary qualification as shown in Figure 3.

---

[18] South Africa's narrow unemployment rate were 29.5 percent in 2001:1, compared to an almost 10 percent higher 32.1 percent in 2023Q4. Employment rates were similar at about 40 percent in both periods.

[19] The wage gains at the bottom end may also be attributable to the introduction of sectional and national minimum wage legislation during the period.

Figure 4: Employment probability and mean real hourly wages by highest level of education, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2 – 2007:2, QLFS 2008Q1 – 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Vertical lines represent, in order, the change of the survey instrument, the onset of the COVID-19 pandemic in March 2020, and the repeal of all remaining COVID-19 pandemic restrictions in June 2022. Wage data not available from 2008Q1 – 2009Q4. Wages adjusted for inflation and expressed in January 2024 Rands.

In 2001, an individual with a tertiary education faced an employment probability of 77 percent – 54 percent larger than that of those with a complete secondary (50 percent) and more than double than those with an incomplete secondary or less. Until 2008, while this ranking remained in place, employment rates for all groups rose, reflecting improvement in aggregate labour market conditions in the run up to the Global Financial Crisis.[20] During this period, the wages of primary- and tertiary -educated workers also rose substantially by 24 and 44 percent in real terms, respectively. The latter estimate is significant but relatively imprecise.[21] Thereafter, employment rates gradually but persistently reduced throughout the period for all groups, but heterogeneously, falling the most for those with a primary education or less (20 percent relative to 2001). On the other hand, real wages continued to rise for all groups until 2015, especially for those at the lower and upper ends of the education distribution. Specifically, relative to 2001, wages rose by over 50 percent for both primary and tertiary-educated workers, in line with the estimates in Figure 3, and just 20 and 11 percent for those with a complete and incomplete secondary education, respectively. Thereafter, wages trended downwards, apart from a spike at the COVID-19 pandemic's onset which is primarily due to a composition effect; that is, a regressive distribution of job loss which mechanically drove wages upwards (Köhler and Bhorat, 2023). While higher than 2001 levels, relative to 2015, wages in 2023 were 17 percent lower

[20] The aggregate employment rate reached a peak of 46 percent in 2008, while concurrently, the narrow unemployment rate shrunk to 21.5 percent – the highest and lowest recorded, respectively, during the post-Apartheid period to date.

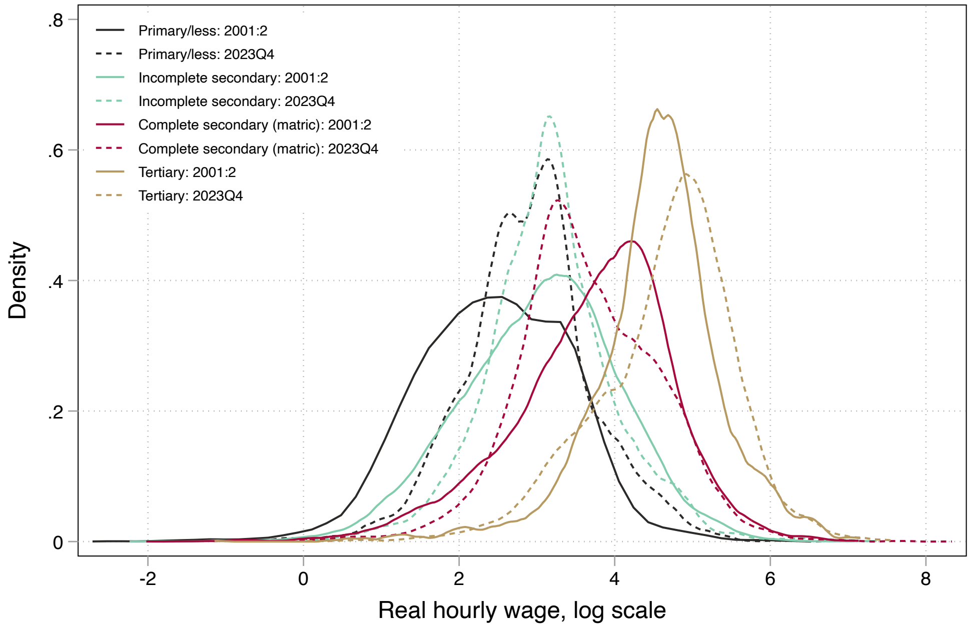
[21] The relevant 95 percent confidence interval is R147 – R230.

for those with at least a complete secondary, while those of lower education levels had remained relatively unchanged.

5.3 The evolving contribution of education to wage inequality

The preceding estimates show that post-Apartheid South Africa has simultaneously experienced a substantial rise in educational attainment accompanied by a reduction in employment probabilities for all education groups, especially for those with a primary level. Simultaneously and importantly, real wages have risen substantially for those with either a tertiary or primary education, while those with incomplete or complete secondary have remained relatively constant. This is suggestive of a U-shaped distribution of wage growth across the education distribution. This, together with the fact that a larger share of the population has acquired tertiary education, suggests that wage inequality may have reduced over the period. I descriptively investigate these dynamics further here. Figure 5 presents kernel density estimates of the real hourly wage distribution by level of education for 2001 and 2023. In both periods, significant wage inequality *between* education levels is clear, as indicated previously in Figure 4. In real terms, in both periods the average tertiary-educated worker earned 5.6 – 6.9 times that of the average worker with a primary education. What is perhaps less clear but still significant is the extent of inequality *within* education levels. In 2001, the CV of wages for primary-educated workers was 1.33, while that for tertiary-educated workers was lower at 0.94. These remained unchanged by 2023. In contrast, the CV for incomplete secondary workers was 1.48 in 2001 but 1.62 in 2023, and similarly for complete secondary workers, 1.17 in 2001 and 1.52 in 2023. In 2001, wage inequality was more pronounced among lower-educated workers and least among higher-educated workers. However, by 2023, this pattern changed. While inequality among lower- and higher-educated workers remained unchanged, that of workers in the middle of the education distribution increased to become the highest.

Figure 5: Distributions of real hourly wages by highest level of education, 2001 versus 2023

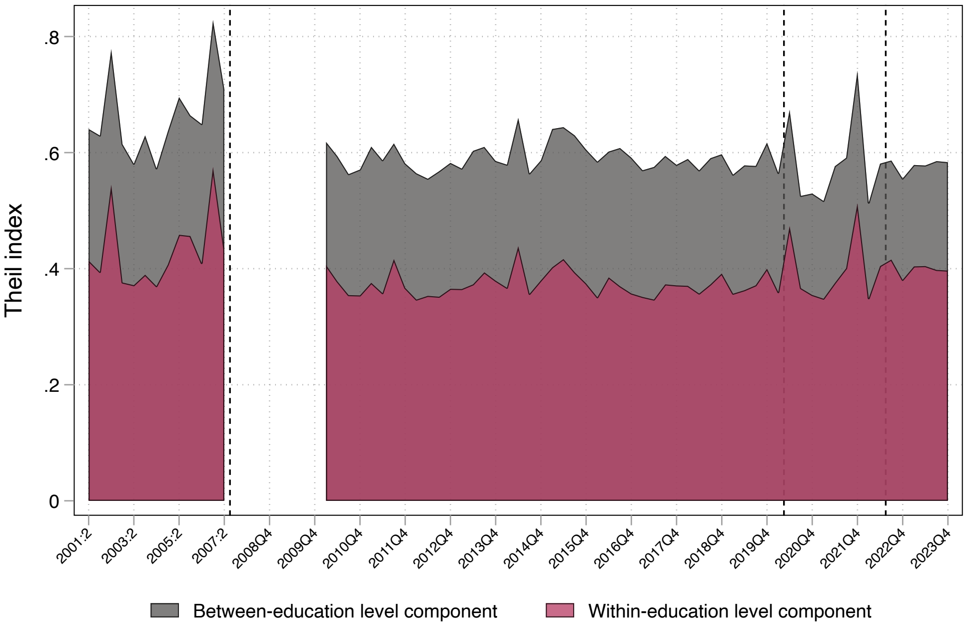


Source: Authors' own calculations using LFS 2001:2 and QLFS 2023Q4. *Notes*: Sample restricted to employed individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Wages adjusted for inflation and expressed in January 2024 Rands.

The aforementioned distributions make it clear the wage inequality with respect to education can be attributed to both inequality between as well as within education levels. Which is dominant? To examine this, I make use of the Generalised Entropy (GE) index with $\alpha = 1$, otherwise known as the Theil T index, which is particularly well-suited for decomposing inequality into within-group and between-group components because it is additively decomposable. The relevant estimates are presented in Figure 6. There are some fluctuations throughout the period, however overall, the estimates suggest that aggregate wage inequality reduced by 9 percent from 0.64 in 2001 to 0.58 in 2023. A smaller but still significant reduction is also clear when the Gini coefficient is alternatively used.[22] While inequality both between and within education levels contribute to total wage inequality throughout the period, as shown above, inequality *within* education levels is consistently the dominant component, explaining a relatively constant 60 – 70 percent of total wage inequality.

[22] I estimate that the Gini coefficient fell from 0.58 in 2001 to 0.56 in 2023. Variation in the amount of change in wage inequality across indices can be explained by variation in the sensitivity of indices to wage changes at different parts of the distribution.

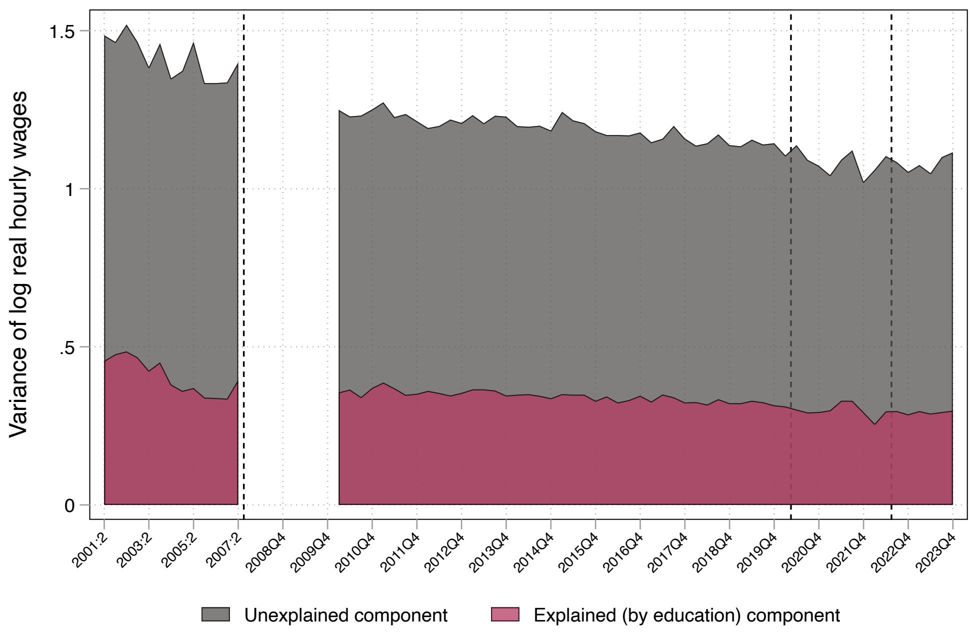
Figure 6: Decomposition of wage inequality into between- versus within-education level inequality, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2 – 2007:2; QLFS 2010Q1 – 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Vertical lines represent, in order, the change of the survey instrument, the onset of the COVID-19 pandemic in March 2020, and the repeal of all remaining COVID-19 pandemic restrictions in June 2022. Wage data not available from 2008Q1 – 2009Q4.

The above suggests that wage inequality within rather than between education levels primarily explain total wage inequality. What is the overall contribution of inequality in educational attainment, regardless of level, to wage inequality? To examine this, I conduct a variance decomposition; specifically, I calculate the variance of the logarithm of real hourly wages for each period, and decompose it into explained and unexplained components. The former is derived from the R^2 value of an OLS regression of the logarithm of real hourly wages on years of education. The latter is then simply $1 - R^2$. Using this measure, I observe a more apparent but gradual reduction in total wage inequality during the period of 25 percent, from a total variance of 1.49 in 2001 to 1.11 in 2023. As such, while inequality has reduced, it however remains high. I estimate that a relatively constant 27 – 31 percent of wage inequality is explained by educational attainment inequality throughout the period. In other words, differences in educational attainment continue to serve as a major driver of wage disparities. Together with the observed substantial increase in educational attainment, the implications of this on the relative returns to education is *ex ante* unclear. I analyse this formally in the next section.

Figure 7: Decomposition of wage inequality into education inequality versus other factors, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2 – 2007:2; QLFS 2010Q1 – 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Vertical lines represent, in order, the change of the survey instrument, the onset of the COVID-19 pandemic in March 2020, and the repeal of all remaining COVID-19 pandemic restrictions in June 2022. Wage data not available from 2008Q1 – 2009Q4.

6 MODELLING RESULTS

6.1 Average returns to education

In this section, I present the formal estimation of the private returns to education. Table 2 presents the relevant conventional and selection-adjusted Mincerian estimates for four periods: the initial period, 2001:2; ten years later in 2011Q4; just before the COVID-19 pandemic in 2019Q4; and five years later in the latest period, 2023Q4. As shown in panel (a), I estimate a conventional return of 20.7 percent in 2001; that is, one additional year of education is associated with 20.7 percent higher real hourly wages on average. By ten years later, the magnitude of this estimate has grown by approximately 10 percent to reach 22.4 percent, and remains relatively constant during the decade thereafter. By the end of the series in 2023, the return is marginally larger at 23.1 percent. The selection-adjusted returns are consistently lower, by a magnitude of approximately seven percentage points, but resemble similar dynamics over the period: From 13.6 percent in 2001 to reach 14.7 percent in 2023. Hence, while their magnitudes of growth marginally differ (12 versus 8 percent for the conventional versus selection-adjusted return, respectively), both set of estimates support the notion that the average returns to education have risen. This is despite the large (25 percent) increase in mean years of education over the period, suggesting that the demand for higher-educated workers has outpaced the increase in supply.

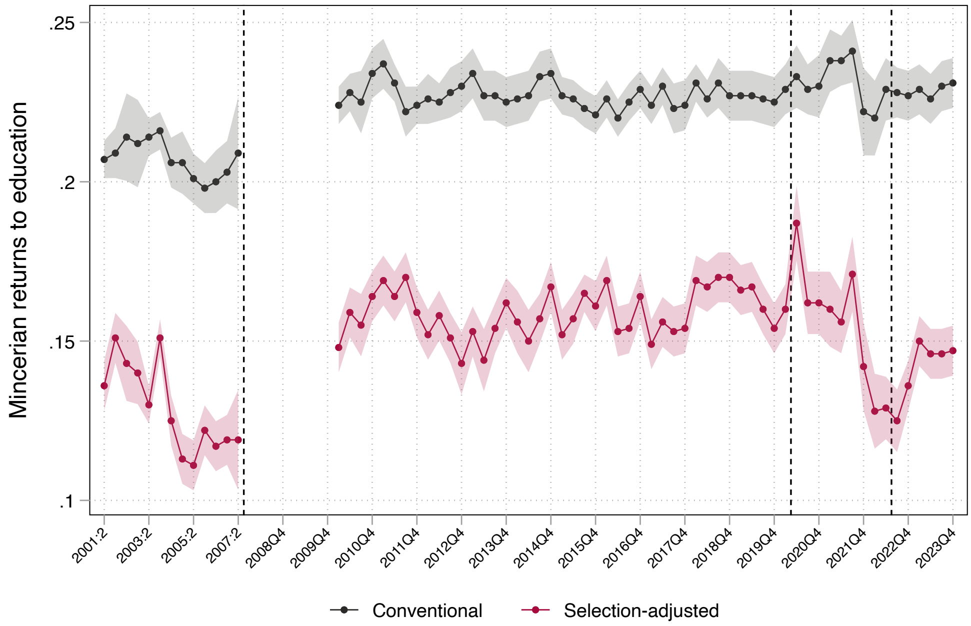
Table 2: Conventional and selection-adjusted Mincerian returns to education estimates, 2001 – 2023

	Period							
	2001:2		2011Q4		2019Q4		2023Q4	
	Return (1)	Mean years (2)	Return (3)	Mean years (4)	Return (5)	Mean years (6)	Return (7)	Mean years (8)
Panel (a): Conventional Mincerian estimates								
Estimate	0.207***	8.507***	0.224***	9.742***	0.225***	10.337***	0.231***	10.613***
	(0.003)	(0.039)	(0.003)	(0.027)	(0.004)	(0.025)	(0.004)	(0.024)
N	25,357		20,263		17,046		15,992	
R^2	0.360		0.373		0.346		0.325	
Panel (b): Selection-adjusted estimates								
Estimate	0.136***	8.507***	0.159***	9.742***	0.154***	10.337***	0.147***	10.613***
	(0.004)	(0.039)	(0.004)	(0.027)	(0.004)	(0.025)	(0.004)	(0.024)
N	25,357		20,263		17,046		15,992	
R^2	0.448		0.413		0.390		0.386	

Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Outcome = real hourly wages (log scale). Estimates weighted using sampling weights. Standard errors are presented in parentheses and adjusted for the complex survey design. All models additionally include *experience* and its squared term. Selection-adjusted models additionally include the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text. * p<0.10; ** p<0.050; *** p<0.001.

Figure 8 presents trends in conventional and selection-adjusted returns over the entire period. The conventional returns vary between 20 – 24 percent, while the selection-adjusted returns vary between 11 – 19 percent. Both sets of estimates indicate that the returns initially declined during the early 2000s until 2005, and rose thereafter to peak during 2010. In the decade approaching the COVID-19 pandemic, they fluctuated but remained relatively constant at 23 and 16 percent for the conventional and selection-adjusted returns, respectively. Dynamics during the pandemic, particularly for the selection-adjusted estimates, largely reflect substantial extensive-margin labour market adjustments which affected the composition of the employed population (Ranchhod and Daniels, 2021; Bassier et al., 2023; Köhler et al., 2022; Köhler, 2023; Köhler and Bhorat, 2023). Following the repeal of all remaining pandemic-related restrictions in June 2022, the conventional estimates returned and remained at their pre-pandemic level. In contrast, the selection-adjusted estimates rose but remained about one percentage point lower. As discussed by Köhler and Bhorat (2023), this change may reflect a pandemic-induced change to the structure of the labour market.

Figure 8: Trends in the conventional and selection-adjusted Mincerian returns to education, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2 – 2007:2; QLFS 2010Q1 – 2023Q4. Notes: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors are adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Selection-adjusted estimates obtained using the conventional Mincerian model specification but additionally including the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text. Vertical lines represent, in order, the change of the survey instrument, the onset of the COVID-19 pandemic in March 2020, and the repeal of all remaining COVID-19 pandemic restrictions in June 2022. Wage data not available from 2008Q1 – 2009Q4.

Table 3 presents the returns estimates for each of South Africa's four racial population groups. While the conventional estimates are presented in panel (a) for completeness, I focus my discussion on my preferred set – the selection-adjusted estimates – presented in panel (b). In line with the literature, I find evidence of substantial heterogeneity in returns by race. In 2001, African/Black and Indian/Asian workers experienced similar, relatively low returns of 12.6 and 12.3 percent, respectively – the lowest across groups. Yet, they exhibited varied levels of education – 7.9 and 10.8 years of education on average, respectively. Coloured workers experienced the highest rate of return across all groups (16.8 percent), but had relatively low levels of education on average (8.5 years). In contrast, White workers had both high levels of education (12 years, the largest across groups) alongside relatively high returns (15.1 percent). From 2001 to 2023, educational attainment across all groups rose by varying degrees, rising the most for African/Black workers (32 percent, from a low base) and the least for White workers (6 percent, from a high base). With the exception of Coloured workers whose rate of return remained unchanged, returns for all groups grew also by varying degrees. Notably, Indian/Asian workers' return to education doubled to 25.4 percent – the largest increase over the period, and the highest return in 2023, in sharp contrast to the group experiencing the lowest return 20 years prior. White workers' average return rose by 11 percent, while that of African/Black workers rose by 25 percent. Hence, despite its improvement and some fluctuation throughout the period, the return for the latter group remains among the lowest.

Table 3: Conventional and selection-adjusted Mincerian returns to education estimates, by race, 2001 – 2023

	Period							
	2001:2		2011Q4		2019Q4		2023Q4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Return	Mean years	Return	Mean years	Return	Mean years	Return	Mean years
Panel (a): Conventional Mincerian estimates								
African/Black	0.158***	7.889***	0.189***	9.353***	0.187***	10.088***	0.198***	10.420***
	(0.004)	(0.045)	(0.004)	(0.036)	(0.004)	(0.028)	(0.005)	(0.028)
Colour-ed	0.191***	8.477***	0.207***	9.761***	0.186***	10.116***	0.204***	10.340***
	(0.006)	(0.103)	(0.008)	(0.079)	(0.008)	(0.091)	(0.009)	(0.075)
Indian/Asian	0.145***	10.775***	0.172***	11.373***	0.239***	11.822***	0.275***	12.127***
	(0.014)	(0.139)	(0.020)	(0.149)	(0.018)	(0.137)	(0.018)	(0.143)
White	0.167***	12.037***	0.177***	12.480***	0.201***	12.735***	0.186***	12.734***
	(0.010)	(0.057)	(0.008)	(0.088)	(0.009)	(0.080)	(0.011)	(0.093)
Panel (b): Selection-adjusted estimates								
African/Black	0.126***	7.889***	0.167***	9.353***	0.164***	10.088***	0.157***	10.420***
	(0.005)	(0.045)	(0.005)	(0.036)	(0.005)	(0.028)	(0.006)	(0.028)
Colour-ed	0.168***	8.477***	0.154***	9.761***	0.161***	10.116***	0.172***	10.340***
	(0.007)	(0.103)	(0.011)	(0.079)	(0.010)	(0.091)	(0.011)	(0.075)
Indian/Asian	0.123***	10.775***	0.181***	11.373***	0.233***	11.822***	0.254***	12.127***
	(0.017)	(0.139)	(0.023)	(0.149)	(0.028)	(0.137)	(0.023)	(0.143)
White	0.151***	12.037***	0.155***	12.480***	0.186***	12.735***	0.168***	12.734***
	(0.011)	(0.057)	(0.009)	(0.088)	(0.012)	(0.080)	(0.011)	(0.093)

Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. Notes: Sample restricted to individuals of working-age (15 – 64 years). Outcome = real hourly wages (log scale). Estimates weighted using sampling weights. Standard errors are presented in parentheses and adjusted for the complex survey design. All models additionally include experience and its squared term. Selection-adjusted models additionally include the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text. * p<0.10; ** p<0.050; *** p<0.001.

6.2 Heterogenous returns to education across education levels

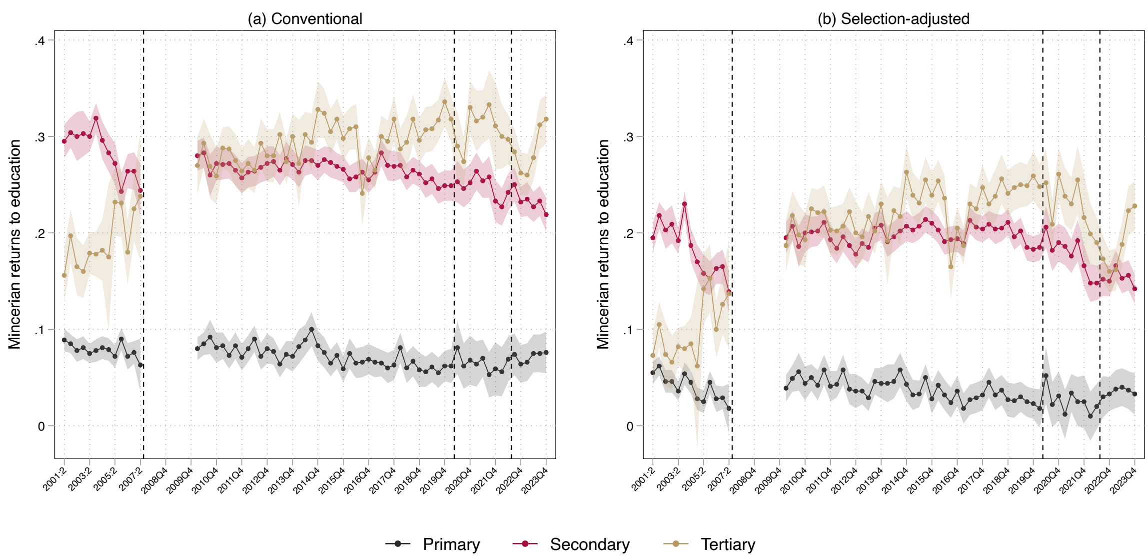
Table 4 presents estimates of the differential returns to education by level of education for the four distinct periods. As before, trends in the estimates for the entire period are presented in Figure 9. I again focus on the preferred set of estimates – the selection-adjusted estimates – presented in columns (2), (4), (6), and (8) in the table, and panel (b) in the figure. The data suggests that, in 2001, the returns to education in South Africa were highest for secondary education (19.5 percent) and lowest for primary education (5.5 percent). Over time, the returns to tertiary education have grown substantially – tripling in size – to reach nearly 23 percent in 2023. Simultaneously, the returns to lower levels have reduced, especially for primary which has shrunk by 40 percent. The set of conventional estimates show a similar pattern. Consequently, the returns structure has shifted to favour tertiary education, followed by secondary and then primary. These dynamics are strongly consistent with both the domestic and international literatures discussed in Section 2. Barring some fluctuation, the higher-frequency estimates in panel (b) of Figure 9, together with the estimates in columns (4) and (6), suggest that the returns to tertiary education steadily increased until approximately 2015 and remained relatively constant until the pandemic's onset.

Table 4: Conventional and selection-adjusted Mincerian returns to education estimates, by level of education, 2001 – 2023

	Period							
	2001:2		2011Q4		2019Q4		2023Q4	
	Conventional	Selection adjusted	Conventional	Selection adjusted	Conventional	Selection adjusted	Conventional	Selection adjusted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Primary	0.089***	0.055***	0.071***	0.041***	0.062***	0.023***	0.076***	0.033***
	(0.006)	(0.006)	(0.007)	(0.008)	(0.008)	(0.008)	(0.011)	(0.011)
Secondary	0.295***	0.195***	0.257***	0.193***	0.249***	0.183***	0.219***	0.142***
	(0.009)	(0.008)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)
Tertiary	0.156***	0.073***	0.264***	0.203***	0.336***	0.259***	0.318***	0.228***
	(0.014)	(0.013)	(0.012)	(0.011)	(0.013)	(0.013)	(0.013)	(0.012)
N	25,357	25,357	20,263	20,263	17,046	17,046	15,992	15,992
R^2	0.403	0.474	0.426	0.453	0.414	0.444	0.380	0.427

Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. Notes: Sample restricted to individuals of working-age (15 – 64 years). Outcome = real hourly wages (log scale). Estimates weighted using sampling weights. Standard errors are presented in parentheses and adjusted for the complex survey design. Estimates obtained by interacting years of education in a conventional Mincerian specification with level of education. All models additionally include experience and its squared term. Selection-adjusted models additionally include the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text. * p<0.10; ** p<0.050; *** p<0.001.

Figure 9: Trends in the conventional and selection-adjusted Mincerian returns to education, by level of education, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2 – 2007:2; QLFS 2010Q1 – 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors are adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Selection-adjusted estimates obtained using the conventional Mincerian model specification but additionally including the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text. Vertical lines represent, in order, the change of the survey instrument, the onset of the COVID-19 pandemic in March 2020, and the repeal of all remaining COVID-19 pandemic restrictions in June 2022. Wage data not available from 2008Q1 – 2009Q4.

Table 5 presents estimates of the differential returns to education by level of education as well as race over time. Panel (a) considers those of African/Black workers, which are strongly consistent with the aggregate dynamics presented in Table 4, albeit with some interesting nuances. In 2001, their returns were again highest for secondary education and lowest for primary education and are of similar magnitudes to the aggregate estimates. A similar ranking holds for Coloured workers, as shown in panel (b), however their returns for all levels were approximately 5 percentage points larger. In contrast, I do not find evidence of any return to either primary or tertiary education for Indian/Asian workers in 2001, as shown in panel (c).[23] These workers only exhibit a positive return to secondary education. Similarly, White workers appear not to experience a return to primary education, in 2001 or any period thereafter, but instead experienced the largest returns to secondary education. Over time, mirroring aggregate dynamics, the returns to tertiary education grew for all groups, while those of primary and secondary education contracted for African/Black, Coloured, and White workers. Notably, the return to tertiary tripled in size for African/Black workers, from 10 percent in 2001 to over 30 percent in 2023, with most of the change occurring from 2001 to 2011. Concurrently, the return to primary for Coloured workers disappeared. The lack of any return to primary education for White workers persisted throughout the period, while their return to secondary education remained their highest, despite the growth in the relevant tertiary coefficient. Interestingly, for Indian/Asian workers, I detect a positive return to primary education in 2019Q4 in contrast to preceding periods, while over the period, their returns to both secondary and tertiary education grew. By 2023, both African/Black and Coloured experienced the highest returns for tertiary education and lowest returns for primary education. In contrast, Indian/Asian and White workers experienced higher returns to secondary relative to other levels.

[23] I do not think this is due to the group comprising a relatively small sample in the data, given the magnitudes of the standard errors are not unlike those for other groups, and the coefficients are all close to zero.

Table 5: Conventional and selection-adjusted Mincerian returns to education estimates, by level of education and race, 2001 – 2023

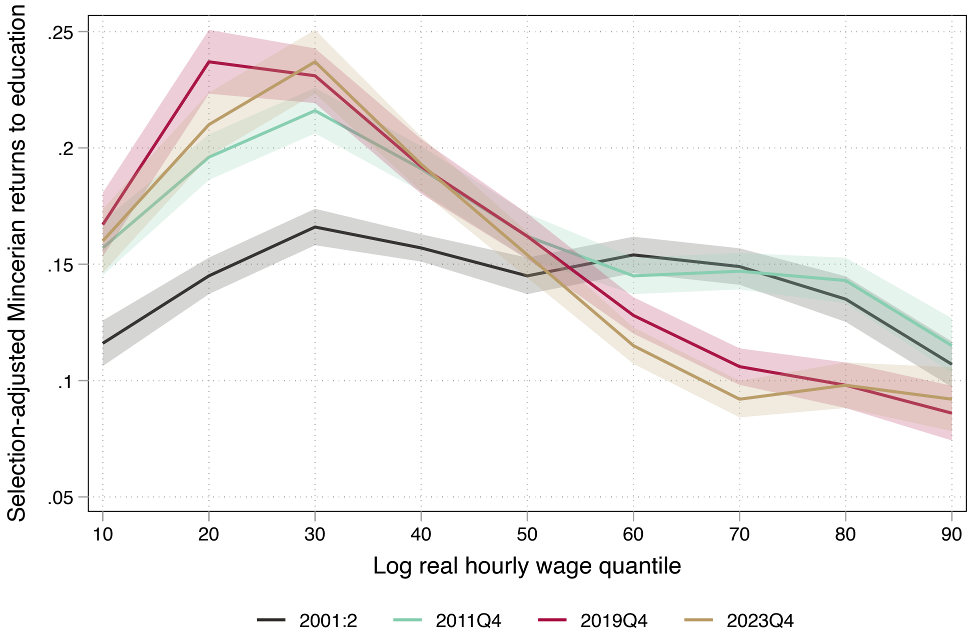
	Period							
	2001:2		2011Q4		2019Q4		2023Q4	
	Conventional (1)	Selection adjusted (2)	Conventional (3)	Selection adjusted (4)	Conventional (5)	Selection adjusted (6)	Conventional (7)	Selection adjusted (8)
Panel (a): African/Black								
Primary	0.073***	0.053***	0.054***	0.049***	0.045***	0.038***	0.061***	0.046***
	(0.006)	(0.006)	(0.008)	(0.008)	(0.009)	(0.009)	(0.012)	(0.012)
Secondary	0.196***	0.162***	0.205***	0.195***	0.204***	0.192***	0.180***	0.151***
	(0.010)	(0.010)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)	(0.010)
Tertiary	0.156***	0.100***	0.316***	0.305***	0.364***	0.350***	0.339***	0.304***
	(0.021)	(0.017)	(0.019)	(0.020)	(0.019)	(0.019)	(0.017)	(0.017)
Panel (b): Coloured								
Primary	0.123***	0.109***	0.085***	0.060***	0.032	0.022	0.041	0.030
	(0.014)	(0.014)	(0.021)	(0.021)	(0.020)	(0.020)	(0.030)	(0.029)
Secondary	0.237***	0.213***	0.220***	0.169***	0.197***	0.179***	0.197***	0.172***
	(0.015)	(0.015)	(0.014)	(0.018)	(0.013)	(0.016)	(0.017)	(0.017)
Tertiary	0.189***	0.153***	0.221***	0.183***	0.254***	0.238***	0.293***	0.276***
	(0.040)	(0.039)	(0.027)	(0.028)	(0.034)	(0.036)	(0.042)	(0.042)
Panel (c): Indian/Asian								
Primary	0.039	0.030	-0.022	-0.023	0.157**	0.157**	.	.
	(0.061)	(0.056)	(0.132)	(0.133)	(0.061)	(0.061)	.	.
Secondary	0.173***	0.154***	0.215***	0.223***	0.258***	0.253***	0.278***	0.252***
	(0.032)	(0.033)	(0.043)	(0.045)	(0.042)	(0.050)	(0.041)	(0.048)
Tertiary	0.014	-0.003	0.087**	0.104**	0.189***	0.185***	0.186***	0.165***
	(0.037)	(0.040)	(0.042)	(0.045)	(0.050)	(0.055)	(0.038)	(0.042)
Panel (d): White								
Primary	0.126	0.122	0.072	0.205	0.045	0.034	-0.106	-0.099
	(0.111)	(0.100)	(0.299)	(0.321)	(0.031)	(0.032)	(0.116)	(0.113)
Secondary	0.183***	0.168***	0.154***	0.128***	0.250***	0.235***	0.165***	0.154***
	(0.027)	(0.026)	(0.024)	(0.024)	(0.027)	(0.029)	(0.036)	(0.037)
Tertiary	0.113***	0.094***	0.183***	0.164***	0.197***	0.182***	0.148***	0.135***
	(0.020)	(0.021)	(0.019)	(0.018)	(0.016)	(0.019)	(0.019)	(0.018)

Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Outcome = real hourly wages (log scale). Estimates weighted using sampling weights. Standard errors are presented in parentheses and adjusted for the complex survey design. Estimates obtained by interacting years of education in a conventional Mincerian specification with level of education. All models additionally include *experience* and its squared term. Selection-adjusted models additionally include the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text. * $p<0.10$; ** $p<0.050$; *** $p<0.001$.

6.3 Heterogenous returns to education across the wage distribution and education levels

The RIF estimates of the differential returns to education across the wage distribution are presented in Figure 10 for the four distinct periods. For brevity, only the selection-adjusted estimates are presented. In 2001, rates of return were relatively constant, averaging around 15 percent per year of education across most of the distribution. The exceptions were at the very bottom and very top where the returns were marginally lower at approximately 10 percent, giving off the impression of an inverse-U shape. This suggests that education may be less effective in improving wages for the least and most advantaged workers during the period. Over time, however, while returns have remained positive across the distribution, they became significantly stronger towards the lower end of the distribution, and simultaneously and weaker towards the upper end. Among the lowest-earning 40 percent of workers, returns grew by 37 percent on average, while among the higher-earning 40 percent of workers, returns shrunk by 26 percent on average. These changes are likely related to both supply-side and demand-side factors, which I discuss in more detail later. Whatever the mechanism, these dynamics suggest that changes in the returns structure have contributed to reducing wage inequality by benefiting lower-wage workers more.

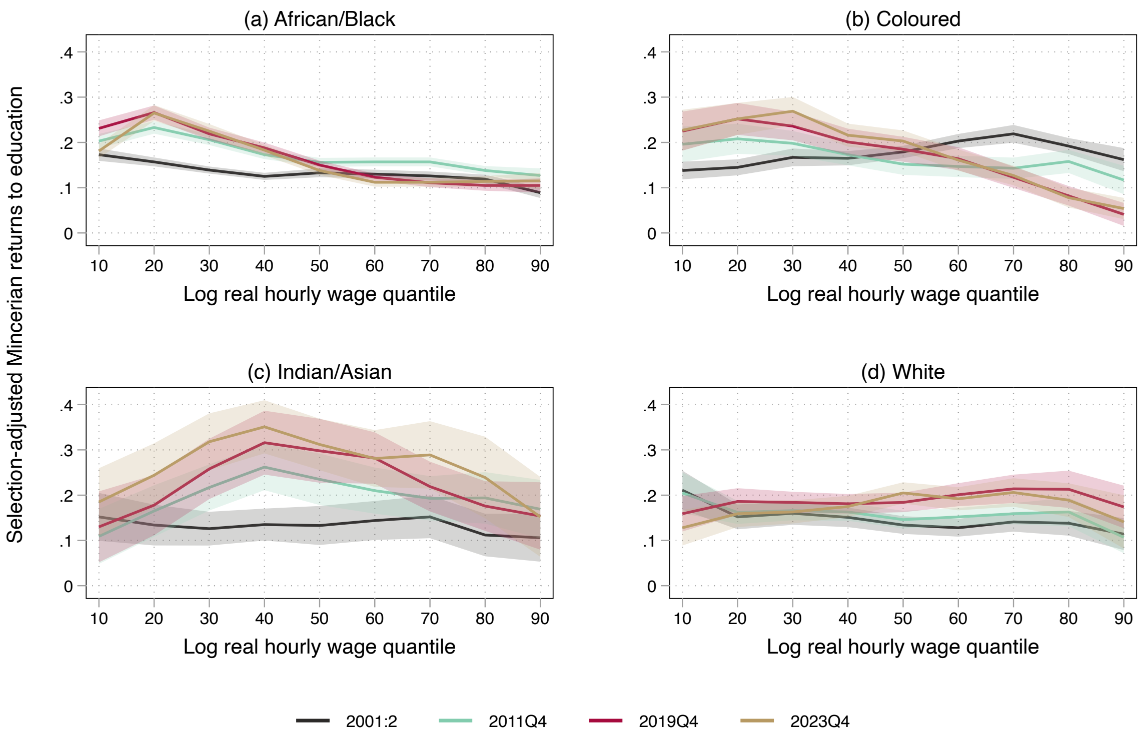
Figure 10: Recentered Influence Function estimates of the selection-adjusted Mincerian returns to education across the wage distribution, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors are adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Selection-adjusted estimates presented and obtained using the conventional Mincerian model specification but additionally including the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text.

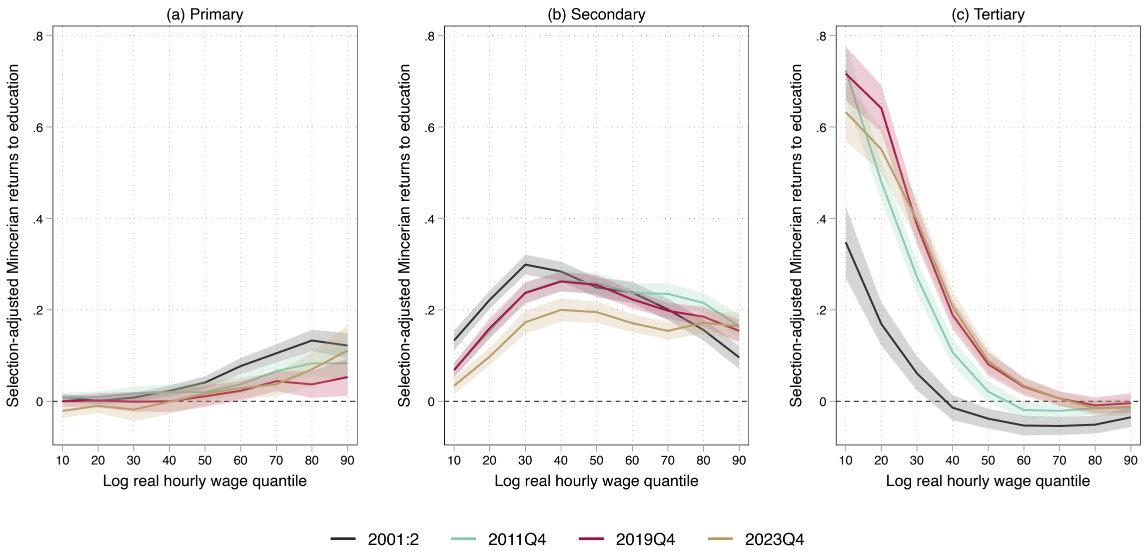
Like Figure 10, in Figure 11 I present the estimates of the differential returns to education across the wage distribution for each racial population group. I find evidence of substantial heterogeneity in returns both within and between groups over time. For African/Black workers, in 2001 their returns were marginally higher at the bottom of the distribution but otherwise relatively uniform. By ten years later, returns across the distribution grew, however those in the bottom half grew considerably more – by up to 50 percent. By 2023, these higher returns in the bottom half persisted, with the exception of the return at the very bottom, while those in the top half returned to their 2001 levels. For Coloured workers, in 2001 returns tended to grow from the bottom through to the 70th percentile and thereafter decrease marginally. Over time, similar to their African/Black counterparts, their returns towards the bottom increased, but unlike them, their returns towards the top decreased, approaching zero at the very top. For Indian/Asian workers, the distribution of returns was relatively constant in 2001. However, over time an inverse-U shape emerged, with returns increasing among those in the middle of the distribution but remaining unchanged at both tails. The distribution of returns for White workers in 2001 was similar to that of African/Black workers, thus favouring lower-wage workers. Over time, this pattern reversed to favour higher-wage workers. Returns reduced at the very bottom, remained unchanged among the remainder in the bottom half, and increased in the top half. Overall, by 2023, at the bottom of the distribution the returns for all groups were similarly higher than White workers, reflecting a significant shift from 2001 when returns were highest for this latter group. In the middle, returns were highest for Indian/Asian workers and lowest for African/Black workers; and at the top, the returns were again highest for Indian/Asian workers and lowest for Coloured workers, the latter of whom experienced the highest returns previously.

Figure 11: Recentered Influence Function estimates of the selection-adjusted Mincerian returns to education across the wage distribution, by race, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors are adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Selection-adjusted estimates presented and obtained using the conventional Mincerian model specification but additionally including the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text.

Figure 12: Recentered Influence Function estimates of the selection-adjusted Mincerian returns to education across the wage distribution, by level of education, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors are adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Selection-adjusted estimates presented and obtained using the conventional Mincerian model specification but additionally including the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text.

The estimates in both Figure 10 and Figure 11 highlight variation in average returns to education at specific points of the wage distribution over time. As shown previously, these returns vary considerably across levels of education. In Figure 12, I present the same estimates but additionally distinguish between primary, secondary, and tertiary-level education. In 2001, recall that at the mean of the distribution returns were highest for secondary and lowest for primary. The estimates here show that this ranking strongly depends on the point of the wage distribution. The returns to tertiary education were highest and increasing towards the bottom end, but were approximately zero within the top half of the distribution. The returns to secondary education were dominant in the remainder of the distribution, while returns to primary education were relatively low but increased towards the top end. Similarly, in 2023, recall that at the mean of the distribution returns became highest for tertiary followed by secondary and then primary. This ranking again varies across the distribution. It holds for the bottom 30 percent of workers, whereas at the 40th percentile, the returns to tertiary and secondary education are similar. For most of the top 50 percent of workers, in contrast, the returns to tertiary education were approximately zero while those to secondary education were highest and relatively uniformly distributed. Over time, the returns to tertiary education have increased substantially, especially for lower-wage workers, while those to lower levels of education have generally reduced across the distribution but to a smaller degree. Figure A5 in the appendix presents the equivalent estimates but additionally by race, revealing another source of interesting variation. Overall, these dynamics are consistent with the preceding estimates in suggesting that changes in the returns structure, by benefiting lower-wage workers more, have placed downward pressure on wage inequality.

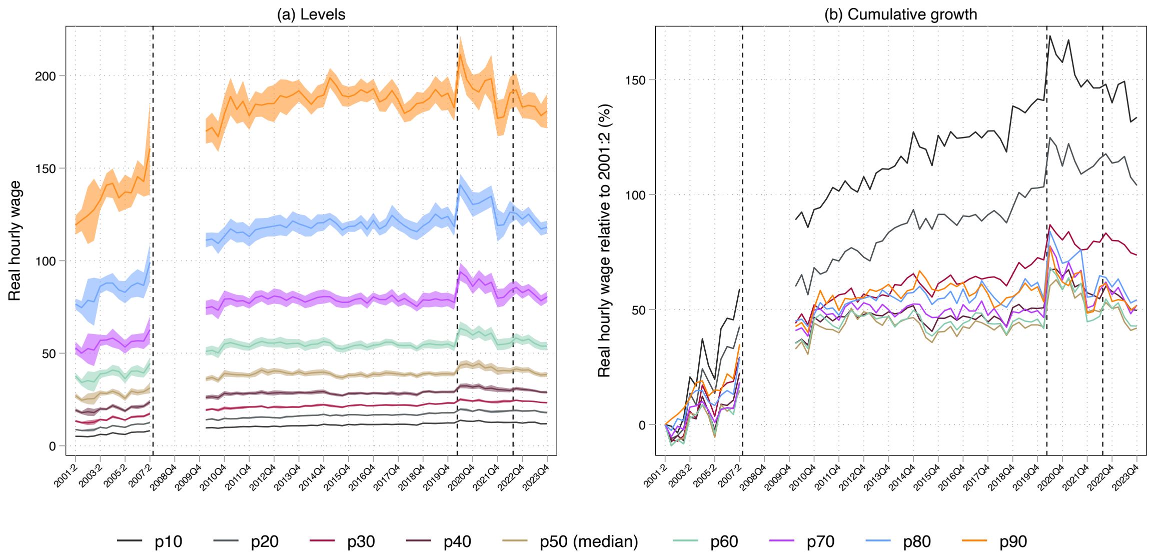
6.4 Decomposition analysis

The preceding estimates show that a larger share of the population is acquiring higher levels of education, particularly tertiary education, while concurrently, the returns to education have changed to benefit both tertiary-educated and lower-wage workers. To what extent does each component explain changes in wages, both on average and across the distribution, over time? In other words, how much of the observed wage changes are due to changes in productive characteristics, specifically with respect to education, versus changes in the returns to these characteristics? Which is dominant? Similarly, to what extent does each component explain differences in wages across racial population groups, and how has this changed over time? To answer these questions, in this section I present the results of the OB and RIF decompositions. For brevity again, only the selection-adjusted estimates are presented.

Before doing so, Figure 13 presents trends in real hourly wages across the wage distribution over the period. Despite the extreme degree of inequality throughout, I observe strong real wage growth concentrated towards the bottom of the distribution. From 2001 to 2023, median real wages grew from approximately R27 to R39 per hour, equivalent to about 2 percent per year on average or a cumulative 42 percent.[24] Growth was much stronger towards the bottom of the distribution. Wages at the 10th and 20th percentiles more than doubled in real terms, recording cumulative growth rates of 133 and 104 percent, respectively. Real wage growth is also evident across the rest of the distribution albeit at a lower rate of approximately 52 percent on average. The distribution of wage growth of the period is presented through the use of growth incidence curves (GIC's) presented in Figure A6 in the appendix. Together with those in Figure 13, the estimates suggest that wages at the bottom of the distribution have partially caught up to those in the middle while those at the top have moved away from the middle. This is strongly consistent with Kerr's (2024) recent analysis of both other household survey and tax administrative data, but stands in contrast to studies which use the problematic, publicly available QLFS data. Moreover, the GIC's suggest that growth was much stronger across the distribution from 2001 to 2011 relative to the period thereafter; however, a pro-poor distribution is evident throughout. In line with the preceding estimates, the consequence of such a heterogenous, largely progressive distribution of wage growth was a decrease in wage inequality.

[24] This former annual growth rate specifically denotes the annualized compound growth rate over the entire period.

Figure 13: Real hourly wages across the wage distribution, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2 – 2007:2, QLFS 2008Q1 – 2023Q4. *Notes*: Sample restricted to employed individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Vertical lines represent, in order, the change of the survey instrument, the onset of the COVID-19 pandemic in March 2020, and the repeal of all remaining COVID-19 pandemic restrictions in June 2022. Wage data not available from 2008Q1 – 2009Q4. Wages adjusted for inflation and expressed in January 2024 Rands.

Table 6 presents the results of the OB decomposition of temporal variation in mean wages. The decomposition is conducted across the four aforementioned periods with each relative to the first (2001). From 2001 to 2023, as shown in column (3), real wages increased by 0.441 natural log points which, in levels, amounts to growth from approximately R50 to R70 per hour. Of this change, the estimates suggest that 0.255 log points (or 58 percent) is explained by the composition effect. In other words, most of the observed change in mean wages over the period is explained by changes in productive characteristics. While this refers to both education and potential experience, it is primarily the former. While the changes to both are statistically significant, mean years of potential experience rose by just 2.8 percent from 18.4 to 18.9 years while, as discussed prior, mean years of education rose by a nine times higher rate – 25 percent. While this component is dominant, the magnitude of the structure effect is non-negligible: 42 percent of the change in mean wages is due to the change in the mean returns to these characteristics – again primarily education which, as shown earlier, rose by 1.1 percentage points (8 percent).[25] The estimates in columns (1) and (2) show that the composition effect was dominant throughout the period, but accounted for a much higher share (80 percent) during the first decade. Conversely then, changes to the mean returns to education have increased in importance considerably over time, from explaining 20 percent of temporal wage variation from 2001 – 2011 to 42 percent from 2001 – 2023.

[25] Estimates of the returns to experience remained close to zero over the period.

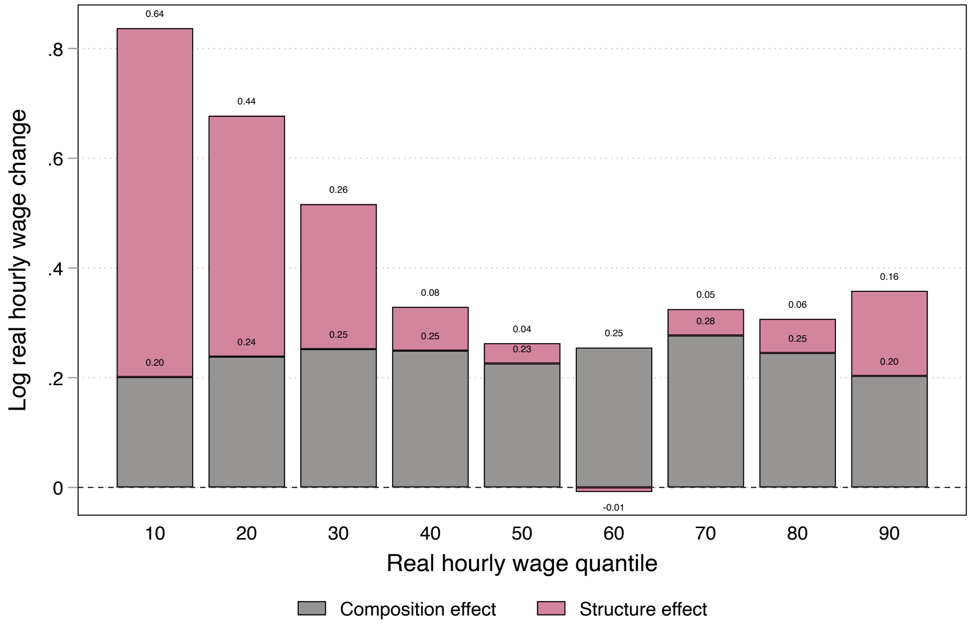
Table 6: Oaxaca-Blinder decomposition estimates of temporal variation in mean real hourly wages, by period

	2001:2 - 2011Q4 (1)		2001:2 - 2019Q4 (2)		2001:2 - 2023Q4 (3)	
Pre mean log real hourly wage	3.238***		3.238***		3.238***	
	(0.010)		(0.010)		(0.010)	
Post mean log real hourly wage	3.654***		3.717***		3.679***	
	(0.009)		(0.009)		(0.009)	
Difference in log real hourly wage	0.416***		0.479***		0.441***	
	(0.014)		(0.014)		(0.014)	
… Due to composition effect (Δ productive characteristics)	0.333***	80%	0.283***	59%	0.255***	58%
	(0.010)		(0.012)		(0.014)	
… Due to structure effect (Δ returns to productive characteristics)	0.083***	20%	0.196***	41%	0.186***	42%
	(0.011)		(0.013)		(0.015)	
N	45,574		42,357		41,303	

Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. Notes: Sample restricted to individuals of working-age (15 – 64 years). Outcome = real hourly wages (log scale). Estimates weighted using sampling weights. Standard errors are presented in parentheses and adjusted for the complex survey design. Models include the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text. * p<0.10; ** p<0.050; *** p<0.001.

Looking beyond the mean, Figure 14 presents the RIF decomposition estimates of temporal variation in wages across the distribution. The estimates show that, in contrast to the mean case, the dominance of the composition effect strongly depends on the point of the distribution. In absolute terms, it explains a constant amount of 0.2 – 0.28 log points in changes in wages across the distribution. In relative terms, while the composition effect primarily explains changes across most of the distribution, the structure effect is notably dominant among lower-wage workers; that is, the group who experienced the largest wage gains. Among workers at the bottom 20 percent of the distribution, 65 – 71 percent of the observed change in wages is due to the change (growth) to their returns to productive characteristics. Recall that, as per the RIF estimates in Figure 10, returns at this point of the distribution became significantly stronger over time, while concurrently weakening towards the upper end.

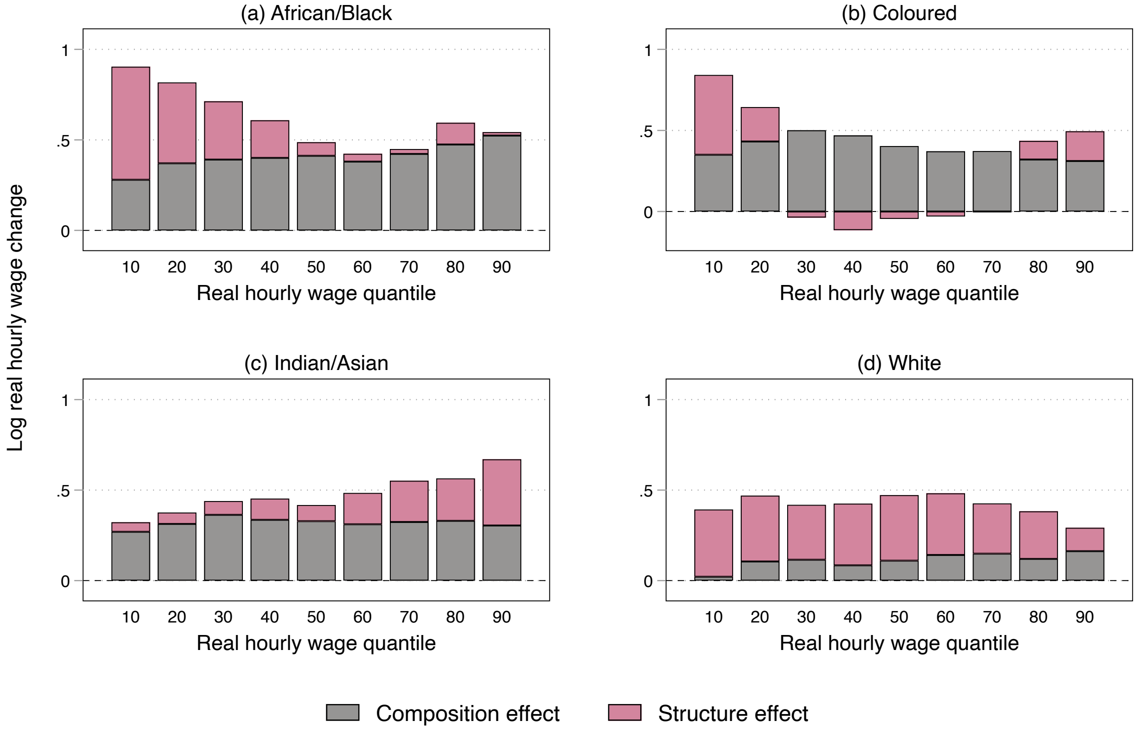
Figure 14: Recentered Influence Function decomposition estimates of temporal variation in real hourly wages across the wage distribution, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2; QLFS 2023Q4. Notes: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Selection-adjusted estimates presented and obtained using the conventional Mincerian model specification but additionally including the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text.

Figure 15 considers the equivalent but race-specific estimates as in Figure 14 for the drivers of temporal variation in wages across the distribution. As before, the estimates show that the dominance of either the composition or structure effect varies across the distribution both within and across racial groups. Notably, for all except White individuals, changes in productive characteristics primarily explain changes in wages across most of the distribution over time. The absolute magnitude of this component is relatively constant both within and between these groups. Among African/Black individuals, the decomposition largely reflects the aggregate one presented in Figure 14, which is unsurprising given their population share. Similarly, then, growth in the returns to productive characteristics was the primary force in driving wages upwards among lower-wage African/Black workers. The estimates for Coloured individuals exhibit a similar pattern, however the structure effect appears only dominant at the very bottom of the distribution. Among Indian/Asian individuals, changes in returns become increasingly important in explaining temporal wage changes towards the top of the distribution. In contrast to these groups, while growth in White individuals' wages is relatively uniform across the distribution, this growth is primarily explained by growth in their returns to education. The relatively small role played by the composition effect for this group is likely explained by their already relatively high education levels, which increased by just 0.7 years (6 percent) in the over-20-year period, in strong contrast to the large growth experienced by other racial groups.

Figure 15: Recentered Influence Function decomposition estimates of temporal variation in real hourly wages across the distribution, by race, 2001 – 2023



Source: Authors' own calculations using LFS 2001:2; QLFS 2023Q4. Notes: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Selection-adjusted estimates presented and obtained using the conventional Mincerian model specification but additionally including the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text.

The previous estimates consider the drivers of wage variation between periods, and hence do not speak to drivers within periods. Table 7 presents the OB decomposition estimates of wage differentials between racial population groups with White individuals serving as the reference group.[26] By analysing the decomposition for the four periods, as before, I can examine how variations in productive characteristics versus the returns to these characteristics explain the evolution of wage differentials over time. Several sources of heterogeneity are clear. The dominance of either the composition or structure effect varies both within and across groups as well as over time. In 2001, most (67 – 73 percent) of the mean wage differentials between African/Black, Coloured, and Indian/Asian versus White individuals were due to differences in productive characteristics. Again, this is primarily with respect to education as opposed to potential experience. On average, White individuals exhibit just 0.4 – 1.5 (2 - 7 percent) more years of potential experience, but 1.3 – 4.2 (11 – 35 percent) more years of education. Conversely, up to one-third (27 – 33 percent) of wage differentials in the period were due to differences in the returns to these characteristics, often referred to as suggestive evidence of discrimination. Again, for every

[26] White individuals are chosen as the reference group because their average wage consistently exceeds those of the other three racial groups throughout the period.

group, the Mincer coefficients suggest this is primarily with respect to education as opposed to potential experience.

Table 7: Oaxaca-Blinder decomposition estimates of inter-race variation in mean real hourly wages, by period, 2001 – 2023

	Period							
	2001:2 (1)		2011Q4 (2)		2019Q4 (3)		2023Q4 (4)	
Panel (a): African/Black								
Log real hourly wage	2.838***		3.374***		3.492***		3.468***	
	(0.011)		(0.010)		(0.009)		(0.010)	
Difference in log real hourly wage relative to White	-1.615***		-1.415***		-1.494***		-1.408***	
	(0.022)		(0.022)		(0.024)		(0.025)	
… Due to composition effect	-1.087***	67%	-0.720***	51%	-0.643***	43%	-0.697***	50%
(Δ productive characteristics)	(0.029)		(0.021)		(0.023)		(0.022)	
… Due to structure effect	-0.528***	33%	-0.695***	49%	-0.851***	57%	-0.711***	50%
(Δ returns or "discrimination")	(0.034)		(0.026)		(0.027)		(0.029)	
Panel (b): Coloured								
Log real hourly wage	3.307***		3.740***		3.778***		3.809***	
	(0.019)		(0.020)		(0.021)		(0.022)	
Difference in log real hourly wage relative to White	-1.146***		-1.049***		-1.208***		-1.066***	
	(0.028)		(0.028)		(0.031)		(0.032)	
… Due to composition effect	-0.834***	73%	-0.747***	71%	-0.585***	48%	-0.589***	55%
(Δ productive characteristics)	(0.030)		(0.032)		(0.036)		(0.034)	
… Due to structure effect	-0.312***	27%	-0.302***	29%	-0.623***	52%	-0.478***	45%
(Δ returns or "discrimination")	(0.035)		(0.039)		(0.040)		(0.038)	
Panel (c): Indian/Asian								
Log real hourly wage	3.951***		4.469***		4.328***		4.403***	
	(0.032)		(0.041)		(0.048)		(0.049)	
Difference in log real hourly wage relative to White	-0.502***		-0.321***		-0.657***		-0.472***	
	(0.038)		(0.046)		(0.053)		(0.054)	
… Due to composition effect	-0.348***	69%	-0.189***	59%	-0.256***	39%	-0.149***	32%
(Δ productive characteristics)	(0.048)		(0.029)		(0.045)		(0.045)	
… Due to structure effect	-0.154***	31%	-0.131***	41%	-0.401***	61%	-0.323***	68%
(Δ returns or "discrimination")	(0.056)		(0.043)		(0.053)		(0.046)	

Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. *Notes*: Sample restricted to individuals of working-age (15 – 64 years). Outcome = real hourly wages (log scale). Estimates weighted using sampling weights. Standard errors are presented in parentheses and adjusted for the complex survey design. Models include the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text. * $p<0.10$; ** $p<0.050$; *** $p<0.001$.

Over time, I observe a decreasing importance of the contribution of the composition effect, falling to 32 – 55 percent by 2023. This is likely a consequence of convergence of educational attainment across racial groups. Over the period, the difference in mean years of education relative to White individuals nearly halved, from up to 4.2 years (35 percent) to 2.4 years (19 percent). Conversely, and concerningly, I observe an increasing importance of the structure effect – that is, differences to the returns to these characteristics – in explaining wage differentials, regardless of the comparison group. For African/Black individuals, I estimate that the contribution of this component increased from 33 percent in 2001 to 50 percent in 2023 – an increase of over 52 percent. That for Coloured individuals grew more – by 67 percent – while that for Indian/Asian individuals more than doubled (119 percent), and consequently became the dominant component.

7 DISCUSSION

By making use of over 20 years of harmonized household survey microdata, including the longest uninterrupted series of reliable earnings data during the post-Apartheid period not available in the public domain, this study contributes to both the international and South African literatures on the evolving labour market returns to education and their drivers. In doing so, it sheds new light on the interconnected, long-term trends of educational attainment and the returns structure – both on average and across levels of education and racial population groups – with greater precision than was previously possible. By further considering variation across the wage distribution and decomposing wage differentials both within and between periods, I examine the implications of these trends on wage inequality.

South Africa has witnessed a substantial increase in educational attainment during the post-Apartheid period, with mean years of education increasing by 25 percent, driven by an expansion of both completed secondary and tertiary education alongside a contraction of primary education. Consequently, I estimate that educational attainment inequality has reduced by 40 percent. Consistent with global evidence (Montenegro and Patrinos, 2014; Patrinos and Psacharopoulos, 2020), strong, positive associations between education and employment and, conditional on employment, earnings have persisted. Concurrently, employment probabilities have reduced across the education distribution, especially for those with lower levels, while real wages have risen, especially for those with higher levels. This likely reflects worsening aggregate labour market conditions over the period alongside changes in the supply of and demand for workers across the education distribution. The consequence is not only a persistence of but an increasingly convex returns structure, with tertiary education being increasingly rewarded in the labour market. Hence, despite the much larger increase in educational attainment, I estimate that the average return to education, adjusted for selection into labour market participation and employment, has still increased by 8 percent. This suggests that the demand for higher-educated workers has outpaced the increase in supply, which is consistent with broader international trends and is plausibly explained by both supply- and demand-side factors such as increased coverage of lower levels of schooling and technological change favouring higher levels of education (Colclough et al., 2010; Psacharopoulos and Patrinos, 2018; Patrinos, 2019).

Changes to the average return to education, however, masks considerable heterogeneity. By race, the average return to education for African/Black individuals rose by a notable 25 percent, but despite this, their return remained the lowest throughout the period. This underscores the persistence of labour market disparities rooted in South Africa's historical inequalities (Mwabu and Schultz, 2000; Branson and Lam, 2021). By education level, in 2001 the returns to secondary education were highest while those to primary were lowest. This holds for most racial groups. Over time, this pattern shifted in favour of tertiary education. The returns to tertiary education grew substantially – tripling in size – while that of primary education shrunk by 40 percent. This pattern was evident for all racial groups, but to varying degrees. I also observe notable variation across the wage distribution. In 2001, rates of return were positive and relatively uniformly distributed, but became significantly stronger at the lower end and simultaneously weaker at the upper end over time. Average returns grew by 37 percent among the lowest-earning 40 percent of workers but shrunk by 26 percent among the higher-earning 40 percent of workers. By education level, in 2001 the returns to tertiary were highest towards the bottom end while those to secondary were dominant in the remainder. Over time, the returns to tertiary grew considerably at the bottom. By benefiting lower-wage workers more, these changes in the returns structure have placed downward pressure on wage inequality. While wage growth is evident across the distribution, it was concentrated towards the bottom where wages more than doubled in real terms. Consequently, wage inequality has decreased – the magnitude of which depends on the measure – but nevertheless remains high. This finding is consistent with Kerr's (2024) recent analysis of both other household survey and tax administrative data, and notably, stands in contrast to the rise in inequality suggested by studies which use the problematic, publicly available QLFS data. This highlights the consequences of using the latter data for any analysis of earnings in South Africa, likely due to the poor quality of imputations therein (Wittenberg, 2017; Kerr and Wittenberg, 2021; Köhler and Bhorat, 2023; Kerr, 2024).

The interplay between rising educational attainment and varying returns is highlighted in the decomposition analysis. Increases in both educational attainment and the returns to education explain the observed growth in mean wages, however educational attainment explains the most (58 percent). While the contribution of growth in returns is smaller, it has however doubled in importance over time. The dominance of the former, however, strongly depends on the point of the distribution. In contrast to the mean case, growth in returns primarily (65 – 71 percent) explains the observed wage growth among lower-wage workers, who experienced the most wage growth and whose returns became significantly stronger over time. These findings contrast those of Lam et al. (2012), who conclude that increased educational attainment reduced wage inequality while increased returns increased inequality.[27] By race, the increase in educational attainment remains the dominant force across most of the distribution for all groups apart from White individuals, which is likely explained by the latter's already relatively high education levels. At the bottom end, growth in returns continue to be the dominant force for both African/Black and Coloured individuals.

Considering inter-race wage differentials, I find that while most (67 – 73 percent) of the mean wage differentials relative to White individuals in 2001 were explained by differences in educational attainment and just up to one-third (27 – 33 percent) by differential returns to these characteristics, over time this pattern has changed. Concerningly, by 2023 the contribution of the former reduced to 32 – 55 percent, likely reflecting the convergence of educational attainment across groups, while that of the latter had risen to 45 – 68 percent. In other words, differences in wages have become

[27] This is potentially explained by Lam et al.'s (2012) use of data for an earlier period (1997 to 2007).

increasingly explained by the labour market's differential valuations of a given level of education across racial groups. While this is suggestive of increasing discrimination, it almost certainly, at least partially, reflects differences in the quality of education which can vary considerably for a given year of education (Salisbury, 2016; Burger and Teal, 2019; Filmer et al., 2020; Branson and Lam, 2021). Unfortunately, the data does not permit us to investigate this further, and as such, it serves as a crucial avenue for future research.

8 CONCLUSION

This paper provides an analysis of the evolution of the labour market returns to education and their drivers in post-Apartheid South Africa. Using over 20 years of harmonized household survey microdata from 2001 to 2023, including the longest uninterrupted series of reliable earnings data not available in the public domain, it sheds new light on the complex interplay between rising educational attainment and a varying returns structure, providing greater precision in understanding its implications for wage inequality than was previously possible.

I show that South Africa has experienced a substantial increase in educational attainment, driven by an expansion of completed secondary and tertiary education alongside a contraction of primary education. Together, these dynamics reduced educational attainment inequality by 40 percent over the period. Despite this large increase in the supply of higher-educated workers, the average return to education also increased, indicating that the demand for these workers has outpaced supply. On average, increases in both educational attainment and the return to education drove real wages upwards. While the former was dominant, the contribution of the latter has doubled over time. This pattern holds for most racial groups with White individuals serving as the exception, which is likely explained by their already relatively high levels of education. Average returns, however, mask notable heterogeneity. I show that the education-earnings relationship has become increasingly convex, driven by higher returns to tertiary education which have tripled in size over time. This pattern is evident for all racial groups to varying degrees. Across the wage distribution, rates of return, particularly for tertiary education, became significantly stronger for lower-wage workers. These higher returns primarily account for the growth in wages experienced by these workers, which more than doubled in real terms. Consequently, overall wage inequality has decreased but nevertheless remains high. Considering inter-race wage inequality, I show that while most is explained by differences in educational attainment for most groups, differential returns have become increasingly important for all groups. While this is suggestive of increasing discrimination, it likely at least partially reflects a growing importance of differences in education quality.

Overall, these findings highlight the dual roles of rising educational attainment and an increasingly convex returns structure in shaping wage dynamics and inequality in post-Apartheid South Africa. The results indicate that while increasing educational attainment among people of colour in the country represents real progress, these gains have become increasingly overshadowed by an unequal distribution of education's labour market benefits. This highlights the important role to be played by policy that addresses both labour market discrimination and disparities in education quality.

9 REFERENCE LIST

Allanson, P., Atkins, J. P. and Hinks, T. 2002. No end to the racial wage hierarchy in South Africa? *Review of Development Economics*, *6*(3): 442-459.

Ashenfelter, O. and Krueger, A. 1994. Estimates of the economic return to schooling from a new sample of twins. *The American Economic Review*, 84(5): 1157-1173.

Barouni, M. and Broecke, S. 2014. The Returns to Education in Africa: Some New Estimates. *The Journal of Development Studies*, 50(12): 1593-1613.

Bassier, I., Budlender, J., Zizzamia, R. and Jain, R. 2023. The labour market and poverty impacts of COVID-19 in South Africa. *South African Journal of Economics*, *91*(4): 419-445.

Becker, G. S. 1964. *Human capital*. New York: National Bureau of Economic Research.

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

Bhorat, H., Lilenstein, K., Oosthuizen, M. and Thornton, A. 2020. Wage polarization in a high-inequality emerging economy. WIDER Working Paper 2020/55. Helsinki: UNU-WIDER.

Bhorat, H., Stanwix, B. and Thornton, A. 2021. 'Changing Dynamics in the South African Labour Market,' in A. Oqubay, F. Tregenna, and I. Valodia (eds), *The Oxford Handbook of the South African Economy*. Oxford Academic.

Branson, N. and Lam, D. 2021. 'The Economics of Education in South Africa,' in A. Oqubay, F. Tregenna, and I. Valodia (eds), *The Oxford Handbook of the South African Economy*. Oxford Academic.

Branson, N. and Leibbrandt, N. 2013. Educational Attainment and Labour Market Outcomes in South Africa, 1994-2010. OECD Economics Department Working Papers No. 1022. Organisation for Economic Co-operation and Development.

Burger, R. and Teal, F. 2019. Convexity or Heterogeneity? Estimates of the Shape of the Earnings Profile. REDI3x3 Working Paper No. 53.

Card, D. 1999. 'The Causal Effect of Education on Earnings,' in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*. Elsevier.

Colclough, C., Kingdon, G. and Patrinos, H. A. 2010. The changing pattern of wage returns to education and its implications. *Development Policy Review*, *28*(6): 733-747.

Deming, D. J. 2022. Four facts about human capital. *Journal of Economic Perspectives*, *36*(3): 75-102.

Díaz Pabon, F. A., Leibbrandt, M., Ranchhod, V. and Savage, M. 2021. Piketty comes to South Africa. *The British Journal of Sociology*, 72(1): 106–124.

Fasih, T., Kingdon, G. G., Patrinos, H. A., Sakellariou, C. and Söderbom, M. 2012. Heterogeneous returns to education in the labor market. World Bank Policy Research Working Paper No. 6170. World Bank Group.

Filmer, D., Rogers, H., Angrist, N. and Sabarwal, S. 2020. Learning-adjusted years of schooling (LAYS): Defining a new macro measure of education. *Economics of Education Review*, *77*: 101971.

Finn, A., Leibbrandt, M. and Ranchhod, V. 2016. Patterns of persistence: Intergenerational mobility and education in South Africa. Tech. rep., SALDRU Working Paper Number 175/ NIDS Discussion Paper 2016/2. SALDRU, University of Cape Town.

Firpo, S., Fortin, N. M. and Lemieux, T. 2009. Unconditional quantile regressions. *Econometrica*, 77(3): 953–973.

Fortin, N., Lemieux, T. and Firpo, S. 2011. 'Decomposition Methods in Economics,' in D. Card and O. Ashenfelter (eds.), *Handbook of Labor Economics*. Elsevier.

Griliches, Z. 1970. Notes on the Role of Education in Production Functions and Growth Accounting. National Bureau of Economic Research.

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

Gunderson, M. and Oreopolous, P. 2020. 'Returns to education in developed countries,' in S. Brandley and C. Green (eds.), *The Economics of Education: A Comprehensive Overview*. Academic Press.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica*, 47(1): 153–161.

Hertz, T. 2003. Upward bias in the estimated returns to education: Evidence from South Africa. *The American Economic Review*, *93*(4): 1354-1368.

Kerr, A. 2024. Earnings and Earnings Inequality in South Africa: Evidence from Household Survey and Administrative Tax Microdata from 1993 to 2020. *Review of Income and Wealth*.

Kerr, A. and Wittenberg, M. 2019. Earnings and employment microdata in South Africa. WIDER Working Paper No. 2019/47. Helsinki: United Nations World Institute for Development Economic Research (UNU-WIDER).

Kerr, A. and Wittenberg, M. 2021. Union wage premia and wage inequality in South Africa. *Economic Modelling*, 97: 255-271.

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

Köhler, T. 2023. What I know about COVID-19 and the South African labour market. Research on Socio-Economic Policy (RESEP) Research Note.

Köhler, T., Bhorat, H., Hill, R. and Stanwix, B. 2022. 'The Short-Term Labor Market Effects of South Africa's National COVID-19 Lockdown,' in M. Qobo, M. Soko and N. Xenia Ngwenya (eds), *The Future of the South African Political Economy Post-COVID 19*. International Political Economy Series. Palgrave Macmillan, Cham.

Köhler, T. and Bhorat, H. 2023. Wages and wage inequality during the COVID-19 pandemic in South Africa. Development Policy Research Unit Working Paper 202308. DPRU, University of Cape Town.

Köhler, T., Bhorat, H. and Hill, R. 2023. The effect of wage subsidies on job retention in a developing country: Evidence from South Africa. WIDER Working Paper 2023/114. Helsinki: UNU-WIDER.

Lam, D., Leibbrandt, M., Garlick, J. and Branson, N. 2012. Education and Inequality: The South African Case. A Southern Africa Labour and Development Research Unit Working Paper Number 75. Cape Town: SALDRU, University of Cape Town.

Leibbrandt, M. and Díaz Pabón, F. 2022. 'Inequality in South Africa,' in A. Oqubay, F. Tregenna, and I. Valodia (eds), *The Oxford Handbook of the South African Economy*. Oxford Academic.

Leibbrandt, M., Finn, A. and Woolard, I. 2012. Describing and decomposing post-apartheid income inequality in South Africa. *Development Southern Africa*, 29(1): 19–34.

Leibbrandt, M., Ranchhod, V. and Green, P. 2020. 'South Africa: The top end, labour markets, fiscal redistribution, and the persistence of very high inequality,' in C. Gradín, M. Leibbrandt and F. Tarp (eds), *Inequality in the Developing World*. Oxford University Press.

Mincer, J. 1958. Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*(4): 281-302.

Mincer, J. 1962. On-the-job training: Costs, returns, and some implications. *Journal of Political Economy*, *70*(5, Part 2): 50-79.

Mincer, J. 1974. *Schooling, Experience and Earnings.* New York: National Bureau of Economic Research.

Montenegro, C. E. and Patrinos, H. A. 2014. Comparable estimates of returns to schooling around the world. World Bank Policy Research Working Paper No. 7020. World Bank Group.

Moses, E., Van der Berg, S. and Rich, K. 2017. How unequal education quality limits social mobility in South Africa. Synthesis report for the Programme to Support Pro-Poor Policy Development. Pretoria: Department of Monitoring, Planning, and Evaluation.

Mwabu, G. and Schultz, T. P. 2000. Wage premiums for education and location of South African workers, by gender and race. *Economic Development and Cultural Change*, *48*(2): 307-334.

Patrinos, H. A. 2019. The changing pattern of returns to education: What impact will this have on inequality? World Bank Policy Research Working Paper No. 8866. World Bank Gorup.

Patrinos, H. A. 2024. Estimating the return to schooling using the Mincer equation. IZA World of Labor.

Patrinos, H. A. and Psacharopoulos, G. 2020. 'Returns to education in developing countries,' in S. Brandley and C. Green (eds.), *The Economics of Education: A Comprehensive Overview*. Academic Press.

Psacharopoulos, G. 2024. Returns to education: A brief history and an assessment. *Education Economics*, 32(5): 561-565.

Psacharopoulos, G. and Patrinos, H. A. 2018. Returns to investment in education: A decennial review of the global literature. *Education Economics*, 26(5): 445-458,

Ranchhod, V. and Daniels, R. C. 2021. Labour market dynamics in South Africa at the onset of the COVID-19 pandemic. *South African Journal of Economics*, 89 (1): 44–62.

Rubin, D. B. 1987. *Multiple imputation for nonresponse in surveys*. Wiley: New York.

Salisbury, T. 2016. Education and inequality in South Africa: Returns to schooling in the post-apartheid era. *International Journal of Educational Development*, *46*: 43-52.

Schultz, T. P. 1988. 'Education investments and returns,' in H. Chenery and T.N. Srinivasan (eds.), *Handbook of Development Economics*. Elsevier.

Sherer, G. 2000. Intergroup economic inequality in South Africa: The post-apartheid era. *American Economic Review*, *90*(2): 317-321.

Spaull, N. 2013. Poverty & privilege: Primary school inequality in South Africa. *International Journal of Educational Development*, *33*(5): 436-447.

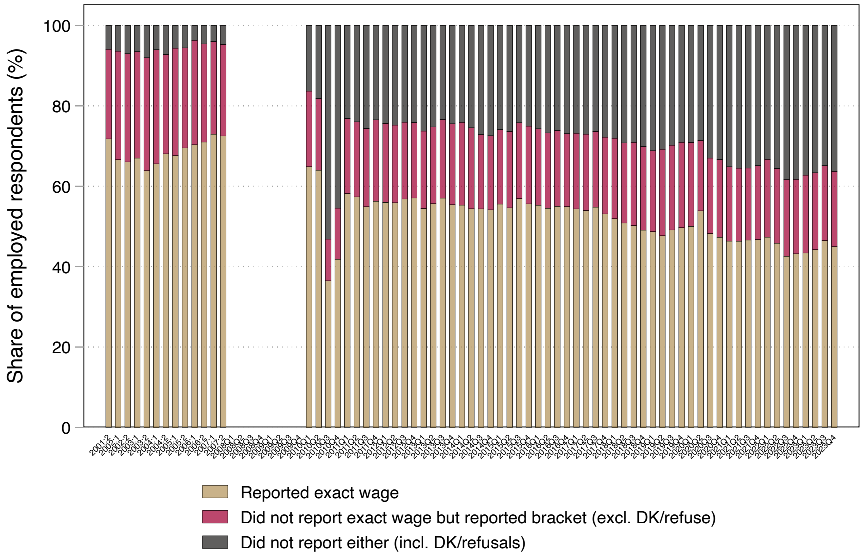
Van der Berg, S. 2014. Inequality, poverty and prospects for redistribution. *Development Southern Africa*, 31(2): 197-218.

Webber, D. 2014. Is the return to education the same for everybody? IZA World of Labor.

Wittenberg, M. 2017. Wages and wage inequality in South Africa 1994-2011: Part 2 – Inequality measurement and trends. *South African Journal of Economics*, 85(2): 298-318.

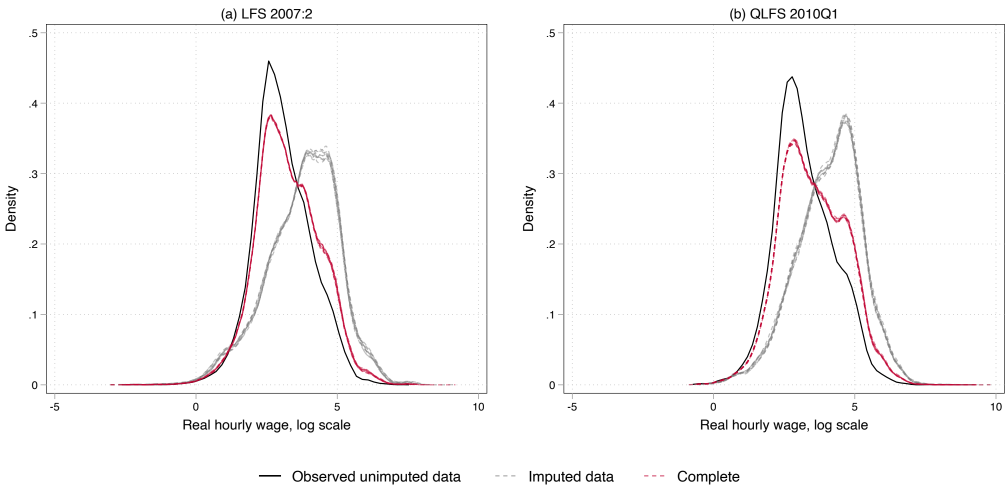
10 APPENDIX

Figure A1: Distribution of earnings responses in the LFS and QLFS, 2001 – 2023



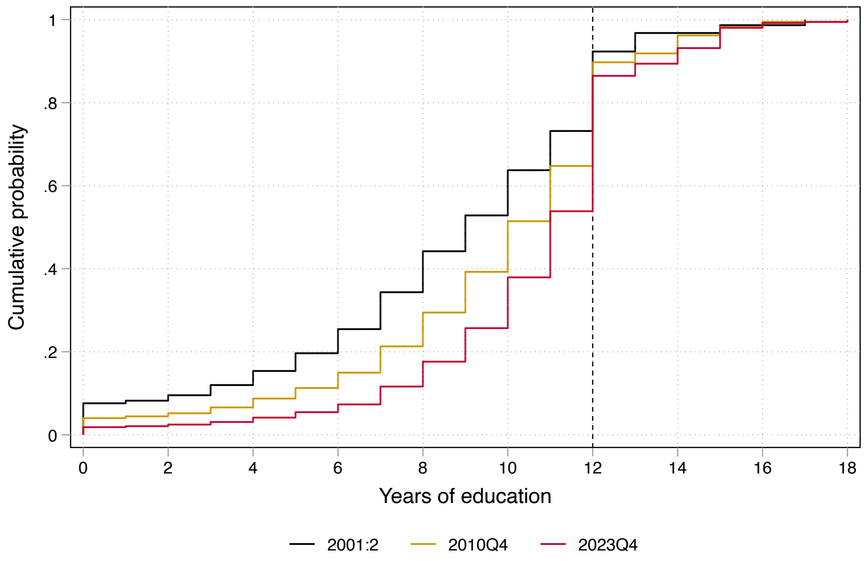
Source: Authors' own calculations using LFS 2001:2 – 2007:2, QLFS 2008Q1 – 2023Q4. *Notes*: Sample restricted to employed individuals of working-age (15 – 64 years). Estimates unweighted. DK = Don't know.

Figure A2: Diagnostic plot of observed, multiply imputed, and complete real hourly wage distributions in the LFS and QLFS



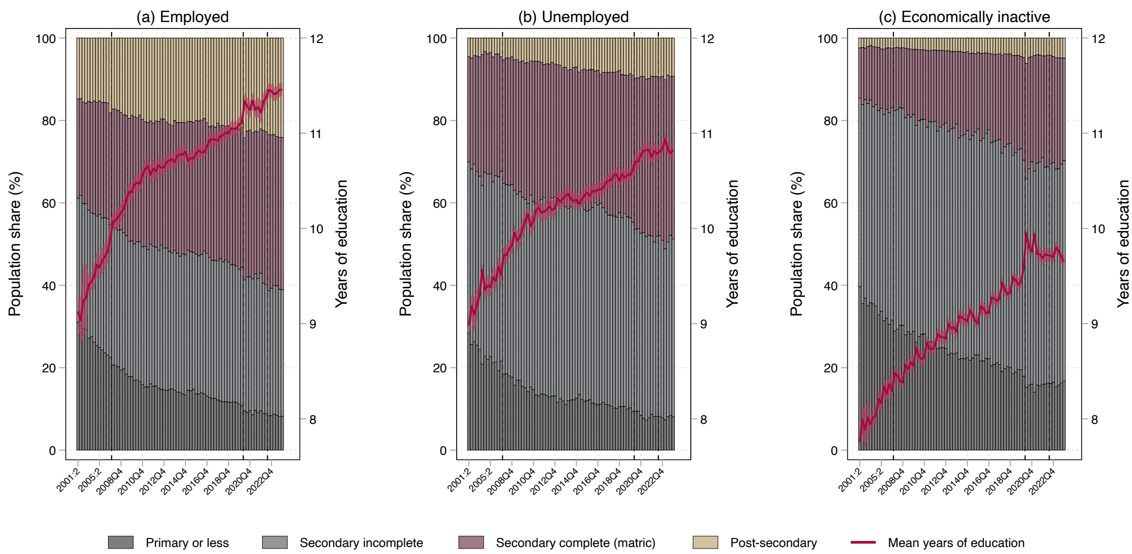
Source: Authors' own calculations using LFS 2007:1; QLFS 2010Q1. *Notes*: Sample restricted to employed individuals of working-age (15 – 64 years). Estimates unweighted. Wages expressed in June 2024 Rands. Observed = reported wage data only; Imputed = imputed wage data only; Completed = combination of observed and imputed data.

Figure A3: Cumulative distribution functions of mean years of education over time



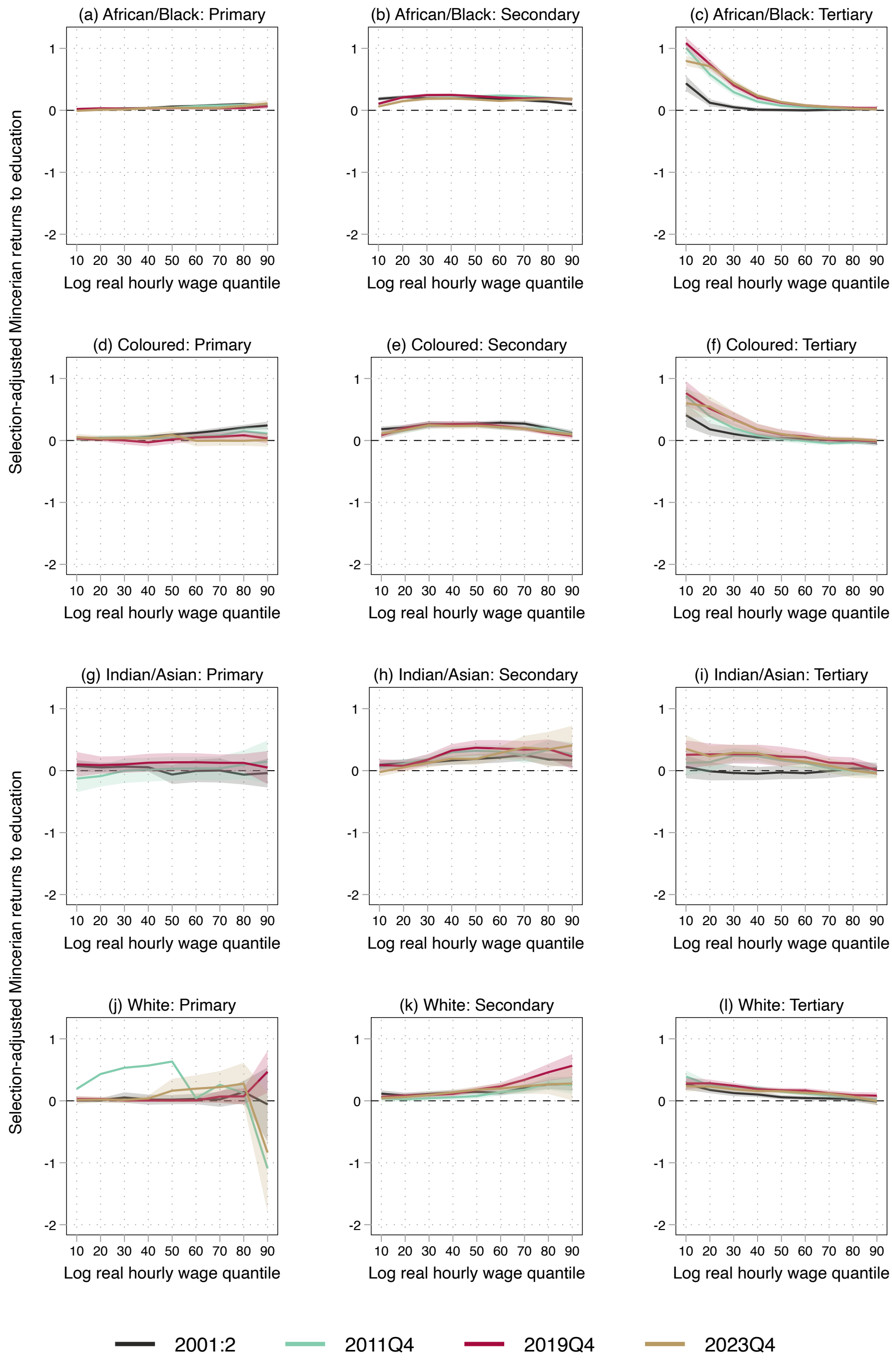
Source: Authors' own calculations using LFS 2001:2 – 2007:2, QLFS 2008Q1 – 2023Q4. Notes: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights.

Figure A4: Distribution of educational attainment, by labour market status, 2001 – 2023



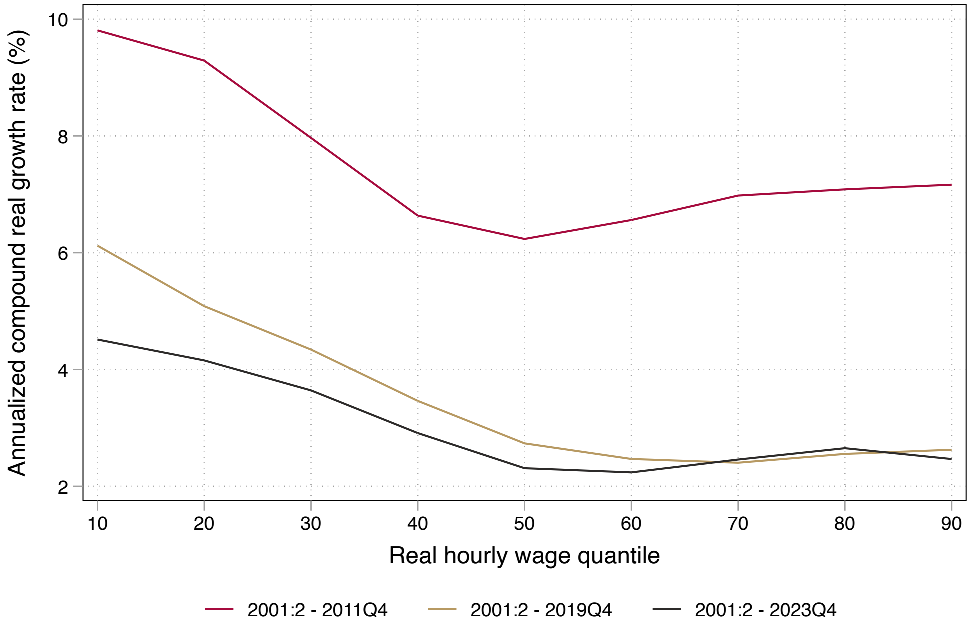
Source: Authors' own calculations using LFS 2001:2 – 2007:2, QLFS 2008Q1 – 2023Q4. *Notes*: Narrow labour market definitions used. Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Vertical lines represent, in order, the change of the survey instrument, the onset of the COVID-19 pandemic in March 2020, and the repeal of all remaining COVID-19 pandemic restrictions in June 2022.

Figure A5: Recentered Influence Function estimates of the selection-adjusted Mincerian returns to education across the wage distribution, by level of education and race, 2001-2023



Source: Authors' own calculations using LFS 2001:2; QLFS 2011Q4; 2019Q4; 2023Q4. Notes: Sample restricted to individuals of working-age (15 – 64 years). Estimates weighted using sampling weights. Standard errors are adjusted for the complex survey design. Shaded area represents 95 percent confidence intervals. Confidence interval for 2011Q4 in panel (j) omitted. Selection-adjusted estimates presented using the conventional Mincerian model specification but additionally including the inverse Mills ratio estimated using Heckman's two-stage estimation procedure described in the text.

Figure A6: Growth incidence curves of real hourly wages, by period



Source: Authors' own calculations using LFS 2001:2 – 2007:2, QLFS 2010Q1 – 2023Q4. *Notes*: Sample restricted to employed individuals of working-age (15 – 64 years). Estimates weighted using sampling weights.

Contact Information

Email:
carinebrunsdon@sun.ac.za
Phone:
(+27) 21 808 2024
Website:
www.resep.sun.ac.za

Address:
Research on Socio-Economic
Policy (RESEP)
Department of Economics
Matieland
7602

Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

forward together
sonke siya phambili
saam vorentoe

RESEP
Research on Socio-Economic Policy