WIDER Working Paper 2018/83

The evolution and determination of earnings inequality in post-apartheid South Africa

Arden Finn and Murray Leibbrandt*

August 2018

Abstract: In this paper we identify some of the drivers of changes in the distribution of earnings and earnings inequality in the South African labour market between 2000 and 2014. Although the overall level of earnings inequality between 2000 and 2011 was high and relatively stable, there were nonetheless some interesting shifts in the factors generating inequality. The earnings data from mid-2012 to 2014, however, show a steep increase in inequality. It is difficult to determine how much of this is a 'real' change, and how much is driven by other factors such as measurement error and changes in data collection and processing. For this reason, all results are presented in a 2000–14, 2000–11, and 2011–14 format. We use RIF regressions to decompose changes in average earnings, as well as changes in the Gini coefficient and different percentile ratios. Our main finding is that changes in the returns to education and changes in the returns to potential experience were the most important determinants of changes in inequality, with the former generally being inequality-enhancing, and the latter inequality-reducing.

Keywords: RIF decomposition, South Africa, wage inequality
JEL classification: D31, J31

Acknowledgements: Arden Finn and Murray Leibbrandt acknowledge support from UNU-WIDER's 'Inequality in the Giants' project, the Research Chairs Initiative of the South African National Research Foundation, and the South African Department of Science and Technology, who fund Leibbrandt's Chair in Poverty and Inequality Research.

This paper was produced when Arden Finn was a postdoctoral researcher at SALDRU, University of Cape Town. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors, and do not necessarily represent the views of the World Bank, its Executive Directors, or the governments of the countries they represent.

* Southern Africa Labour and Development Research Unit, University of Cape Town, Cape Town, South Africa; corresponding author: murray.leibbrandt@uct.ac.za.

This study has been prepared within the UNU-WIDER project on 'Inequality in the Giants'.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-525-1 https://doi.org/10.35188/UNU-WIDER/2018/525-1

Typescript prepared by Gary Smith.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

1 Introduction

The consistently high level of wage inequality in South Africa is a phenomenon that has attracted a great deal of attention over the last decade and a half. Although there have been significant changes in the composition of the labour force in post-apartheid South Africa—among other things, workers are more educated and female labour market participation has risen—the overall level of wage inequality, as measured in high-frequency labour force surveys, has not decreased. This is important both because the structure of the labour market seems to be replicating high levels of inequality year after year, and also because of the central role of wage inequality in driving overall household income inequality in the country.

In this paper we identify and unpack some of the key drivers of inequality in the labour market in South Africa. In doing so we highlight the different roles that endowments and the returns to those endowments play in generating persistent inequality. This means that if we are comparing inequality in 2000 and 2014, for example, we need to be able to attribute the relative contributions of, for example, the level of education of workers as distinct from the returns to education of those same workers. We do this by making use of recentred influence function (RIF) regression analysis and generalized Oaxaca–Blinder (OB) decompositions in order to better understand how demographics and human capital determine wages differently at different points of the earnings distribution.

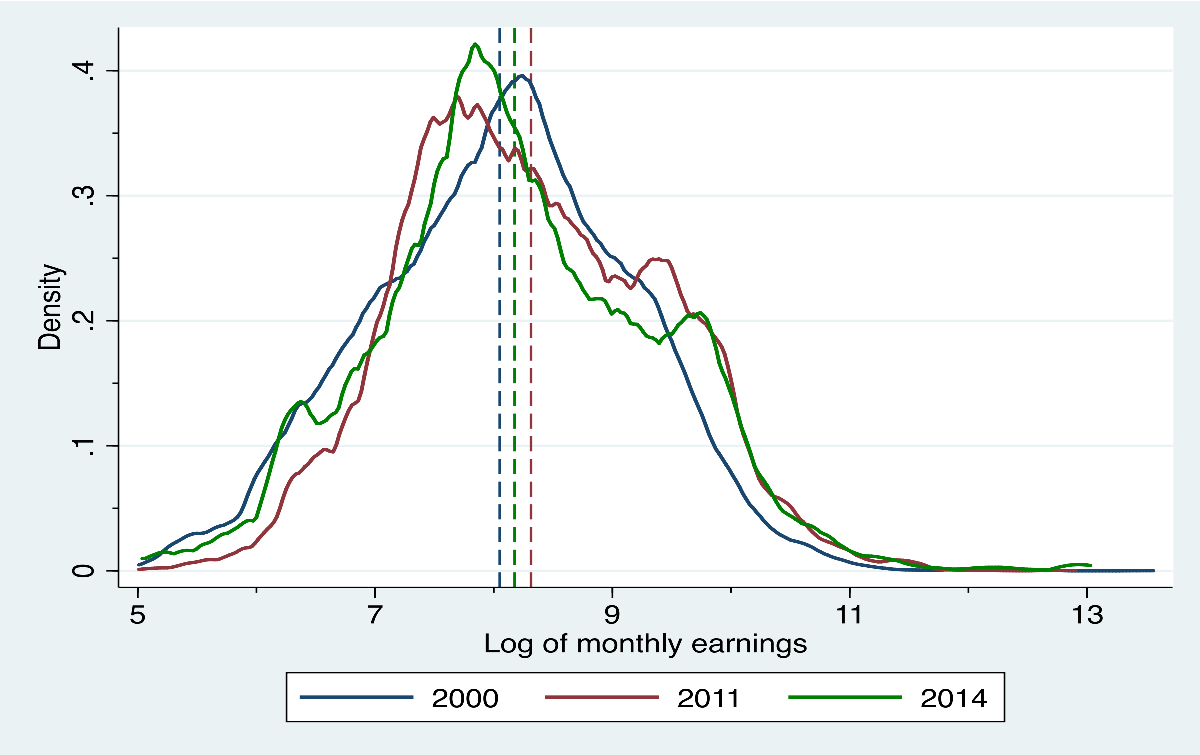
One of the major challenges facing our analysis is a methodological change in the construction of earnings data from the third quarter of 2012 onwards, which was implemented by Statistics South Africa (StatsSA). As we show later, this change caused significant shifts in the cross-sectional level of earnings inequality in all subsequent periods, and is something that we will return to in detail when describing the data. Because of this, we divide our analysis into three windows: 2000–14, 2000–11, and 2011–14.

Figure 1 presents the first view of changes in the distribution of (log) earnings for three different years—2000, 2011, and 2014. There are a number of notable differences between the distributions of 2000 and 2011. The first is that there was a strong rightward shift at the bottom of the distribution, suggesting that wage growth was robust for low earners during that period. In addition, there was a rightward shift in the density of the distribution of log wages between 8.5 and 10. The mean of the log distribution rose, as indicated by the dashed lines, and the overall mean rose from R5,710 to R7,420. As noted by Wittenberg (2017a), from 2000 to the end of 2011, the annual increases in the mean and the median were 2.45 per cent and 1.26 per cent respectively, suggesting that overall increases accrued disproportionately to those with higher wages to begin with.

The extent of the challenge posed by the change in methodology in the construction of earnings data is apparent when comparing the 2011 and 2014 distributions (the red and green lines, respectively). The leftward shift at the lower end of the distribution seemingly wipes out almost all the gains made between 2000 and 2011, and there is also a significant compression in the middle of the distribution of log earnings. For log earnings of approximately 10 and above, the 2011 and 2014 distributions are very similar. In short, what this figure suggests is that, methodological issues aside, one would expect to see a sharp increase in earnings inequality from 2011 to 2014.

The changes from 2000 to 2011 to 2014 are robust to whether we consider all workers (as in Figure 1), or whether we restrict our analysis to full-time workers only (for various reasonable definitions of full-time work). In general, measured inequality is slightly lower if we consider only full-time workers, but this restriction is not enough to change any of the qualitative conclusions about shifts in the earnings distributions, and changes in inequality. How much of this is a ‘real’ change, and how much is driven by StatsSA’s methodological shift is unclear.

Figure 1: Kernel densities of log wages in South Africa



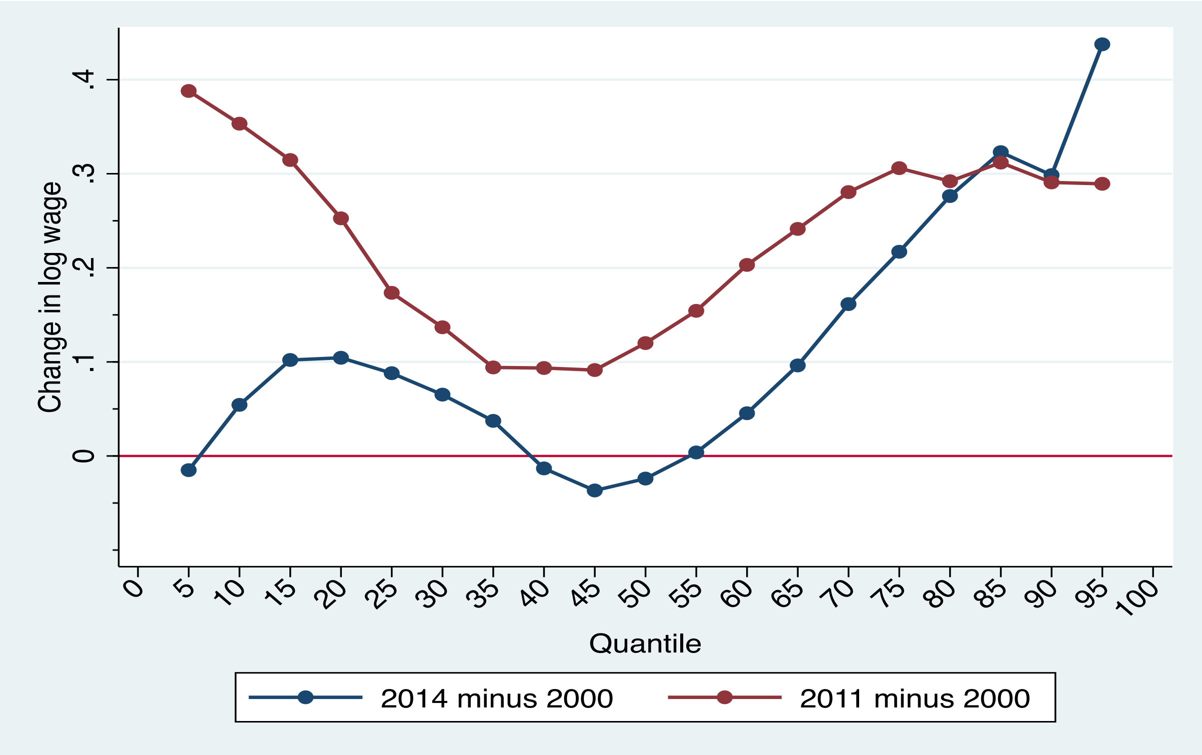
Source: authors' calculations using PALMS V3.1 dataset.

Figure 2 complements our understanding of the kernel densities in Figure 1 by plotting the change in log wages over the percentiles of the earnings distribution. The red line shows the log earnings in 2011 minus the log earnings in 2000 for various percentiles. The blue line does the same, but now for 2014 minus 2000. The pattern of wage growth between 2000 and 2011 is clear—growth was robust at the lower and upper ends of the distribution, and was a lot weaker for those who earned around the middle of the distribution. The 2000 to 2014 difference, however, is very different, and shows very low growth at the bottom of the distribution, declines in real earnings around the median, and increasingly strong growth towards the top of the distribution. To reiterate, it is not clear how much of this difference is 'real' and how much is a result of methodological changes, but the impact on measured earnings inequality is obvious, as will be shown in Figure 3.

Figures 1 and 2 suggest that earnings inequality must have increased between 2000 and 2014. Figure 3 shows the trend in inequality from 2000 to 2014 by plotting Gini coefficients for each year of available earnings data. This is simply an updated version of what appears in Wittenberg (2017a) and Finn (2015), as earnings data from 2012 to 2014 have been added.

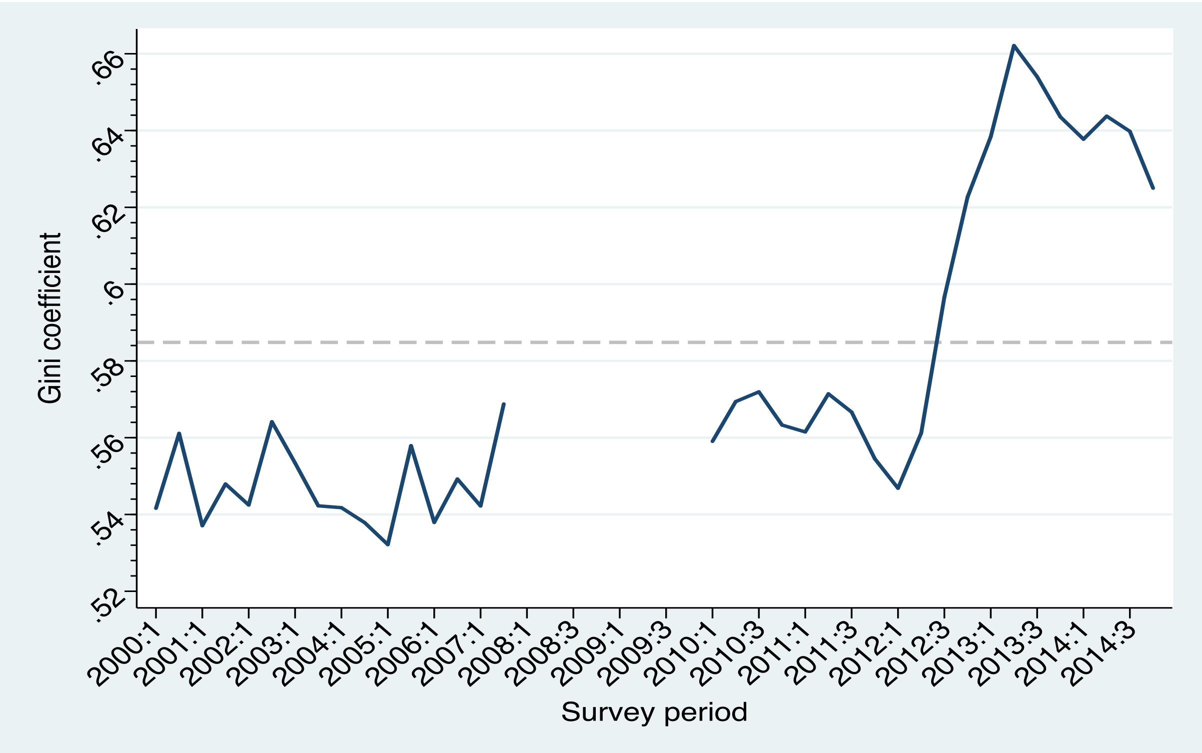
The Gini coefficient was generally within the band of 0.54 to 0.57 from the first quarter of 2000 until the first quarter of 2012. In fact, if taken as two snapshots, one might conclude that there was only a small increase in wage inequality in South Africa between the start of 2000 and the end of 2011. This relatively small change in the Gini coefficient may partially be a function of the construction of the measure itself—Atkinson (1970) shows that the Gini coefficient is most sensitive to changes in the middle of the distribution, which is precisely where changes were the most muted in South Africa between 2000 and 2011. The sharp rise in the Gini coefficient from 0.561 in the second quarter of 2012 to 0.662 in the second quarter of 2013 shows some, if not all, of the effect of StatsSA's change in methodology in the construction of the earnings data. The Gini coefficient tends to be a slow-moving statistic, and it is highly unlikely that it would experience a real change of 18 per cent over 12 months. The fact that the Gini coefficient is between 0.63 and 0.66 for all eight quarters of 2013–14 shows that the change in methodology has caused a substantial and lasting increase on the measurement of earnings inequality in the country. In fact, the increase is so substantial that the dashed horizontal line showing the average Gini coefficient for the full period is always above every single period's Gini coefficient until the third quarter of 2012, which is exactly when the change in methodology occurred.

Figure 2: Changes in log wage over the distribution of earnings



Source: authors' calculations using PALMS V3.1 dataset.

Figure 3: The Gini coefficient of earnings over time



Source: authors' calculations using PALMS V3.1 dataset.

The three panels of Figure 4 complement what we know about the trend in the Gini coefficient by presenting trends in the 90/10, 90/50 and 50/10 ratios from 2000 to 2014. If we focus just on the 2000–11 period, we see evidence of a slight decrease in the ratio of earnings at the 90th percentile to earnings at the 10th percentile. The relatively sluggish growth in the middle of the distribution is clear in the second panel, which shows the top of the distribution steadily pulling away from the median earner. This effect is also evident in the third panel, which shows, for the 2000–11 period, that the ratio of the 50th percentile to the 10th percentile fell.

Figure 4: Three percentile ratios over time

Source: authors' calculations using PALMS V3.1 dataset.

Extending our analysis from 2011 to 2014 shows some quite dramatic differences. The ratio between earners at the 90th percentile and earners at the 10th percentile increased from 15 in the first quarter of 2012 to 25 at the end of 2014. The explosion in this ratio is driven by both the collapse in earnings at the bottom of the distribution and the rapid growth at the top, rather than by only one of these factors. Once again, how much of this is driven by 'real' changes versus how much is a vestige of changing methodology is a debatable point, though it seems highly unlikely that the ratio could change so much over such a short period of time if methodological issues do not shoulder most of the blame.

Something that was consistent between the two curves in Figure 2 was the fact that growth in earnings was at its weakest in the middle of the distribution. This is again highlighted in the second panel of Figure 4 in which the 90/50 ratio is shown to have steadily increased over the 2000–11 and 2011–14 periods. Likewise, the third panel shows that the ratio of the median earner declined relative to the earner in the 10th percentile in the 2000–11 period. This 50/10 ratio, however, increased steadily from 2012 onwards, and this was mainly because of the leftward shift at the bottom of the 2014 distribution, as was shown in Figure 1.

In summary, earnings inequality seems to have been relatively stable in the 2000–11 period. However, the factors driving the generation of this inequality may have changed over the period—it is possible that very different underlying processes can produce the same level of inequality. A methodological change in the generation of the earnings variable by StatsSA in the third quarter of 2012 led to a sharp rise in measured inequality for the period ending in the fourth quarter of 2014. Assessing how much of this change is real and how much is an artefact driven by this change is not possible at this point in time. Given this, all analysis of the post-2012 Q3 period must be treated with caution.

The rest of this paper is structured as follows. Section 2 outlines some of the research on post-apartheid earnings inequality, with special attention to the work using RIF techniques. Section 3 describes the RIF methodology which forms the backbone of the analysis of this paper, while Section 4 presents a detailed overview of the dataset being used, as well as descriptive statistics. Section 5 reports a range of decompositions based on the RIF regressions for the Gini coefficient, and at various points of the earnings distribution. Section 6 provides some concluding remarks.

2 Background literature on South Africa

There is a large body of literature that covers the measurement of earnings and household income inequality in South Africa in the post-apartheid era. Much of this literature has focused on decomposing inequality into various subgroups (for example by race or by income source), but less is known about what drove changes in inequality over this period.

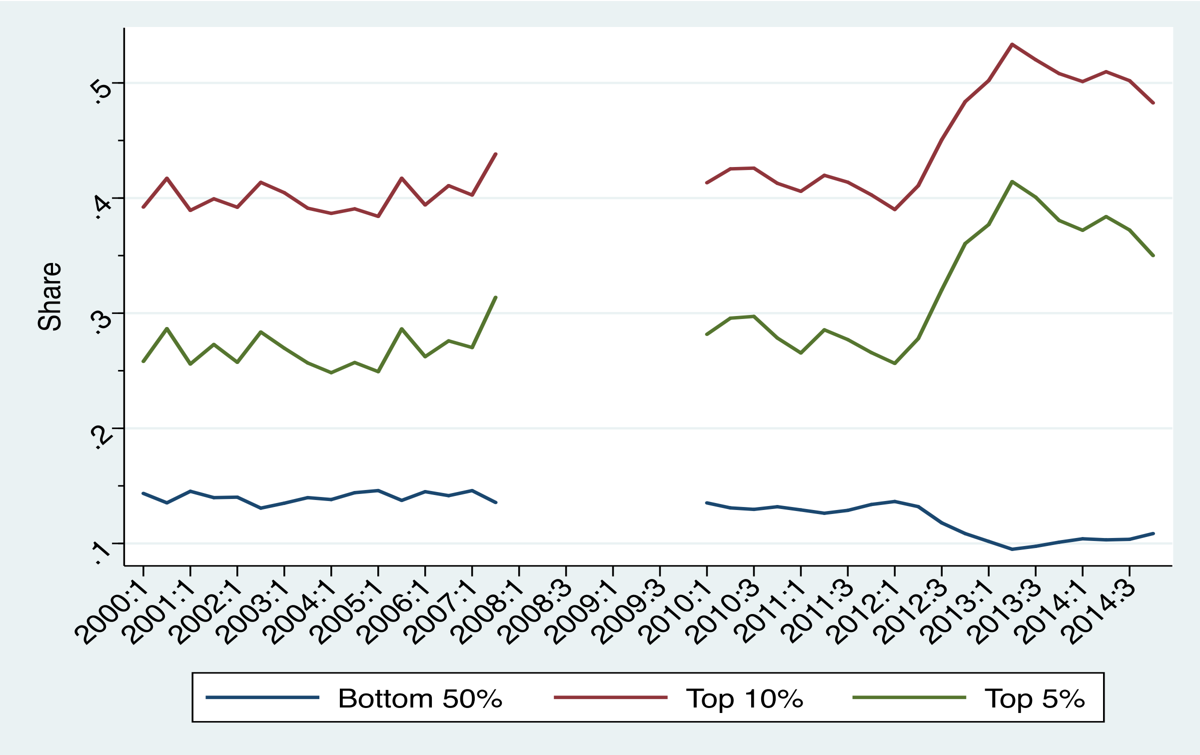
The most comprehensive recent studies on trends in earnings inequality in South Africa can be found in the work of Wittenberg (2017a,b). The first of these papers studies the entire 1994–2011 period, with a focus on employees only (meaning that the self-employed are excluded from the analysis). The broad conclusions reflect what was shown in the previous figures in this paper—the distribution of earnings below the median became more compressed over time, while the top of the distribution moved away from the median.

Wittenberg (2017b) shows that wage inequality increased between 1994 and 2011 according to most of the standard inequality measures, and that Gini coefficients post-March 2007 are noticeably higher than they were between 2000 and 2006. This paper takes measurement issues and changes in the survey instrument and methodology very seriously, and offers the reader the most consistent possible interpretation of trends, given the available data. A decomposition of earnings inequality by race and education shows that within-race inequality accounts for an increasing proportion of total earnings inequality in South Africa, and also that there has been some wage compression among the least skilled workers in the labour force.

An important set of graphs presented by Wittenberg (2017a) show the various wage shares going to the top 10 per cent, 5 per cent, and 1 per cent, and the bottom 50 per cent of earners over the full period. These figures reinforce the finding that the top part of the distribution gained shares in the post-apartheid era. These findings are reproduced and extended in Figure 5, which plots the shares of the bottom 50 per cent, as well as the top 10 per cent and the top 5 per cent from 2000 to the final quarter of 2014. The figure shows that up to the middle of 2012, the shares going to each of these groups were relatively consistent—about 15 per cent to the bottom half, 28 per cent to the top 5 per cent, and 41 per cent to the top 10 per cent of earners. The dramatic break in the trend that was shown in previous figures is shown here as well, with the share of the bottom half seeming to collapse from late 2012 onwards, at the same time as the share of the top 5 per cent increased to around 40 per cent, and the top 10 per cent to around 52 per cent.

The Wittenberg (2017a,b) papers give a broad outline of the trends in inequality in the country. There are some recent studies which complement these analyses by applying a RIF methodology in an attempt to answer more specific questions about the nature of earnings inequality in South Africa. Naidoo et al. (2014) investigate racial discrimination in the labour market at three points in time—1994, 2001, and 2011—and run RIF regressions in order to model the impact of discrimination over the earnings distribution. This allows them to build on the pre-existing discrimination literature, which was focused on the mean, rather than specific percentiles, of the wage distribution. The authors break down the wage gaps into an explained component and an unexplained component, and use changes in the unexplained com-

Figure 5: Total wage shares



Source: authors' calculations using PALMS V3.1 dataset.

ponent as proxies for changes in discrimination. The broad findings are that discrimination decreased over the period under study, and that the greatest decrease took place at the top of the wage distribution. These findings build on the work of Bhorat and Goga (2013), which focuses on gender discrimination in the wages of Africans in the labour market. The authors use the 2007 Labour Force Survey (LFS) in their analysis, and find that the gender wage gap is largest for low earners, and that a 'pure' discrimination effect accounts for most of the measured difference in earnings from the 20th percentile to the median worker.

Kwenda and Ntuli (2015) move the focus off demographic differences and use RIF regression techniques to unpack the underlying driver of the public–private wage gap in South Africa. The authors use PALMS data from 2000 to 2007 and find that the wage gap displays an inverted-U shape over the distribution of earnings. The gender pay gap—to the disadvantage of women—tends to be larger in the private sector, while the wage gap at the bottom end of the distribution is driven almost entirely by the inferior endowments of private sector workers.

In research that covers the labour market from 2001 to 2011, Hosking (2016) acknowledges that although earnings inequality was relatively stable, the nature of that inequality changed substantially. The author reiterates the point about the hollowing out of the wage distribution over the period, with strong relative growth at the two ends of the distribution, but very sluggish growth in the middle. The study uses measures of task intensity to capture the role of technological change on the distribution of wages. RIF regressions decompose earnings differentials into composition and structure effects, and it is found that technological change, minimum wage coverage, and changing returns to tertiary education were the key contributors to the wage structure component of the decomposition.

3 Background to estimating RIF regressions for South Africa

There are two main technical challenges facing our attempts to decompose earnings inequality over time. The first of these is how to attribute the role of different variables at different points of the earnings distribution in determining inequality. The second is how to divide each variable's contribution into a 'composition' effect and a 'wage structure' effect at each percentile in the distribution of earnings. Essentially, we would like to generalize the Oaxaca–Blinder decomposition of the mean to any unconditional quantile of the distribution. This is important because the distributions of the characteristics of workers have changed over time, as have the partial correlations of these characteristics to earnings. The estimating methodology best suited to addressing these challenges is RIF regression, which is introduced to the literature by Firpo et al. (2009) and then expanded on with specific attention to functionals of income/earnings distributions (for example, percentiles and measures of dispersion such as the Gini coefficient or the 90/10 ratio) in Fortin et al. (2011). Our exposition of how to relate changes in the distributions of covariates and their premiums to changes in the distribution of earnings is based on Ferreira et al. (2017), and follows below.

If a researcher is interested in comparing earnings in two time periods, $t = 1$ and $t = 2$, then, following Oaxaca (1973) and Blinder (1973), if one assumes that earnings are linear and separable in observables and unobservables, then:

$$Y_t = X_t'\beta_t + \varepsilon_t, \quad \text{for} \quad t = 1,2 \tag{1}$$

where X_t is a vector of covariates, and β is the vector of parameters associated with the coefficients. A model of pooled earnings is expressed as follows:

$$Y = X'\beta + \varepsilon$$

Let the indicator variable D take a value of 1 if $t = 2$ and 0 if $t = 1$. Invoking the usual assumption about orthogonal observables and unobservables, the mean earnings gap can be expressed as:

$$\begin{aligned}
&E[Y|D_t = 1] - E[Y|D_t = 0] = \\
&E[X|D_t = 1]'(\beta_2 - \beta) + E[X|D_t = 0]'(\beta - \beta_1) + \\
&(E[X|D_t = 1] - E[X|D_t = 0])'\beta
\end{aligned} \tag{2}$$

This can be estimated as:

$$\bar{Y}_2 - \bar{Y}_1 = \left[\bar{X}_2'\left(\hat{\beta}_2 - \hat{\beta}\right) + \bar{X}_1'\left(\hat{\beta} - \hat{\beta}_1\right)\right] + \left(\bar{X}_2' - \bar{X}_1'\right)\hat{\beta} = \hat{\Delta}_S^{\mu} + \hat{\Delta}_X^{\mu} \tag{3}$$

where the horizontal bars represent sample averages. The first term in this equation is the estimated earnings structure effect, $\hat{\Delta}_S^{\mu} = \bar{X}_2'\left(\hat{\beta}_2 - \hat{\beta}\right) + \bar{X}_1'\left(\hat{\beta} - \hat{\beta}_1\right)$, while the second term is the estimated composition effect, $\hat{\Delta}_X^{\mu} = (\bar{X}_2' - \bar{X}_1')\hat{\beta}$. For the remainder of this paper, the first term is synonymous with the role played by changing premiums in changes in earnings, while the second term is synonymous with the role played by changing distributions of covariates in changes in earnings.

The two effects can be expressed as the sum over the covariates used in the estimation:

$$\hat{\Delta}_S^{\mu} = \sum_{j=1}^{k} \bar{X}_{2,j}'\left(\hat{\beta}_{2,j} - \hat{\beta}_j\right) + \bar{X}_{1,j}'\left(\hat{\beta}_j - \hat{\beta}_{1,j}\right) \tag{4}$$

$$\hat{\Delta}_X^{\mu} = \sum_{j=1}^{k} \left(\bar{X}_{2,j}' - \bar{X}_{1,j}'\right)\hat{\beta}_j \tag{5}$$

where covariates are indexed by j, and $\left(\bar{X}_{2,j}' - \bar{X}_{1,j}'\right)\hat{\beta}_j$ and $\bar{X}_{2,j}'\left(\hat{\beta}_{2,j} - \hat{\beta}_j\right) + \bar{X}_{1,j}'\left(\hat{\beta}_j - \hat{\beta}_{1,j}\right)$ are the contributions of covariate j to the composition and wage structure effects, respectively.

The exposition above is specific to the decomposition of average earnings. Generalizing this to enable an analogous decomposition for any part of the earnings distribution, as well as for functionals such as percentile ratios or the Gini coefficient, means adopting the RIF framework (sometimes also called the unconditional quantile regression (UQR) framework) of Fortin et al. (2011). The key insight underlying RIF estimation is that as long as a statistic admits an influence function, this influence function can approximate non-linear functionals of the distribution of earnings by an expectation.

Estimating RIF regressions works the same way as estimating Oaxaca–Blinder decompositions, except for the fact that the dependent variable becomes the recentred influence function of the statistic of interest. If v is a functional of the distribution of earnings, then the structure and composition effects of that functional (for example the mean, the Gini coefficient, and any percentile) are expressed as sums over the covariates as follows:

$$\hat{\Delta}_S^v = \sum_{j=1}^{k} \hat{\Delta}_{S,j}^v = \sum_{j=1}^{k} \bar{X}'_{2,j}\left(\hat{\beta}_{2,j}^v - \hat{\beta}_j^v\right) + \bar{X}'_{1,j}\left(\hat{\beta}_j^v - \hat{\beta}_{1,j}^v\right) \tag{6}$$

$$\hat{\Delta}_X^v = \sum_{j=1}^{k} \hat{\Delta}_{X,j}^v = \sum_{j=1}^{k} \left(\bar{X}'_{2,j} - \bar{X}'_{1,j}\right)\hat{\beta}_j^v \tag{7}$$

where $\hat{\beta}_{t,j}^v$ and $\hat{\beta}_j^v$ are the coefficients of covariate j when the RIF is regressed on X in period t, and for the pooled periods respectively. As noted by Ferreira et al. (2017), when the functional is the mean, then the RIF simply becomes Y, and the RIF regression is identical to the usual Oaxaca–Blinder decomposition.

When we implement RIF regressions in this paper, a number of our covariates are categorical (for example, race and occupational sector). Oaxaca–Blinder, and therefore RIF decompositions, are not invariant to the choice of which category serves as the base. In all of our estimations we have chosen the 'best performing' category as the base for each categorical variable (for example, those with postsecondary education, and those working in the mining sector).[1]

In this paper we examine the roles played by eight different possible explanations for changes in earnings inequality in South Africa. They are: education, potential experience, unionization, formality, race, gender, geographic location, and sector of employment. We separate out the roles played by changes in the distributions of these variables, and in the changes in the returns associated with them.

The next section describes the dataset used in running RIF decompositions on South African earnings, before presenting some descriptive statistics and taking a closer look at some of the covariates used in the decompositions.

4 Data and descriptive statistics

This paper uses version 3.1 of the Post-apartheid Labour Market Series (PALMS) dataset (Kerr et al., 2016) for the descriptive statistics and the analysis. Any presentation using South African labour market data from different points in time is going to be subject to comparability concerns, and PALMS offers by far the most comprehensive and harmonized labour market data available in South Africa since 1994. Wittenberg (2017a,b) provides a very clear and comprehensive discussion of what assumptions

[1] Results with different base categories are available from the authors.

are needed in order to make defensible comparisons over time. These two papers also present trends in earnings, inequality, and the composition of the South African labour force from 1994 to 2012.[2]

We use PALMS data from the first quarter of 2000 to the final quarter of 2014. The harmonized data are drawn from the LFS from 2000 to 2007, and the Quarterly Labour Force Surveys (QLFSs) from 2008 to 2014.

We restrict our analysis to wage earners only (meaning that the self-employed are not part of the sample), and to those who are aged between 18 and 65. Earnings are deflated to their real 2015 equivalents. We make use of the weights provided in the PALMS dataset which correct for bracket responses, and which pay careful attention to changes in the population distribution, so as to allow for comparisons of changes over time. PALMS also flags outliers in the earnings distribution using a studentized regression approach (see Wittenberg, 2017a for more details), and we drop these from our analysis.

The introductory section of this paper discussed the concerns we have with measures of inequality in the period from the third quarter of 2012 onwards. One methodological change that took place at this stage of the data production was how data were imputed. As noted in the PALMS V3.1 user manual (Kerr and Wittenberg, 2016) and by Kerr and Wittenberg (2017), in the 2000 to 2012 Q2 QLFSs, both refusals and bracket responses were imputed. In fact, over 99 per cent of those who were employed have earnings data. However, from the third quarter of 2012 onwards there is only partial imputation—bracket responses are given rand amounts, but complete refusals are not imputed. We are unable to quantify the full impact of changing methodology on the earnings distribution and measures of inequality because we cannot tell which observations have been imputed, and, if imputed, which imputation method has been used. Nonetheless, it appears that even though inequality measures from 2012 Q3 should be treated with caution, inequality has increased since 2000 in South Africa, and that this has been driven in large part by a stagnation of income growth in the middle of the distribution, especially compared to growth in the top 10 per cent.

4.1 Descriptive statistics of key variables

Table 1 presents summary statistics of the key variables that are used in our decompositions for each of the years 2000, 2011, and 2014. Wage earners in 2011 and 2014 had about two more years of education than those in 2000, on average. This translated into a slightly lower level of potential experience—defined as age minus years of education minus six. Unionization coverage was about 5 percentage points down in 2014 compared to 2000, and the proportion of wage earners in formal jobs was almost the same in the first period and the last period, despite experiencing a slight dip in 2011.

The proportion of earners who were African rose by about 3 percentage points between 2000 and 2014, and there was also a relatively large increase in the share of women, and those working in urban areas. Turning our attention to the sectoral composition of earners, the table shows a big drop in the proportion of earners working in the agricultural sector—from 10 per cent to 6 per cent. There was also a large relative fall in the share of mining, while the share of those working in the trade sector rose slightly. The finance and services sectors experienced the largest gains. The shares of both of these sectors rose by 5 percentage points over the 2000–14 period, and the services sector made up one-quarter of all employees in the last year of our study. Finally, the proportion of workers in domestic services fell from 13 per cent to 10 per cent over the period.

[2] The composition of the labour force extends to the final quarter of 2012, though the earnings and inequality analysis ends in the final quarter of 2011.

The third-last and second-last rows of the table give the mean of earnings and the mean of the log of earnings, respectively. There was fairly strong growth in the real mean—from R5,740 in 2000, to R7,418 in 2011, to R7,951 in 2014. Interestingly, though, although the mean increased from 2011 to 2014, the mean of the log distribution fell during this period. This unusual finding comes about because of the different growth rates in the distribution, and was shown by the blue line in Figure 2. The relatively high growth rates at the top of the distribution between 2011 and 2014 were almost solely responsible for pulling the mean up, given the low growth at the bottom and stagnation in the middle of the distribution.

Table 1: Summary statistics of key variables in 2000, 2011, and 2014

	2000	2011	2014
Years of education	8.77	10.54	10.75
Potential experience	22.25	21.33	21.52
Union member	0.34	0.30	0.29
Formal employment	0.79	0.75	0.78
African	0.70	0.70	0.73
Coloured	0.12	0.12	0.12
Asian/Indian	0.03	0.03	0.02
White	0.14	0.14	0.13
Female	0.40	0.45	0.46
Urban	0.68	0.79	0.77
Sectoral shares			
Agriculture	0.10	0.05	0.06
Mining	0.07	0.03	0.03
Manufacturing	0.14	0.14	0.12
Utilities	0.01	0.01	0.01
Construction	0.06	0.07	0.07
Trade	0.15	0.18	0.17
Transport	0.05	0.06	0.06
Finance	0.08	0.13	0.13
Services	0.20	0.24	0.25
Domestic services	0.13	0.10	0.10
Earnings	5740	7418	7951
Log earnings	8.06	8.31	8.18
Observations	24 276	67 235	63 845

Source: authors' calculations using PALMS V3.1 dataset.

4.2 Changes in labour market premiums

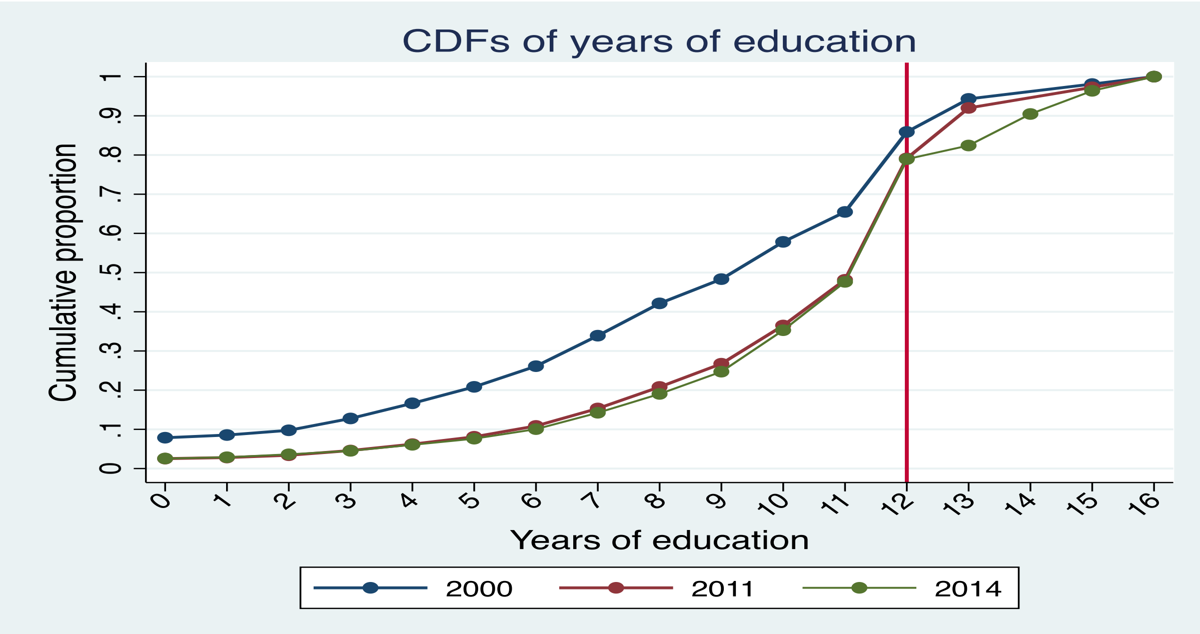
Table 1 provided a very brief overview of the composition of the South African labour market. In Table 2 we take these same variables and present the coefficients of an extended Mincerian regression.[3] Like Ferreira et al. (2017), we loosely refer to these coefficients as the returns to, or premiums of, each of the characteristics. These coefficients provide values at the mean, which we use to anchor our discussion below and we discuss each variable in turn.

Starting with education, a fuller picture of changes in the educational attainment of the labour force is presented in Figure 6, in which cumulative distribution function (CDFs) of years of education are plotted for 2000, 2011, and 2014. As expected from the mean years of education shown in Table 1, there was a sizeable increase in educational attainment for South African workers between 2000 and 2011. The biggest gains came from shifts up from primary to incomplete secondary education. For example, in

[3] Although not discussed in detail in this paper, Figure A1 in the Appendix presents coefficients from unconditional quantile regressions across the full distribution of earnings for eight key variables.

2000, approximately 35 per cent of workers had a grade 7 or below. In 2011 this had decreased to about 17 per cent. The proportion of workers who completed matric (shown by the vertical red line) or below was 86 per cent in 2000, and was 79 per cent in both 2011 and 2014. The 2011 and 2014 CDFs match each other very closely, diverging only in the postsecondary part of the education distribution, where the proportion of workers who completed a tertiary degree was higher in 2014.

Figure 6: CDFs of educational attainment

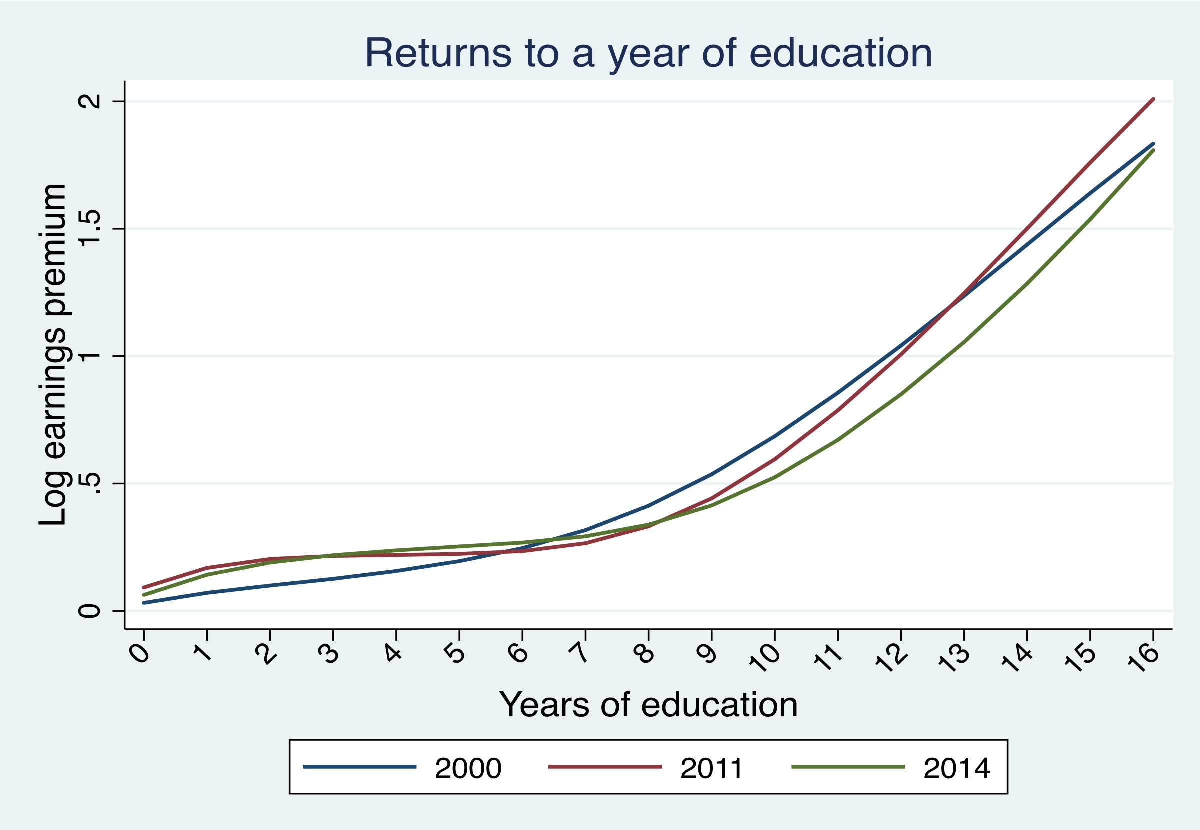


Source: authors' calculations using PALMS V3.1 dataset.

Perhaps the simplest way of comparing the returns to an additional year of education in all three years is to look at the predicted education–earnings profiles based on the quartic polynomial specification of years of education in the regression results reported in Table 2. The lines in Figure 7 are smoothed predictions of the returns to an additional year of schooling, relative to no schooling. The convexity of the education–earnings profile increased between 2000 and 2011/2014; indeed the curves for the later period in our study are practically flat until they reach the incomplete secondary part of the education distribution. The premium associated with a matric, relative to no, education was about the same in 2000 and 2011, which were both higher than was the case in 2014. The returns to postsecondary education were at their highest in 2011, before decreasing in 2014 to below the comparative returns in 2000. The effect of changing returns to education on wage inequality depends on where the changes took place. Lam et al. (2015) identify the year of education at which mean log earnings is earned, and show that if returns increased for education levels below that point, then inequality would decrease. The opposite holds true for increases in inequality above that point. It is therefore something of a puzzle that we see such a dramatic increase in inequality from 2011 to 2014 occurring at the same time as a relative decrease in the returns associated with postsecondary education. The RIF decompositions presented later in this paper will help us to unpack this effect in more detail.

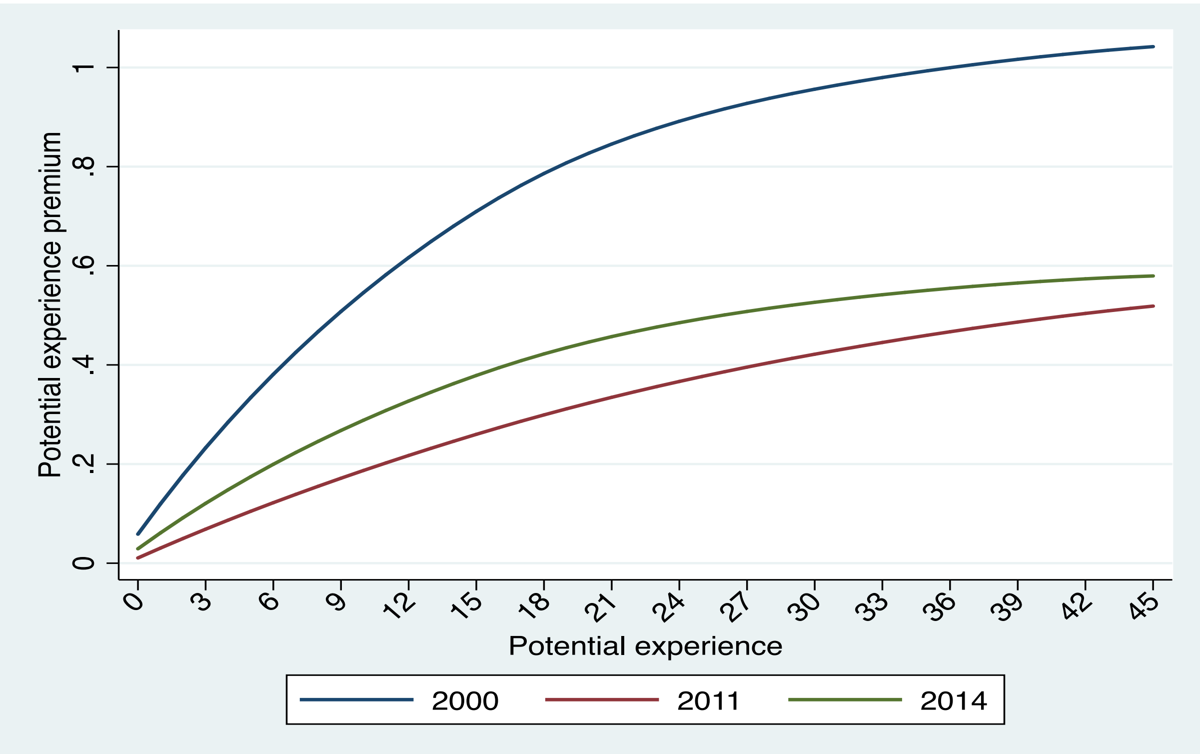
Another potentially important driver of inequality change is the changing returns to potential experience. Ferreira et al. (2017) find that the main factor behind the decline in the wage Gini coefficient for Brazil between 1995 and 2012 is a reduction in the returns to potential experience. The experience–earnings profiles for South Africa in 2000, 2011, and 2014 are presented in Figure 8. Again, these are the predicted values from an extended Mincerian regression where potential experience enters as a quartic polynomial. The figure shows a dramatic reduction in the returns to potential experience between 2000 and 2011. As we show later in this paper, the equalizing effect of the falling returns to potential experience was more than offset by the increasing convexity of the returns to education. The upward

Figure 7: Returns to an additional year of education relative to no schooling



Source: authors' calculations using PALMS V3.1 dataset.

Figure 8: Returns to an additional year of potential experience



Source: authors' calculations using PALMS V3.1 dataset.

shift in the experience–earnings profile between 2011 and 2014 is an early candidate explanation for the sharp rise in earnings inequality in this period.

Returning to Table 2, the conditional penalty to working in the informal sector was fairly consistent over the period, as was the share of informal employment in total employment. Holding all other explanatory variables constant, non-unionized workers earned 21–31 per cent less than their unionzed counterparts, on average. There are some interesting conditional changes in the race coefficients relative to the base group of white workers. Holding all other covariates constant, the conditional average difference between African and white workers was almost identical in 2000 and 2011 at 64 per cent. This dropped to 53 per cent in 2014. The conditional difference between coloured workers and Indian/Asian and white workers increased between 2000 and 2014 by 5.8 per cent and 7.3 per cent, respectively. The changing sectoral composition of the labour force, as well as the premiums associated with each sector, ought to give us some ideas about the shifting relative demand for different skills in different sectors. We have already seen the falling proportions of agriculture and mining in total employment, along with the concomitant increases in the share of the labour force employed in the finance and services sectors. The conditional average difference between those working in the agricultural sector and those working in the mining sector (the base category) fell from over 92 per cent to 56 per cent, showing at least some of the effects of a large increase in the minimum wage in the agricultural sector. The relatively large influx of workers into the finance and services sectors saw decreases in the conditional wages of these sectors relative to mining over the period under study. One final point to take away from Table 2 is the fact that the covariates do a far better job of explaining the total variance of log wages in 2000 and 2011 than they do in 2014, as the R-squared falls from 0.65 to 0.52 to 0.31.

Table 2: Labour market premiums: OLS regression with log earnings as the dependent variable

	2000	2011	2014
Years of education	0.124***	0.295***	0.183***
	(0.0165)	(0.0160)	(0.0221)
(Years of education)^2/100	–3.926***	–9.144***	–4.734***
	(0.463)	(0.404)	(0.557)
(Years of education)^3/1,000	4.731***	9.314***	4.281***
	(0.447)	(0.366)	(0.502)
(Years of education)^4/10,000	–1.457***	–2.677***	–1.006***
	(0.139)	(0.109)	(0.149)
Potential experience	0.0852***	0.0249***	0.0387***
	(0.00706)	(0.00511)	(0.00687)
(Potential experience)^2/100	–0.341***	–0.0755**	–0.145***
	(0.0485)	(0.0363)	(0.0484)
(Potential experience)^3/1,000	0.0676***	0.0164	0.0297**
	(0.0129)	(0.0100)	(0.0132)
(Potential experience)^4/10,000	–0.00522***	–0.00151	–0.00250**
	(0.00116)	(0.000933)	(0.00122)
Not union member	–0.307***	–0.205***	–0.244***
	(0.0110)	(0.00765)	(0.0105)
Informal	–0.353***	–0.337***	–0.316***
	(0.0154)	(0.00928)	(0.0120)
Non-urban	–0.229***	–0.210***	–0.236***
	(0.0111)	(0.00793)	(0.0106)
African	–0.679***	–0.681***	–0.572***
	(0.0149)	(0.00955)	(0.0138)
Coloured	–0.440***	–0.479***	–0.509***
	(0.0186)	(0.0122)	(0.0174)
Asian/Indian	–0.318***	–0.240***	–0.388***
	(0.0267)	(0.0180)	(0.0292)
Female	–0.271***	–0.252***	–0.225***
	(0.0101)	(0.00641)	(0.00887)
Agriculture	–0.748***	–0.510***	–0.410***
	(0.0231)	(0.0227)	(0.0304)
Manufacturing	–0.173***	–0.279***	–0.348***
	(0.0209)	(0.0194)	(0.0265)
Utilities	0.242***	–0.125***	–0.200***
	(0.0458)	(0.0377)	(0.0499)

Construction	–0.216***	–0.259***	–0.340***
	(0.0254)	(0.0212)	(0.0285)
Trade	–0.355***	–0.337***	–0.316***
	(0.0215)	(0.0193)	(0.0260)
Transport	–0.0591**	–0.204***	–0.286***
	(0.0267)	(0.0217)	(0.0291)
Finance	–0.107***	–0.221***	–0.291***
	(0.0243)	(0.0197)	(0.0266)
Services	–0.0796***	–0.244***	–0.392***
	(0.0206)	(0.0187)	(0.0251)
Domestic services	–0.741***	–0.545***	–0.673***
	(0.0263)	(0.0222)	(0.0292)
Constant	7.991***	8.490***	8.425***
	(0.0450)	(0.0379)	(0.0518)
Observations	23,125	65,469	62,063
R-squared	0.647	0.520	0.312

Note: base categories are: white; union member; formal employment; urban; male; mining sector. Robust standard errors in parentheses.
Source: authors' calculations using PALMS V3.1 dataset.

So far we have discussed some of the broad changes in the distributions of and returns to a number of covariates in the earnings equation. We now shift our focus to a narrower discussion of how changes in these covariates determined changes in inequality over time.

5 Results

We follow the same structure as Ferreira et al. (2017) in presenting our results. We begin with a standard Oaxaca-Blinder decomposition for mean earnings, before moving to generalized decompositions of the Gini coefficient, and then different percentile ratios. In each case we present three panels of results—2000 to 2014, 2000 to 2011, and 2011 to 2014.

5.1 Oaxaca–Blinder decomposition of the mean

Table 3 presents the Oaxaca–Blinder decompositions of mean earnings for each window of time covered by our study. The three columns of results are first for the full 2000–14 period, then for the 2000–11 period (so as to avoid using some of the more questionable data from the later period), and finally from 2011–14. There are four panels of results in the table. The first of these reports the means of log earnings in the relevant first and second periods, the difference between them (post minus pre), and how much of the difference is explained by changes in the distribution of covariates (endowments) compared to changes in the returns to these covariates (structure). The second panel separates the full endowment effect into the contributions from each group of covariates—education, experience, unionization, formality, race, gender, urban–rural differences, and sector of employment. The third panel does the same, except this time for the structure effect. The final panel combines the endowment and structure effects of each group of covariates, allowing us to see the overall contribution of each.

Table 3: Oaxaca–Blinder decomposition of changes in mean earnings

	2000–14	2000–11	2011–14
		Overall	
Post	8.18	8.32	8.18
	(0.01)	(0.00)	(0.01)
Pre	8.04	8.04	8.32
	(0.01)	(0.01)	(0.00)
Difference	0.14	0.27	–0.14
	(0.01)	(0.01)	(0.01)
Endowments	0.12	0.17	–0.01
	(0.01)	(0.01)	(0.00)
Structure	0.02	0.10	–0.13
	(0.01)	(0.01)	(0.01)
		Endowment	
Education	0.14	0.17	0.00
	(0.00)	(0.00)	(0.00)
Potential experience	0.00	–0.01	0.00
	(0.00)	(0.00)	(0.00)
Union	–0.01	–0.01	0.00
	(0.00)	(0.00)	(0.00)
Formal status	0.00	–0.02	0.01
	(0.00)	(0.00)	(0.00)
Race	–0.01	0.00	–0.01
	(0.00)	(0.00)	(0.00)
Gender	–0.01	–0.01	0.00
	(0.00)	(0.00)	(0.00)
Urban	0.02	0.02	–0.01
	(0.00)	(0.00)	(0.00)
Sector	0.00	0.01	0.00
	(0.00)	(0.00)	(0.00)
		Structure	
Education	–0.02	–0.22	0.16
	(0.03)	(0.03)	(0.02)
Potential experience	–0.34	–0.43	0.09
	(0.06)	(0.05)	(0.04)
Union	0.04	0.08	–0.04
	(0.01)	(0.01)	(0.01)
Formal status	0.01	0.00	0.01
	(0.01)	(0.01)	(0.00)
Race	0.07	0.03	0.03
	(0.03)	(0.03)	(0.02)
Gender	0.02	0.00	0.01
	(0.01)	(0.01)	(0.01)
Urban	0.00	0.01	–0.01
	(0.01)	(0.01)	(0.00)
Sector	0.12	0.11	0.01
	(0.03)	(0.03)	(0.02)
Constant	0.13	0.52	–0.39
	(0.07)	(0.07)	(0.05)
		Total	
Education	0.12	–0.05	0.17
	(0.03)	(0.03)	(0.02)
Potential experience	–0.35	–0.44	0.09
	(0.06)	(0.05)	(0.04)
Union	0.03	0.07	–0.04
	(0.01)	(0.01)	(0.01)
Formal status	0.00	–0.01	0.02
	(0.01)	(0.01)	(0.00)
Race	0.05	0.03	0.02
	(0.03)	(0.03)	(0.02)
Gender	0.00	–0.01	0.01
	(0.01)	(0.01)	(0.01)

Urban	0.02	0.03	-0.01
	(0.01)	(0.01)	(0.00)
Sector	0.12	0.12	0.01
	(0.03)	(0.03)	(0.02)
N	85,996	89,459	128,497

Note: robust standard errors in parentheses.
Source: authors' calculations using PALMS V3.1 dataset.

Overall the log mean went up over the period, though it first increased by 0.27 log points from 2000 to 2011, and then decreased by 0.14 log points from 2011 to 2014. The fall in the mean of log earnings was discussed in a previous section. Changes in endowments explain almost all (0.12 log points out of a total of 0.14 log points) of the increase in the mean of log wages from 2000 to 2014. This overall change of 0.14 log points is, however, much smaller than the change between 2000 and 2011, which was 0.27 log points. Interestingly, the 2000–11 change was more evenly explained by a combination of endowment and structure effects—-the former accounting for 0.17 log points and the latter for 0.10 log points. The surprisingly large drop in the mean of log earnings from 2011 to 2014 is almost entirely accounted for by changes in the returns to various characteristics, rather than by changes in the distributions of the characteristics themselves. This is unsurprising, given that the distributions of the characteristics themselves did not change much in the relatively short time between 2011 and 2014 (see, for example, the similarity in the educational attainment CDFs in Figure 6).

Given that changes in endowments move more slowly than changes in premiums, it is unsurprising that there is almost no action in the role of changing endowments on changes in the mean for the 2011–14 period (column 3 of panel 2). Over the full period, however (column 1 in panel 2) we see that the overall increase in educational attainment contributed the full amount to the increase of the mean of the log distribution (0.14 log points). The positive contribution made by shifts of workers to urban areas (0.2) was offset by the decrease in unionization (see Table 1), as well as by changes in the gender composition of wage earners. A very similar dynamic is present in the slightly shorter 2000–11 period, with the exception being that the higher share of informality led to a 0.2 log point lowering of the mean. The role of changes in the distribution (endowment effect) of potential experience is negligible in each window of time.

Changes in the mean of the log distribution that came about because of changes in wage structure, or returns to different characteristics, can be seen in the third panel of the table. The effects of changes in the returns to education on changes of the mean were not statistically significant over the full 2000–14 period—in contrast to the effects of changes in the educational distribution discussed earlier. This, however, hides two very different dynamics for each sub-period. In column 2 we see that changes in the returns to education led to a 0.22 log point lowering of the mean between 2000 and 2011, but a 0.16 log point increase of the mean in the short time between 2011 and 2014. The changes in the returns to potential experience are now the main drivers of changes in the mean. This is in contrast to the entirely muted role of changes in the distribution of potential experience as shown in panel 2. Changes in returns to potential experience appear to have contributed negatively to changes in the mean over the full period, although this was slightly offset by a positive contribution in the 2011–14 period. Increasing returns to worker unionization raised the mean by 0.04 log points over the full period, though this effect is doubled if we consider only the 2000–11 window. Wage structure changes by race contributed positively to changes in mean log earnings for each of the three windows we consider, though the standard errors on these are relatively high. Finally, wage structure changes by sector led to a 0.12 log mean increase over the full period, though this was driven almost entirely by changes over the 2000–11 period, rather than over the 2011–14 period. The role of unobserved factors in changes in the mean of the log distribution are given by the constant, which appears in the final row of panel 3. These unobserved factors were large and positive in the 2000–11 period, but had the opposite sign in the 2011–14 period. The result was that their effect on changes in the mean over the full period was of the same magnitude as the race variables (0.07 log points), though with larger standard errors.

The final panel of Table 3 provides the joint endowment and structure contributions of each of the covariates, along with the joint standard errors.[4] The overall rise in the mean of log earnings over the full period was mainly due to educational factors and changes in the distribution and wage structure of the sectoral composition of the labour force. The other human capital variable—potential experience—was associated with a significant drop in mean earnings. Changes in unionization, formality, gender, and urban versus rural earnings were less important in explaining changes at the mean over the full period. Later in this paper we will unshackle ourselves from the mean and investigate the effects of changes in all of these covariates over the full distribution of earnings.

So far we have presented a standard Oaxaca–Blinder decomposition of the mean for the South African labour market post-2000. In the following subsection we exploit the RIF framework, which was outlined in the methodology section of the paper, in order to better understand the role of each of our covariates in determining changes in earnings inequality as measured by the Gini coefficient.

5.2 Explaining changes in the Gini coefficient

The results shown in Table 4 should be interpreted in the same way as the results in Table 3, except this time the decomposition is of the Gini coefficient rather than the mean of log earnings from 2000–14. The results are presented in four panels (Overall, Endowments, Structure, Total) and seven columns, with each successive column providing decomposition results for a richer specification of the model.

There was an 8.23 percentage point (about 15 per cent) increase in the Gini coefficient over the full period—from 55.2 to 63.4 This equates to an average increase of about 1 per cent per year, which, if the data are to be believed, is surprising, given that the Gini is usually quite a slow-moving statistic.

The first column of results includes only education and potential experience as covariates in the model. In this first column we see that changes in the distribution of education and potential experience actually lowered the Gini coefficient by 0.78 percentage points. Changes in the returns to these variables between 2000 and 2014, however, raised the Gini coefficient by 9 percentage points. This column is, in fact, the last time we see the endowment effect lowering inequality over the full period. As we add more covariates to our specification, the contribution of the endowment effect to increasing inequality remains very small, and in some cases is not statistically significant at the 10 per cent level.

Changes in the returns to the demographic, geographic, and labour market variables generally contributed almost all of the 8 percentage point increase in the Gini coefficient over the period. Let us now focus on column 7 of the final panel (Total). This is the total contribution of the endowment and structure effects of each variable in the full model. We see that, holding everything else constant, the effect of changes in education was to increase the Gini coefficient over the period by 11.6 percentage points. This was driven almost entirely by changes in the returns to education, as shown in the 'Structure' panel. The direction of this effect is not surprising, especially given the increasing convexity of the education–earnings profile shown in Figure 7, but its magnitude is notably large. The role of changes in potential experience in driving changes in inequality is very small. The total effect is negative in sign, but is not statistically significant at the 10 per cent level. Even though the endowment effect of changes

[4] Note that in order to obtain the full effect, these need to be interpreted along with the constant that is reported in the final row of the third panel.

in potential experience is negative and statistically significant in isolation, its economic significance is very small.

The total impact of changes in union status on the Gini coefficient is not statistically significant, but this hides interesting findings if one considers the endowment and structure effects separately. The decreasing proportion of the sample who were union members (see Table 1) contributed to a 0.45 percentage point increase in measured earnings inequality. However, changes in the returns to unionization change inequality in the opposite direction, and reduce the Gini coefficient by just over 2 percentage points. The total impact of changes in the distributions of and returns to being in formal employment, racial and gender differences, and whether a worker was in an urban or rural work environment were all statistically insignificant. However, the increasing share of female workers, in isolation, lowered the Gini coefficient by a fifth of 1 percentage point. One final point of interest is that although there were significant changes in the sectoral composition in the labour market between 2000 and 2014, the endowment, structure, and total effects of the sector categorical variable are all not statistically significant at conventional levels for this time period. This result holds regardless of which sector is selected as the base category. In short, the Gini coefficient increased by over 8 percentage points from 2000 to 2014. Almost all of this change can be attributed to inequality-enhancing changes in the return to education, although there were some small mitigating effects through the channels of unionization, changes in the distribution of potential experience, and the increased share of female workers.

Table 4: RIF regression decomposition of the Gini coefficient: 2000–14

	1	2	3	4	5	6	7
				Overall			
Post	63.43	63.43	63.43	63.43	63.43	63.43	63.43
	(0.37)	(0.37)	(0.37)	(0.37)	(0.37)	(0.37)	(0.37)
Pre	55.20	55.20	55.20	55.20	55.20	55.20	55.20
	(0.59)	(0.58)	(0.58)	(0.58)	(0.58)	(0.58)	(0.57)
Difference	8.23	8.23	8.23	8.23	8.23	8.23	8.23
	(0.69)	(0.69)	(0.69)	(0.69)	(0.69)	(0.69)	(0.68)
Endowments	–0.78	0.66	1.12	0.40	0.37	0.30	0.25
	(0.20)	(0.26)	(0.25)	(0.25)	(0.24)	(0.24)	(0.25)
Structure	9.02	7.58	7.11	7.83	7.86	7.93	7.99
	(0.64)	(0.63)	(0.63)	(0.65)	(0.65)	(0.65)	(0.65)
				Endowments			
Education	–0.73	0.14	0.60	0.04	0.09	0.21	0.42
	(0.21)	(0.24)	(0.23)	(0.22)	(0.23)	(0.23)	(0.24)
Potential experience	–0.05	–0.11	–0.12	–0.10	–0.10	–0.10	–0.08
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Union		0.63	0.56	0.49	0.50	0.50	0.45
		(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Formal status			0.09	0.10	0.10	0.10	0.07
			(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
Race				–0.14	–0.14	–0.15	–0.17
				(0.05)	(0.05)	(0.05)	(0.05)
Gender					–0.09	–0.09	–0.20
					(0.05)	(0.05)	(0.05)
Urban						–0.18	–0.19
						(0.04)	(0.04)
Sector							–0.06
							(0.13)
				Structure			
Education	6.13	7.19	7.39	5.89	6.22	6.63	11.17
	(3.65)	(3.91)	(3.85)	(3.26)	(3.44)	(3.41)	(4.06)
Potential experience	1.06	–1.77	–1.26	–1.38	–1.39	–1.62	–2.53
	(3.43)	(3.72)	(3.71)	(3.85)	(3.85)	(3.83)	(3.87)
Union		–5.03	–4.09	–4.14	–4.19	–4.08	–2.09
		(1.18)	(1.28)	(1.15)	(1.20)	(1.21)	(1.22)

Formal status			–1.16	–1.18	–1.35	–1.36	–0.22
			(0.20)	(0.18)	(0.18)	(0.17)	(0.21)
Race				3.31	3.46	4.12	3.47
				(3.12)	(3.14)	(3.13)	(3.13)
Gender					0.71	0.64	0.62
					(0.64)	(0.65)	(0.62)
Urban						–1.27	–0.62
						(0.23)	(0.23)
Sector							–0.87
							(2.80)
Constant	1.82	7.18	6.23	5.33	4.41	4.87	–0.95
	(4.00)	(3.70)	(3.68)	(4.68)	(4.99)	(5.00)	(5.37)
				Total			
Education	5.40	7.33	7.99	5.94	6.31	6.85	11.59
	(3.61)	(3.86)	(3.79)	(3.20)	(3.37)	(3.35)	(4.00)
Potential experience	1.00	–1.88	–1.39	–1.48	–1.49	–1.72	–2.61
	(3.42)	(3.71)	(3.71)	(3.85)	(3.84)	(3.83)	(3.87)
Union		–4.39	–3.53	–3.65	–3.69	–3.58	–1.65
		(1.20)	(1.31)	(1.18)	(1.23)	(1.23)	(1.24)
Formal status			–1.07	–1.08	–1.25	–1.26	–0.15
			(0.21)	(0.19)	(0.18)	(0.18)	(0.22)
Race				3.17	3.32	3.97	3.30
				(3.14)	(3.15)	(3.14)	(3.14)
Gender					0.62	0.55	0.42
					(0.67)	(0.67)	(0.65)
Urban						–1.45	–0.81
						(0.21)	(0.21)
Sector							–0.92
							(2.78)
N	85,996	85,996	85,996	85,996	85,996	85,996	85,996

Note: robust standard errors in parentheses.
Source: authors' calculations using PALMS V3.1 dataset.

Given our concerns about how much of the change in inequality between 2000 and 2014 was driven by 'real' changes and how much was driven by changes in StatsSA's data collection and processing methodology, we now restrict our analysis to the 2000–11 time period. Table 5 presents results in the same form as Table 4, but this time only up to the final quarter of 2011.

Table 5: RIF regression decomposition of the Gini coefficient: 2000–11

	1	**2**	**3**	**4**	**5**	**6**	**7**
				Overall			
Post	56.30	56.30	56.30	56.30	56.30	56.30	56.30
	(0.21)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)
Pre	55.20	55.20	55.20	55.20	55.20	55.20	55.20
	(0.59)	(0.58)	(0.58)	(0.58)	(0.58)	(0.58)	(0.57)
Difference	1.10	1.10	1.10	1.10	1.10	1.10	1.10
	(0.62)	(0.62)	(0.62)	(0.61)	(0.61)	(0.61)	(0.61)
Endowments	–2.62	–1.38	–0.44	–0.77	–0.78	–1.06	–1.48
	(0.12)	(0.15)	(0.15)	(0.14)	(0.13)	(0.14)	(0.15)
Structure	3.72	2.48	1.54	1.87	1.88	2.16	2.58
	(0.61)	(0.60)	(0.60)	(0.60)	(0.60)	(0.60)	(0.60)
				Endowments			
Education	–2.60	–1.82	–1.18	–1.57	–1.51	–1.29	–0.98
	(0.12)	(0.13)	(0.12)	(0.11)	(0.12)	(0.12)	(0.12)
Potential experience	–0.02	–0.09	–0.10	–0.07	–0.07	–0.08	–0.05
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Union		0.53	0.43	0.39	0.39	0.39	0.36
		(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)

Formal status			0.41	0.43	0.45	0.44	0.28
			(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Race				0.04	0.04	0.04	0.04
				(0.02)	(0.02)	(0.02)	(0.02)
Gender					–0.08	–0.08	–0.13
					(0.02)	(0.02)	(0.02)
Urban						–0.48	–0.44
						(0.04)	(0.04)
Sector							–0.56
							(0.09)
				Structure			
Education	10.09	11.50	11.04	9.02	9.28	9.53	13.77
	(3.50)	(3.75)	(3.68)	(3.05)	(3.22)	(3.20)	(3.87)
Potential experience	–6.79	–10.21	–9.59	–10.05	–10.12	–10.23	–11.19
	(3.33)	(3.55)	(3.55)	(3.70)	(3.69)	(3.68)	(3.70)
Union		–5.49	–5.10	–4.89	–4.96	–4.86	–2.55
		(1.06)	(1.16)	(1.02)	(1.07)	(1.07)	(1.09)
Formal status			–0.75	–0.86	–1.00	–1.02	–0.10
			(0.18)	(0.15)	(0.15)	(0.15)	(0.19)
Race				5.81	5.91	6.32	6.03
				(2.80)	(2.81)	(2.80)	(2.79)
Gender					0.58	0.50	0.66
					(0.59)	(0.59)	(0.55)
Urban						–0.53	–0.06
						(0.21)	(0.21)
Sector							–0.79
							(2.62)
Constant	0.43	6.69	5.94	2.84	2.18	2.44	–3.20
	(3.99)	(3.54)	(3.51)	(4.42)	(4.79)	(4.79)	(5.00)
				Total			
Education	7.49	9.69	9.86	7.45	7.77	8.24	12.79
	(3.48)	(3.73)	(3.66)	(3.03)	(3.20)	(3.18)	(3.85)
Potential experience	–6.82	–10.31	–9.69	–10.12	–10.19	–10.31	–11.24
	(3.36)	(3.58)	(3.58)	(3.72)	(3.72)	(3.70)	(3.72)
Union		–4.97	–4.67	–4.50	–4.57	–4.47	–2.19
		(1.07)	(1.17)	(1.03)	(1.08)	(1.08)	(1.10)
Formal status			–0.35	–0.43	–0.55	–0.58	0.18
			(0.19)	(0.17)	(0.16)	(0.16)	(0.20)
Race				5.86	5.95	6.36	6.07
				(2.80)	(2.81)	(2.80)	(2.79)
Gender					0.51	0.43	0.53
					(0.59)	(0.59)	(0.56)
Urban						–1.01	–0.50
						(0.19)	(0.20)
Sector							–1.35
							(2.61)
N	89,459	89,459	89,459	89,459	89,459	89,459	89,459

Note: robust standard errors in parentheses.
Source: authors' calculations using PALMS V3.1 dataset.

Unlike the decompositions reported in Table 4, there was only a small increase in the Gini coefficient of earnings between 2000 and 2011—from 55.2 to 56.3. However, this seemingly small change hides very large changes in the dynamics driving the generation of inequality.

Overall (see panel 1), the endowment and structure effects worked in opposite directions. Changes in the distribution of all covariates lowered the Gini coefficient by 1.48 percentage points (see column 7). Changes in the returns to these covariates were inequality-enhancing, and more than compensated for the inequality-reducing endowment effect.

The main feature of the results in this table is the presence of countervailing effects of changes in education and in returns to potential experience. Let us once again focus on column 7 in the fourth panel (Total) in order to understand the role of each variable in generating the small increase in inequality over this period. Like Brazil over a similar period, changes in the returns to potential experience served to decrease the Gini coefficient (Ferreira et al., 2017). However, these inequality-reducing effects were more than counteracted by inequality-enhancing changes in the returns to education. Unsurprisingly, the general increase in the educational attainment of the labour force reduced inequality by about 1 percentage point (see column 7 of the Endowments panel). However, the changes in returns to education—highlighted by the increasing convexity of the education–earnings profile and the higher relative returns to postsecondary education—meant that changes in the returns to education were once again strongly inequality-enhancing overall. We saw that over the 2000–14 period (Table 4), the total effect of changes in the distribution and returns to potential experience was negligible. Therefore, there was not much to mitigate the role of the education variable in driving the large increase in inequality in our decomposition. The story is very different if we end our analysis at 2011, however. Now the 12.8 percentage point increase in the Gini coefficient that is attributable to education is counteracted by an 11.24 percentage point decrease that is attributable to changes in the returns to potential experience. The reason for this result is that the returns to each year of potential experience are a lot flatter (see Figure 8), and the average level of potential experience in the labour market was about one year lower in 2011 than it was in 2000. This second point is driven, of course, by the fact that the workers of 2011 spent more years in education than the workers of 2000.

Although the education and potential experience variables dominate this decomposition, there are a few interesting points to make about the role of other variables as well. Lower union coverage was inequality-enhancing, but the returns to being a union member reduced the wage Gini coefficient by a little over 2.5 percentage points. The overall effect of the union variable in the decomposition was to reduce the Gini coefficient by just over 2 percentage points, and this effect was statistically significant at the 5 per cent level. Changes in the returns to non-white workers relative to white workers increased the Gini coefficient by 6 percentage points. This is in contrast to what was shown in Table 4, where the race variable was positive but not statistically significant. Finally, the total effects of changes in formalization, the gender of workers, and the distribution and returns to different sectors were not significant drivers of the change in inequality over the 2000–11 period.

In the first of the Gini coefficient decompositions, we saw that the increase in inequality between 2000 and 2014 was largely driven by inequality-enhancing changes in the returns to education. In the next set of decompositions, from 2000 to 2011, we saw that the (far smaller) increase was again driven by changes in the returns to education, but that these were muted by inequality-reducing changes in the returns to potential experience. In our final table of Gini coefficient decompositions, Table 6, we investigate what drove the increase in measured earnings inequality for the short window of time between 2011 and 2014.

Table 6: RIF regression decomposition of the Gini coefficient: 2011–14

	1	2	3	4	5	6	7
				Overall			
Post	63.43	63.43	63.43	63.43	63.43	63.43	63.43
	(0.37)	(0.37)	(0.37)	(0.37)	(0.37)	(0.37)	(0.37)
Pre	56.30	56.30	56.30	56.30	56.30	56.30	56.30
	0.21	0.20	0.20	0.20	0.20	0.20	0.20
Difference	7.13	7.13	7.13	7.13	7.13	7.13	7.13
	(0.42)	(0.42)	(0.42)	(0.42)	(0.42)	(0.42)	(0.42)
Endowments	–0.02	0.10	–0.12	–0.34	–0.36	–0.32	–0.32
	(0.03)	(0.04)	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)
Structure	7.15	7.03	7.25	7.47	7.50	7.45	7.46
	(0.42)	(0.41)	(0.41)	(0.42)	(0.43)	(0.43)	(0.43)

				Endowments			
Education	–0.03	0.00	0.01	0.00	0.00	0.00	0.01
	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Potential experience	0.01	0.02	0.03	0.02	0.02	0.02	0.01
	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Union		0.08	0.07	0.06	0.06	0.06	0.05
		(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Formal status			–0.23	–0.26	–0.26	–0.26	–0.18
			(0.02)	(0.03)	(0.03)	(0.03)	(0.02)
Race				–0.15	–0.15	–0.16	–0.18
				(0.05)	(0.05)	(0.05)	(0.05)
Gender					–0.03	–0.03	–0.06
					(0.01)	(0.01)	(0.02)
Urban						0.05	0.05
						(0.01)	(0.01)
Sector							–0.04
							(0.04)
				Structure			
Education	–2.05	–2.35	–1.88	–1.51	–1.46	–1.40	–1.21
	(1.36)	(1.45)	(1.43)	(1.37)	(1.42)	(1.41)	(1.44)
Potential experience	7.81	8.40	8.27	8.63	8.68	8.57	8.62
	(2.61)	(2.69)	(2.69)	(2.70)	(2.69)	(2.69)	(2.74)
Union		0.50	1.07	0.79	0.82	0.83	0.48
		(0.74)	(0.80)	(0.75)	(0.78)	(0.78)	(0.80)
Formal status			–0.50	–0.39	–0.43	–0.42	–0.15
			(0.16)	(0.16)	(0.16)	(0.16)	(0.19)
Race				–2.54	–2.48	–2.22	–2.59
				(1.86)	(1.86)	(1.87)	(1.89)
Gender					0.14	0.15	–0.05
					(0.41)	(0.40)	(0.44)
Urban						–0.49	–0.37
						(0.12)	(0.12)
Sector							0.46
							(1.30)
Constant	1.40	0.49	0.29	2.50	2.23	2.43	2.25
	(2.91)	(2.97)	(2.98)	(3.51)	(3.58)	(3.58)	(3.87)
				Total			
Education	–2.08	–2.36	–1.87	–1.51	–1.46	–1.40	–1.20
	(1.36)	(1.45)	(1.42)	(1.37)	(1.42)	(1.40)	(1.44)
Potential experience	7.82	8.42	8.30	8.64	8.69	8.59	8.63
	(2.61)	(2.69)	(2.69)	(2.70)	(2.69)	(2.69)	(2.74)
Union		0.57	1.14	0.85	0.88	0.89	0.54
		(0.74)	(0.80)	(0.76)	(0.78)	(0.78)	(0.80)
Formal status			–0.72	–0.65	–0.70	–0.68	–0.33
			(0.15)	(0.15)	(0.15)	(0.15)	(0.17)
Race				–2.69	–2.63	–2.38	–2.77
				(1.89)	(1.89)	(1.91)	(1.92)
Gender					0.12	0.12	–0.10
					(0.42)	(0.42)	(0.46)
Urban						–0.44	–0.32
						(0.13)	(0.13)
Sector							0.42
							(1.30)
N	128,497	128,497	128,497	128,497	128,497	128,497	128,497

Note: robust standard errors in parentheses.
Source: authors’ calculations using PALMS V3.1 dataset.

From the first panel of Table 6, it is clear that the very large increase in the Gini coefficient of 7.13 percentage points was driven entirely by structure rather than endowment effects. In fact, racial, job formality, and gender endowment effects were all inequality-reducing over the period. Over this time period, the main driver of the increasing Gini coefficient was not changes in the returns to education. It was, in fact, inequality-enhancing changes in the returns to potential experience. This is the single most important change in dynamics between the 2000–11 and 2000–14 periods. In the shorter window (2000–11) we saw that although there was a small increase in the Gini coefficient, this masked the very large countervailing forces between inequality-enhancing changes in the returns to education, and inequality-reducing changes in the returns to potential experience. This dynamic changes completely in the 2011–14 period. Now we see that changes in returns to education did not have a statistically significant effect on changes in the Gini coefficient (see the Structure panel). Changes in the returns to potential experience, however, are solely responsible for the increase in the Gini coefficient that was observed over this period. Recall from Figure 8 that there was an increase in the returns to potential experience profile in 2014 compared to 2011, in contrast to the very large decrease in this profile between 2000 and 2011. The 2011–14 rise in this profile is surprising, but even if we accept it as reflecting a 'real' change, it is difficult to imagine that this could have been solely responsible for such a large rise in earnings inequality. This stark reversal in the role of potential experience in explaining inequality strikes us as suspicious, and we interpret this as an anomaly, rather than as an actual change in the true underlying determination of wage inequality.

Tables A1, A2, and A3 in the Appendix present a range of robustness tests for each window by running the Gini decomposition without weights, restricting to male earners only, using hourly earnings instead of monthly earnings, and by restricting the analysis to full-time workers only (those working at least 35 hours per week). The qualitative findings are not sensitive to any of these changes.

We have spent some time discussing the decomposition of the Gini coefficient, but have not yet exploited one of the advantages of the RIF methodology in that it allows us to investigate endowment, structure and total effects at any percentile of the distribution. In the following section we present the role of each variable in the decomposition graphically, before turning to a decomposition of changes in different percentile ratios.

5.3 RIF decomposition at different percentiles, and of percentile ratios

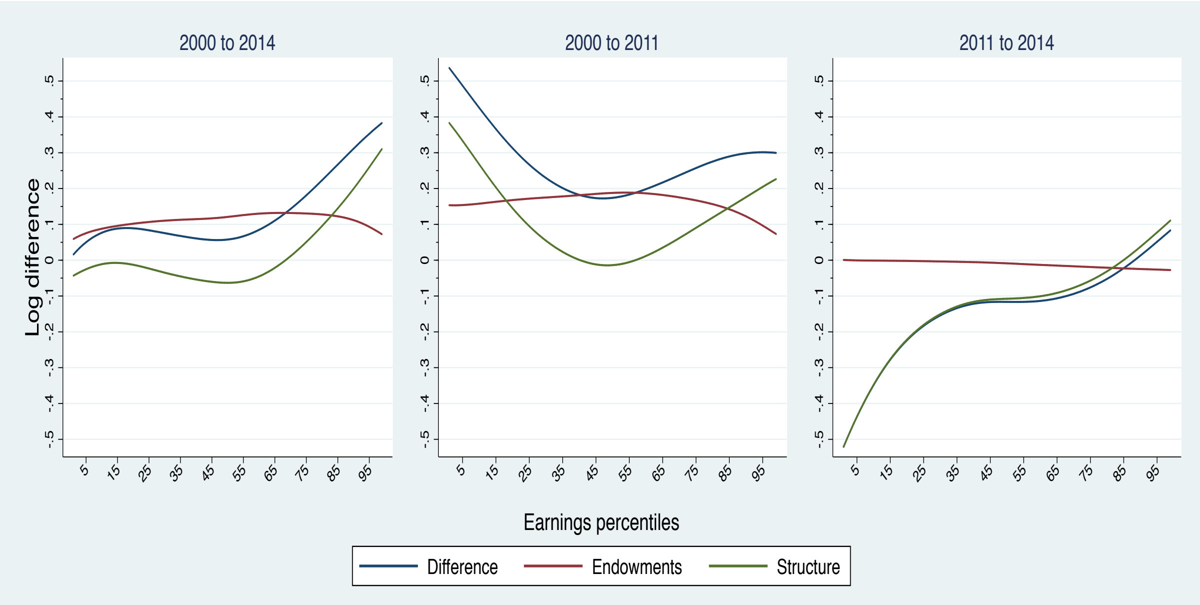
The key innovation of RIF/UQR regression is that it enables us to perform Oaxaca–Blinder-style decompositions at any point of the distribution of earnings. In Figure 9 we decompose changes at each percentile of the wage distribution into endowment and structure effects. The three panels of the figure are for the full 2000–14 period, the 2000–11 period, and the 2011–14 period, respectively.

The blue line shows the observed difference between earnings at each percentile. In effect, this is a growth incidence curve, analogous to what was shown previously in Figure 2. The slight difference between the curves in Figure 2 and Figure 9 comes about because the latter are smoothed over the distribution of earnings, while the former are not.

In the first panel of the figure we see that the shape of the overall changes curve (the blue line) is driven by the shape of the structure effect curve (the green line). The endowment effect (red line) is fairly consistent over the whole of the distribution, and contributes little to the observed difference between the overall curve and the structure curve.

The same general relationship between the endowment effect and the overall effect can be seen in the second panel, which is for the 2000–11 period. The difference, however, is that growth rates broadly followed a U-shape over the distribution of earnings, and were lowest in the middle of the log distribution, as previously shown in Figure 2. As was shown in panel 1 of Table 5, the endowment effect lowered earnings inequality, in contrast to the structure effect which drove increases in the Gini coefficient over the period. In this second panel, the endowment effect is positive everywhere, and in fact plays a more

Figure 9: Changes in log wages, endowment effects and structure effects over the distribution of earnings



Source: authors' calculations using PALMS V3.1 dataset.

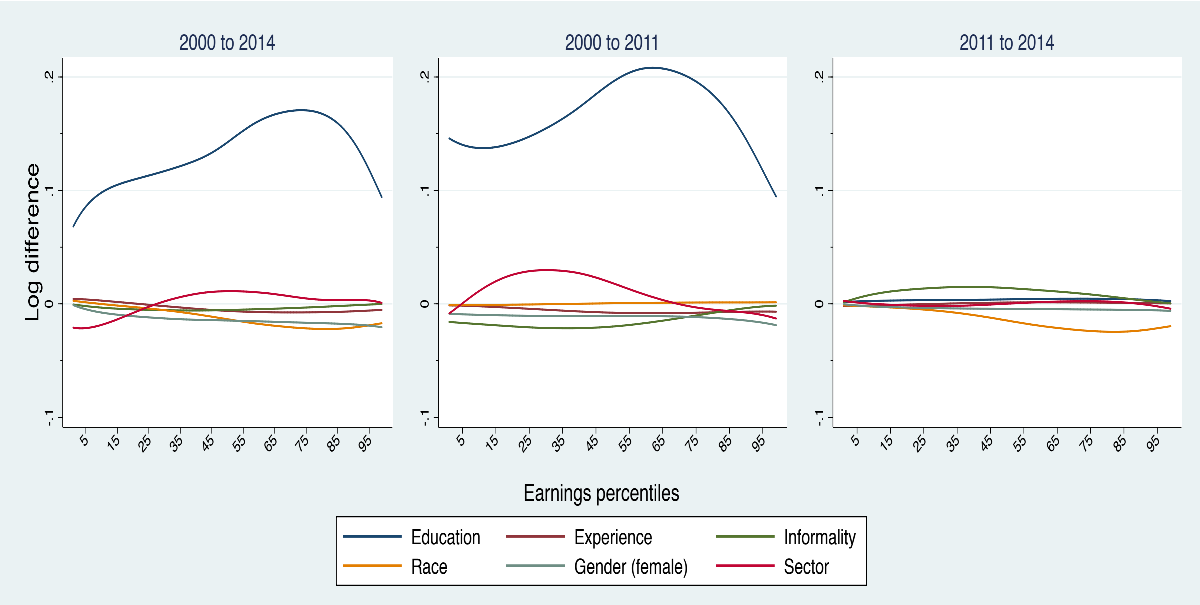
important role than the structure effect in determining changes in log earnings between the 20th and the 85th percentiles.

The final panel in Figure 9, for 2011–14, highlights once again that changes in the distribution of the covariates in our model explain almost none of the overall changes. The biggest negative growth rates over this period took place at the bottom of the distribution, and this is something that we highlighted as a major concern in Figure 1. The overall line crosses zero only at the 85th percentile, and even after this, the level of growth is very low. This collapse of earnings at the bottom of the distribution is the main driver of the inequality jump between 2011 and 2014.

The three panels of Figure 10 are once again for each of the three windows of analysis. In this case we present the effect of endowments (or changes in the distribution of each covariate) at each point along the distribution of earnings.[5] As one might expect, given the discussion in the section of the paper dealing with decompositions of the Gini coefficient, the effect of changes in the distribution of education dominates the effect of changes in the distributions of all other covariates across the whole earnings distribution in both the 2000–14, and the 2000–11 periods (panels 1 and 2 respectively). This difference was, in general, more pronounced towards the top of the earnings distribution. The endowment effect for all other covariates was relatively muted, although changes in the sectoral composition of the labour market contributed to a very slight rise in wages in the bottom half of the distribution. The very small changes in the distributions of the covariates in our model mean that the endowment effect is practically zero across the entire distribution of earnings for all of the six variables shown in the figure.

[5] In this and later figures we do not show the relevant effects of changes in union status or the urban–rural dummy variable on the change in earnings, as these are very small and are not statistically significant.

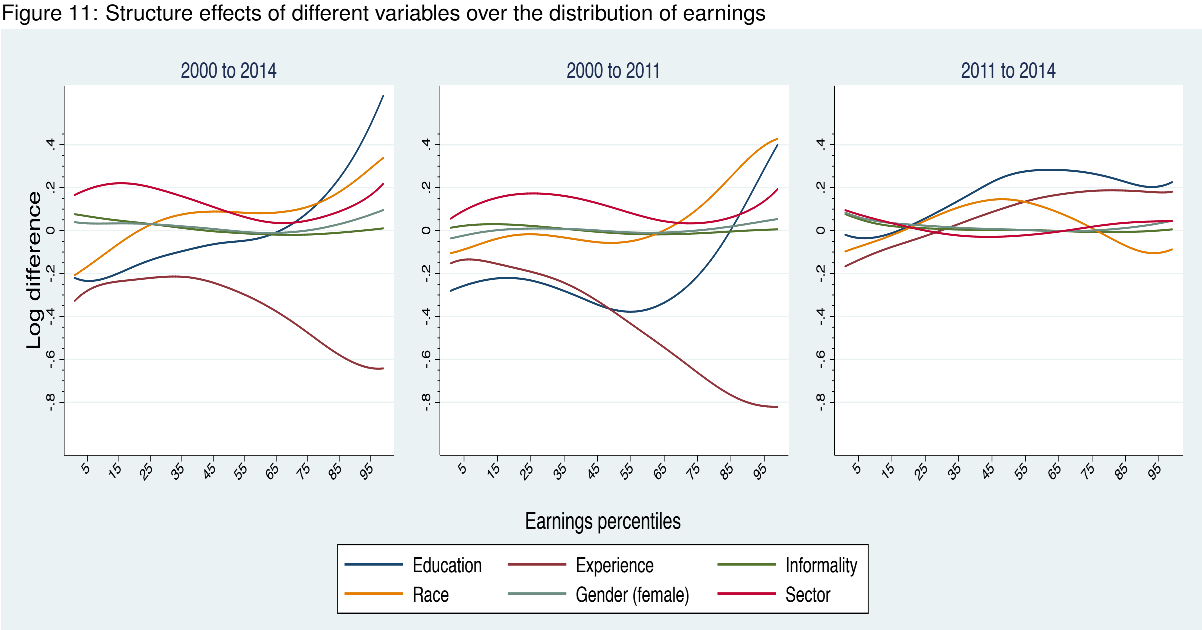
Figure 10: Endowment effects of different variables over the distribution of earnings



Source: authors' calculations using PALMS V3.1 dataset.

The three panels in Figure 11 present the structure effects from the decomposition of earnings changes over the distribution of earnings itself. There is a lot more action in each of these panels compared to what we saw in Figure 10, and this is to be expected, given the relatively large role of structure effects in the Gini coefficient decomposition presented earlier.

Figure 11: Structure effects of different variables over the distribution of earnings



Source: authors' calculations using PALMS V3.1 dataset.

The first panel once again presents results for the full 2000–14 period. Interestingly, changes in the returns to potential experience resulted in lower wages at every percentile. These changes had a stronger negative effect on wages than any other variable in our model, and this was true at every percentile. We can complement the findings of Table 4 and the inequality-reducing role that potential experience played over this time period by noting that the steepest negative effect of the changing returns took place at the very top of the earnings distribution. This is in stark contrast to the inequality-enhancing effects of changes in the return to education, which had a negative impact on wages from the bottom of the distribution all the way to the 70th percentile, and then became increasingly positive for each successively higher percentile. These two curves, which are close to mirroring each other on the horizontal, do the bulk of the work in explaining the increase in earnings inequality that was decomposed in Table 4.

The same general pattern holds in the 2000–11 period, though now the negative effect of changes in the return to education is greater than that of changes in the return to potential experience from the bottom of the distribution up to the median wage earner. Immediately following this middle part of the distribution, the structure effect of education reaches a turning point, and becomes positive at around the 85th percentile. The fact that changes in the return to education had their largest positive effect on those who were earning the highest wages explains why the structure effect of the education variable was so high in the decomposition of the Gini coefficient for the 2000–11 window. Although the second panel in Figure 11 shows that this effect was offset somewhat by the strong negative effect of changes in the returns to potential experience for high earners, the overall result was that the Gini coefficient rose by about 1 percentage point.

One way of explaining this is that those with the highest levels of education are also those who are likely to be at the top of the earnings distribution, thus benefiting the most from changes in the returns to education, while those with the most experience may be spread around the earnings distribution more evenly. Thus, the top of the earnings distribution benefited fully from changes in the returns to education, but were not fully penalized by the negative changes in the returns to potential experience.

The shapes of the education and potential experience curves look very different in the final panel of the figure, which is for the 2011–14 period. Both of these had negative effects on log wages at the bottom of the distribution, but then rose together as they moved rightward. The potential experience curve was always below the education curve, but in the 2011–14 period these converged rather than diverging as they did in the 2000–14 and 2000–11 windows.

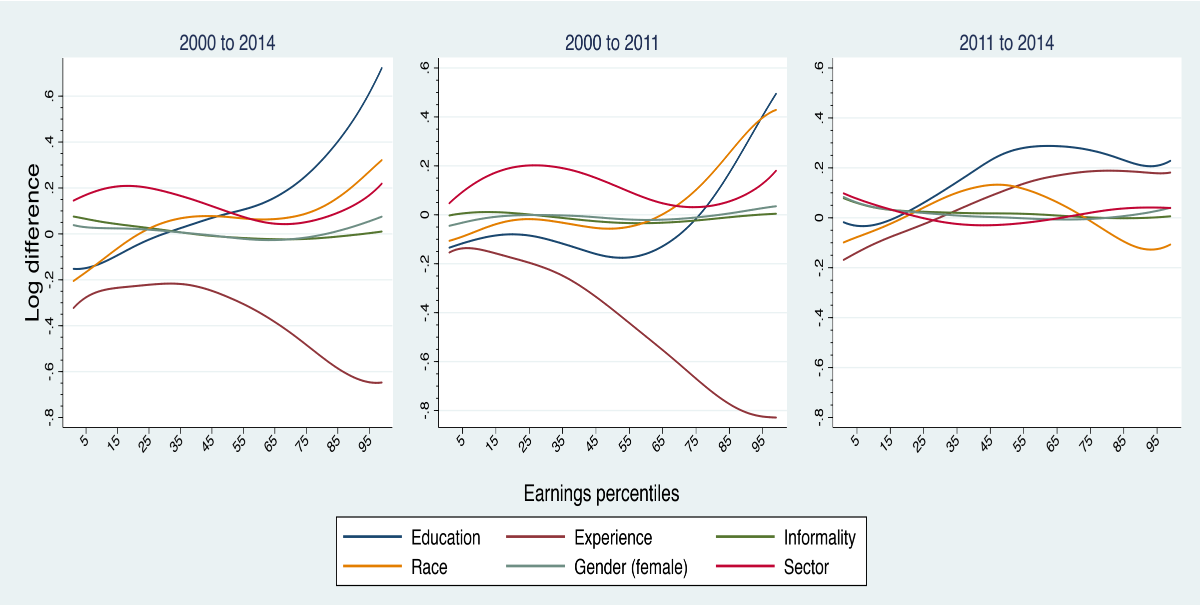
Finally, we present the total combined endowment and structure effects of the variables in our model across the distribution of earnings. The patterns in Figure 12 are, expectedly, mainly driven by the shapes of the curves in Figure 11.

In the first two panels the total impact of changes in education on changes in log earnings is even more strongly positive at the top of the distribution compared to either one of the endowment or structure effect graphs. Countering this, the effect of potential experience is pushed further downwards when the endowment and structure effects are combined. In fact, the single largest effect in absolute terms for any variable is the negative effect of potential experience at the very top of the distribution in the 2000–11 panel. The total effect of the sector variable was positive over the full distribution in the first two panels, and this was largely because of the structure effect. Finally, the total effect of the race variable, relative to the base category of white, was negative in the lower part of the distribution, and positive in the higher part of the distribution in both the 2000–14 and 2000–11 panels in Figure 12. The pattern changed in the 2011–14 period, where the effect relative to the base category had an inverted-U shape, and was at its highest around the 45th percentile.

The previous four figures give us an interesting perspective of where the points of greatest impact of the variables in our model are across the entire distribution. In the final part of this section of the paper we turn our attention to decomposing different percentile ratios, in order to highlight some of the patterns that have already been presented.

Table 7 presents the decompositions of three sets of percentile ratios for each of the three time windows that form part of this study. The log differences for the 90/10, 90/50, and 50/10 percentile ratios are decomposed into endowment, structure, and overall effects.

Figure 12: Total effects of different variables over the distribution of earnings



Source: authors' calculations using PALMS V3.1 dataset.

The 90/10 ratio decreased between 2000 and 2011, reflecting the relatively higher growth at the bottom of the distribution for this period that was presented in Figure 2. This suggests that the rise in earnings inequality for this period was driven by the increases at the very top of the wage distribution. Over the full 2000–14 period, however, the 90/10 ratio increased by 0.2 log points (see the Overall panel), and this is in line with our expectations, given what happened to the Gini coefficient, and the collapse of earnings at the bottom of the distribution post 2012 Q3. The endowment effects of education worked in the opposite direction to the structure effects in determining changes in the 90/10 ratio over each of the three periods. Changes in the distribution of education increased the 90/10 ratio in all three windows, while changes in the returns to education decreased the ratio.

Changes in the returns to potential experience have different impacts on the percentile ratios, depending on which time window is in question. This variable has a large negative effect on the 90/10 ratio for the 2000–14 and 2000–11 periods, but no effect in the 2011–14 period. In general, changes in the returns to potential experience had less of an impact in determining changes in the 50/10 ratio than they did for the 90/10 and 90/50 ratios. The experience and education variables worked in different directions in the 2000–11 period—increasing and decreasing the 50/10 ratio, respectively. This pattern was reversed in the 2011–14 window, where changes in the returns to education had a very large negative effect on the 50/10 ratio, and changes in the returns to potential experience had a large and positive effect.

Changes in formality in the labour market between 2000 and 2011 increased earnings at the 90th percentile relative to the median, and decreased the ratio of earnings at the median compared to the 10th percentile. Changes in the wages accruing to racial groups relative to the white base category increased the 90/10 and 50/10 ratios over the full period. Although the total effect of the race variable in the 2011–14 time period was not statistically significant in explaining changes in the 90/10 or 90/50 ratio, it had a large and positive impact on the ratio of median earnings to earnings at the 10th percentile. As was the case for most of the Gini coefficient decompositions, although there were significant changes in the sectoral composition of the labour market, the effect of these changes was not significant in any economic or statistical sense.

Table 7: RIF regression decomposition of different percentile ratios

	2000–14			2000–11			2011–14		
	90/10	90/50	50/10	90/10	90/50	50/10	90/10	90/50	50/10
					Overall				
Post	3.12	1.72	1.40	2.81	1.58	1.23	3.12	1.72	1.40
	(0.03)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.03)
Pre	2.92	1.36	1.56	2.92	1.36	1.56	2.81	1.58	1.23
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
Difference	0.20	0.37	–0.16	–0.11	0.22	–0.33	0.31	0.15	0.17
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)
Endowments	0.08	0.06	0.01	0.03	0.01	0.02	0.05	0.04	0.01
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)
Structure	0.12	0.30	–0.18	–0.14	0.21	–0.35	0.27	0.11	0.16
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)	(0.01)	(0.03)
					Endowments				
Education	0.20	0.10	0.10	0.11	0.03	0.08	0.06	0.04	0.02
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.00)	(0.00)	(0.00)
Potential experience	–0.01	0.00	–0.01	–0.01	0.00	–0.01	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Union	0.01	0.02	–0.01	0.01	0.02	–0.01	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Formal status	0.01	0.00	0.00	0.02	0.01	0.01	–0.01	–0.02	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Race	–0.02	0.00	–0.02	0.00	0.00	0.00	–0.02	0.01	–0.02
	(0.01)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Gender	–0.01	–0.01	0.00	–0.01	–0.01	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Urban	–0.03	–0.01	–0.02	–0.04	–0.01	–0.03	0.01	0.01	0.01
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Sector	–0.07	–0.04	–0.03	–0.07	–0.04	–0.03	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
					Structure				
Education	–0.76	–0.19	–0.57	–0.27	–0.33	0.07	–0.46	0.17	–0.64
	(0.21)	(0.08)	(0.20)	(0.11)	(0.10)	(0.10)	(0.19)	(0.06)	(0.19)
Potential experience	–0.50	–0.50	0.00	–0.66	–0.43	–0.23	0.16	–0.07	0.23
	(0.16)	(0.13)	(0.11)	(0.15)	(0.14)	(0.10)	(0.10)	(0.08)	(0.08)
Union	0.07	–0.01	0.08	0.05	–0.02	0.07	0.02	0.01	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Formal status	–0.08	0.00	–0.09	–0.02	0.04	–0.06	–0.07	–0.03	–0.03
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)
Race	0.38	0.12	0.25	0.37	0.46	–0.09	–0.01	–0.35	0.35
	(0.11)	(0.10)	(0.04)	(0.09)	(0.09)	(0.04)	(0.07)	(0.05)	(0.03)
Gender	0.10	0.08	0.02	0.10	0.08	0.02	0.00	0.00	0.00
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Urban	–0.01	0.02	–0.03	0.03	0.03	0.00	–0.04	–0.01	–0.03
	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sector	0.00	0.11	–0.11	0.01	0.00	0.01	–0.01	0.10	–0.11
	(0.08)	(0.07)	(0.05)	(0.08)	(0.08)	(0.05)	(0.05)	(0.04)	(0.04)
Constant	0.93	0.68	0.25	0.26	0.39	–0.13	0.67	0.29	0.38
	(0.27)	(0.21)	(0.22)	(0.22)	(0.21)	(0.16)	(0.24)	(0.13)	(0.21)
					Total				
Education	–0.56	–0.09	–0.47	–0.15	–0.30	0.15	–0.40	0.21	–0.62
	(0.10)	(0.10)	(0.04)	(0.09)	(0.09)	(0.04)	(0.06)	(0.05)	(0.03)
Potential experience	–0.50	–0.50	0.00	–0.66	–0.43	–0.23	0.16	–0.07	0.23
	(0.16)	(0.13)	(0.11)	(0.15)	(0.14)	(0.10)	(0.10)	(0.08)	(0.08)
Union	0.08	0.01	0.07	0.05	0.00	0.06	0.02	0.01	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)
Formal status	–0.08	0.01	–0.08	0.00	0.06	–0.05	–0.08	–0.05	–0.03
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Race	0.36	0.12	0.24	0.38	0.46	–0.09	–0.02	–0.35	0.32
	(0.10)	(0.10)	(0.04)	(0.09)	(0.09)	(0.04)	(0.06)	(0.05)	(0.03)
Gender	0.08	0.07	0.02	0.09	0.07	0.02	0.00	0.00	0.00

	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Urban	–0.04	0.01	–0.05	–0.01	0.02	–0.02	–0.03	0.00	–0.03
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sector	–0.07	0.07	–0.13	–0.06	–0.04	–0.03	0.00	0.10	–0.11
	(0.08)	(0.08)	(0.05)	(0.09)	(0.09)	(0.05)	(0.05)	(0.04)	(0.04)
N	85,996	85,996	85,996	89,459	89,459	89,459	128,497	128,497	128,497

Note: robust standard errors in parentheses.
Source: authors' calculations using PALMS V3.1 dataset.

6 Conclusion

A stylized fact of the post-apartheid labour market in South Africa is that wage inequality started off high, and remained high and stable until at least the early 2010s. The first aim of this paper was to investigate whether the roles of variables generating this inequality were themselves stable, or whether there were changes in the underlying dynamics. The second aim was to try to make some sense of the labour market data in the period following the third quarter of 2012.

We began by showing that there was a strong rightward shift in the distribution of earnings between 2000 and 2011, but that a lot of the gains at the bottom of the distribution were, surprisingly, reversed between 2011 and 2014. The implications for rising wage inequality were fairly clear, as this period saw low to negative wage growth rates at the bottom and middle of the distribution, and higher growth rates at the top. This was in stark contrast to the dynamics over the 2000–11 period, where there was growth at both ends of the distribution relative to the middle. The result of this was that although inequality was relatively stable from 2000 to the middle of 2012, it shot up and remained high thereafter.

The fact that the Gini coefficient jumped by 18 per cent in 12 months, and then remained at the new higher level, raised concerns about changes in the quality of and/or methodology used in processing the data. As it turns out, there was a change in the imputation procedure used by StatsSA in preparing the data for public release. However, we are unable to identify the full effect of this change in methodology as (1) we cannot identify which respondents have imputed earnings and (2) we do not know which imputation procedure (or combination or procedures) was used for each of these respondents. This effectively meant that our analysis was broken down into three windows: 2000-14 (the full period), 2000–11 (the period for which we have consistent data), and 2011–14 (to try to see if we can identify which variables drove the sharp increase in inequality).

Having highlighted the dramatic changes after 2012, we distilled some of the key results that remained clear across this fault line. We showed how the composition of the labour market changed in South Africa over the 2000–14 period, with years of education increasing, female labour market participation rising, and sectoral composition shifting away from agriculture and mining and into finance and services. We also showed how the education–earnings profile became more convex over the period, and how the experience–earnings profile flattened considerably between 2000 and 2011, but then rose again between 2011 and 2014.

Then, we used RIF regressions to unpack the roles played by changes in the distributions of key labour market variables and in the returns associated to these variables in generating wage inequality. We found that over the 2000–11 period the returns to education and the returns to potential experience worked in the opposite directions. Changes in the returns to education increased the Gini coefficient, while changes in the returns to potential experience lowered the Gini coefficient over this period. The net result was a small increase in measured wage inequality of about 1 percentage point. The story was very different in the 2011–14 period, however. Although we treat the 2014 Gini coefficient with a great deal of suspicion, it is clear that the 7 percentage point rise over the period is driven almost exclusively by changes in the returns to potential experience. This might be a plausible part of an explanation over a longer period of time, but for such a sharp change to occur over such a short period of time suggests that the shift has more to do with data quality issues than it has to do with any real underlying changes.

RIF regressions allowed us to unshackle ourselves from a standard Oaxaca–Blinder decomposition of the mean, and to decompose earnings at any unconditional quantile of the distribution. Doing so showed how the education and experience variables worked in opposite directions over the full distribution from 2000 to 2011, and from 2000 to 2014, with the differences at their highest at the top of the distribution. The roles of other variables were more muted, but when the total effect was significant it was generally because of the contribution of structure, rather than the endowment effects on the decomposition of changes in earnings.

The increase in the average level of education in the workforce is not yet having a dampening effect on earnings inequality. In large part, this is because more workers are shifting to the steepest part of the returns-to-education function. Thus, paradoxically, the increase in educational attainment has been inequality-enhancing to some extent, as the increased supply of relatively skilled workers has not had a dampening effect on the education premium in the labour market. For the 2000–11 period the change in the distribution of education actually reduced the Gini coefficient, but this was more than offset by changes in the returns to education. Decompositions of the Gini coefficient and different percentile ratios show that the dynamics of earnings inequality in South Africa since 2000 essentially boil down to the relationship between returns to education and returns to experience. Further investigation of why there have been such shifts in the returns to potential experience, and what the role of skills- or age-biased technical change has been in determining wage inequality, is left for future research.

References

Atkinson, A. (1970). 'On the Measurement of Inequality'. *Journal of Economic Theory*, 2(3): 244–63.

Bhorat, H., and S. Goga (2013). 'The Gender Wage Gap in Post-Apartheid South Africa: A Re-examination'. *Journal of African Economies*, 22(5): 827–48.

Blinder, A. (1973). 'Wage Discrimination: Reduced Form and Structural Estimates'. *Journal of Human Resources*, 8(4): 436–55. .

Ferreira, F., S. Firpo, and J. Messina. (2017). 'Ageing Poorly? Accounting for the Decline in Earnings Inequality in Brazil, 1995–2012'. Policy Research Working Paper 8018. Washington, DC: World Bank.

Finn, A. (2015). 'A National Minimum Wage in the Context of the South African Labour Market'. SALDRU Working Paper 153. Cape Town: Southern Africa Labour and Development Research Unit, University of Cape Town.

Firpo, S., N. Fortin, and T. Lemieux (2009). 'Unconditional Quantile Regressions'. *Econometrica*, 77(3): 953–73.

Fortin, N., T. Lemieux, and S. Firpo (2011). 'Decomposition Methods in Economics'. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics, Volume 4A*. Amsterdam: Elsevier.

Hosking, S. (2016). 'Labour Market Inequality in South Africa: A Decomposition of Changes in Earnings from 2001 to 2011'. Master's thesis. Cape Town: University of Cape Town.

Kerr, A., and M. Wittenberg (2016). *A Guide to Version 3.1 of the Post-Apartheid Labour Market Series (PALMS). PALMS User Guide 3.1*. Cape Town: DataFirst, University of Cape Town.

Kerr, A., and M. Wittenberg (2017). 'QLFS earnings data quality'. Presentation at DataFirst data quality workshop, Cape Town.

Kerr, A., D. Lam, and M. Wittenberg (2016). 'Post-Apartheid Labour Market Series 1994–2015' [Dataset]. Version 3.1 of the harmonized dataset based on Statistics SA's OHS, LFS and QLFS surveys, 1994–2015. Cape Town: DataFirst [data producer and distributor], 2016.

Kwend, P., and M. Ntuli (2015). 'A Detailed Decomposition Analysis of the Public–Private Sector Wage Gap in South Africa'. IZA Discussion Paper 9271. Bonn: Institute for the Study of Labor.

Lam, D., A. Finn, and M. Leibbrandt (2015). 'Schooling Inequality, Returns to Schooling, and Earnings Inequality: Evidence from Brazil and South Africa'. WIDER Working Paper 2015/050. Helsinki: UNU-WIDER.

Naidoo, K., B. Stanwix, and D. Yu (2014). 'Reflecting on Racial Discrimination in the Post-Apartheid South African Labour Market'. DPRU Policy Briefs 14/39. Cape Town: Development Policy Research Unit, University of Cape Town.

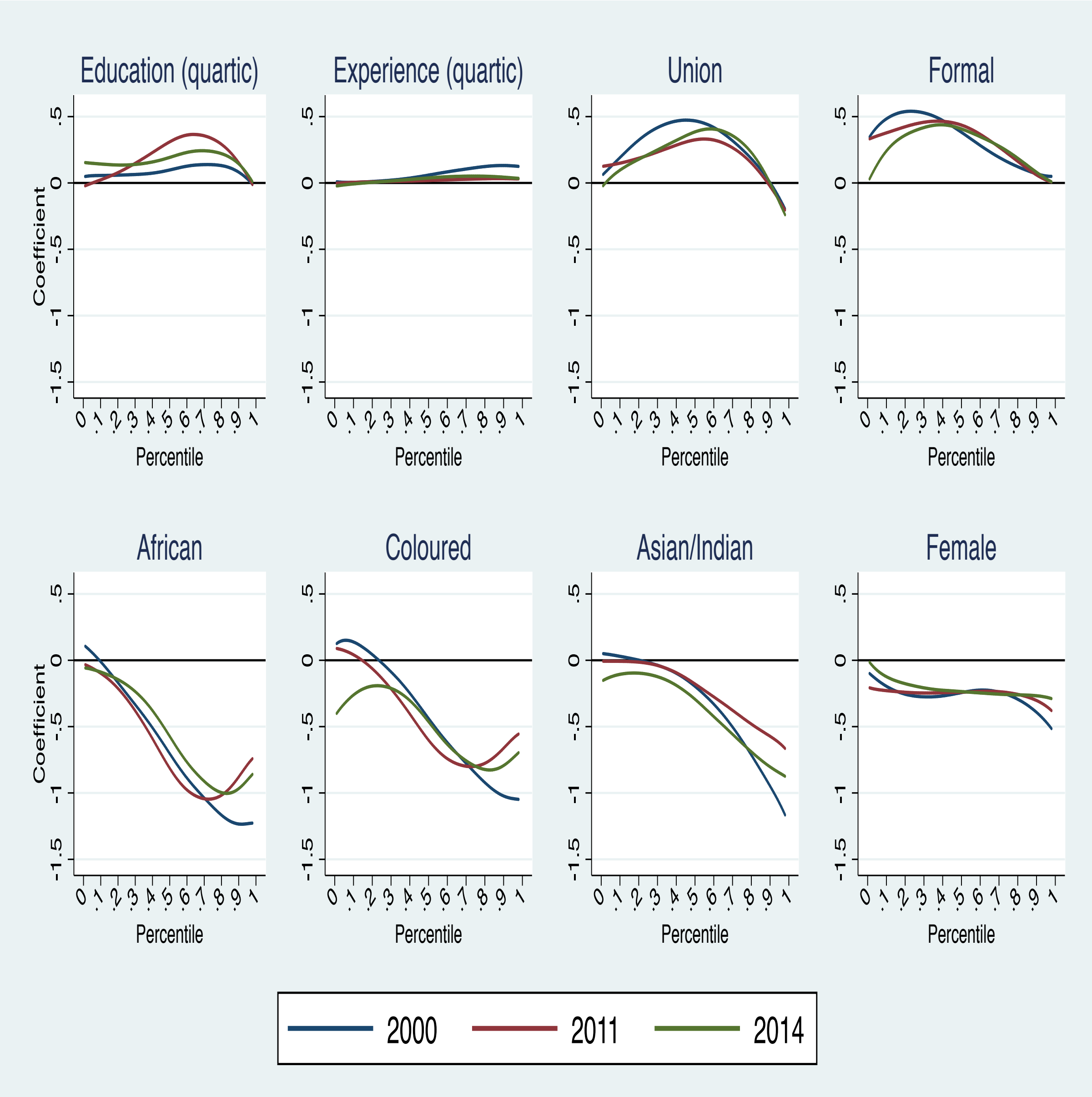
Oaxaca, R. (1973). 'Male–Female Wage Differentials in Urban Labor Markets'. *International Economic Review*, 14(3): 693–709.

Wittenberg, M. (2017a). 'Wages and Wage Inequality in South Africa 1994–2011: Part 1—Wage Measurement and Trends'. *South African Journal of Economics*, 85(2): 279–97.

Wittenberg, M. (2017b). 'Wages and Wage Inequality in South Africa 1994–2011: Part 2—Inequality Measurement and Trends'. *South African Journal of Economics*, 85(2): 298–318.

Appendix

Figure A1: Coefficients from unconditional quantile regressions over the distribution of earnings



Source: authors' calculations using PALMS V3.1 dataset.

Table A1: Robustness checks—RIF regression decomposition of the Gini coefficient: 2000–14

	Monthly earnings	Without weights	Males only	Hourly earnings	Hourly for full-time
			Overall		
Post	63.43	64.35	63.22	63.94	63.04
	(0.37)	(0.37)	(0.52)	(0.38)	(0.40)
Pre	55.20	55.80	54.15	57.14	55.56
	(0.57)	(0.36)	(0.81)	(0.63)	(0.51)
Difference	8.23	8.56	9.07	6.80	7.49
	(0.68)	(0.52)	(0.96)	(0.73)	(0.65)
Endowments	0.25	0.36	2.33	0.80	0.37
	(0.25)	(0.26)	(0.36)	(0.29)	(0.27)
Structure	7.99	8.19	6.74	6.00	7.12
	(0.65)	(0.46)	(0.89)	(0.70)	(0.61)
			Endowments		
Education	0.42	0.75	1.15	0.79	0.71
	(0.24)	(0.28)	(0.34)	(0.28)	(0.28)
Potential experience	–0.08	–0.08	–0.10	–0.10	–0.09
	(0.03)	(0.03)	(0.07)	(0.04)	(0.05)
Union	0.45	0.03	0.61	0.44	0.35
	(0.06)	(0.03)	(0.10)	(0.06)	(0.05)
Formal status	0.07	–0.10	0.47	0.09	0.02
	(0.02)	(0.02)	(0.06)	(0.02)	(0.02)
Race	–0.17	0.28	–0.02	–0.15	–0.05
	(0.05)	(0.06)	(0.06)	(0.05)	(0.04)
Gender	–0.20	–0.15		–0.19	–0.24
	(0.05)	(0.05)		(0.05)	(0.05)
Urban	–0.19	–0.20	–0.10	–0.24	–0.22
	(0.04)	(0.06)	(0.06)	(0.04)	(0.05)
Sector	–0.06	–0.18	0.31	0.17	–0.11
	(0.13)	(0.18)	(0.18)	(0.12)	(0.15)
			Structure		
Education	11.17	16.56	14.54	7.37	9.32
	(4.06)	(3.02)	(6.30)	(3.32)	(3.61)
Potential experience	–2.53	–5.22	–4.52	–5.07	–3.76
	(3.87)	(2.70)	(5.94)	(3.89)	(3.78)
Union	–2.09	–1.36	–2.51	–2.91	–2.22
	(1.22)	(1.07)	(1.41)	(1.35)	(1.09)
Formal status	–0.22	–0.05	0.16	0.29	0.14
	(0.21)	(0.20)	(0.21)	(0.22)	(0.18)
Race	3.47	8.36	10.38	–1.96	1.51
	(3.13)	(3.60)	(4.60)	(2.77)	(2.84)
Gender	0.62	0.50		0.06	–0.49
	(0.62)	(0.52)		(0.64)	(0.55)
Urban	–0.62	–0.34	–1.04	–0.86	–0.85
	(0.23)	(0.30)	(0.32)	(0.29)	(0.25)
Sector	–0.87	–3.35	0.84	0.57	–0.23
	(2.80)	(2.15)	(3.79)	(2.79)	(2.67)
Constant	–0.95	–6.90	–11.12	8.50	3.68
	(5.37)	(5.58)	(7.60)	(5.75)	(5.38)
			Total		
Education	11.59	17.31	15.69	8.16	10.04
	(4.00)	(2.94)	(6.22)	(3.24)	(3.53)
Potential experience	–2.61	–5.30	–4.63	–5.17	–3.85
	(3.87)	(2.69)	(5.93)	(3.88)	(3.76)
Union	–1.65	–1.33	–1.90	–2.48	–1.87
	(1.24)	(1.07)	(1.46)	(1.38)	(1.12)
Formal status	–0.15	–0.15	0.63	0.38	0.16
	(0.22)	(0.20)	(0.24)	(0.22)	(0.18)
Race	3.30	8.65	10.36	–2.10	1.46
	(3.14)	(3.57)	(4.61)	(2.79)	(2.86)
Gender	0.42	0.36		–0.13	–0.73
	(0.65)	(0.56)		(0.67)	(0.59)

Urban	–0.81	–0.54	–1.13	–1.10	–1.07
	(0.21)	(0.26)	(0.29)	(0.27)	(0.22)
Sector	–0.92	–3.53	1.16	0.74	–0.34
	(2.78)	(2.10)	(3.74)	(2.76)	(2.63)
N	85,996	85,996	44,861	84,041	73,489

Note: robust standard errors in parentheses.
Source: authors' calculations using PALMS V3.1 dataset.

Table A2: Robustness checks—RIF regression decomposition of the Gini coefficient: 2000–11

	Monthly earnings	Without weights	Males only	Hourly earnings	Hourly for full-time
			Overall		
Post	56.30	56.42	56.12	57.64	56.21
	(0.20)	(0.17)	(0.29)	(0.29)	(0.22)
Pre	55.20	55.80	54.15	57.14	55.56
	(0.57)	(0.36)	(0.81)	(0.63)	(0.51)
Difference	1.10	0.63	1.96	0.49	0.65
	(0.61)	(0.40)	(0.86)	(0.69)	(0.56)
Endowments	–1.48	–2.34	–0.17	–0.84	–1.43
	(0.15)	(0.14)	(0.23)	(0.17)	(0.16)
Structure	2.58	2.97	2.14	1.34	2.08
	(0.60)	(0.38)	(0.85)	(0.68)	(0.55)
			Endowments		
Education	–0.98	–1.16	–0.58	–0.57	–0.92
	(0.12)	(0.12)	(0.17)	(0.18)	(0.13)
Potential experience	–0.05	–0.05	–0.15	–0.16	–0.07
	(0.08)	(0.07)	(0.14)	(0.12)	(0.10)
Union	0.36	–0.07	0.53	0.38	0.31
	(0.04)	(0.03)	(0.06)	(0.05)	(0.04)
Formal status	0.28	0.11	0.57	0.36	0.27
	(0.03)	(0.02)	(0.05)	(0.04)	(0.03)
Race	0.04	0.35	0.13	0.05	0.06
	(0.02)	(0.04)	(0.05)	(0.03)	(0.02)
Gender	–0.13	–0.12		–0.14	–0.13
	(0.02)	(0.02)		(0.03)	(0.02)
Urban	–0.44	–0.52	–0.38	–0.49	–0.50
	(0.04)	(0.04)	(0.07)	(0.06)	(0.05)
Sector	–0.56	–0.89	–0.30	–0.27	–0.45
	(0.09)	(0.08)	(0.13)	(0.08)	(0.08)
			Structure		
Education	13.77	22.16	16.93	9.43	12.20
	(3.87)	(2.73)	(6.00)	(3.12)	(3.38)
Potential experience	–11.19	–13.09	–18.69	–14.41	–11.68
	(3.70)	(2.29)	(5.63)	(3.87)	(3.68)
Union	–2.55	–1.69	–3.07	–2.79	–2.46
	(1.09)	(0.89)	(1.24)	(1.32)	(0.95)
Formal status	–0.10	–0.04	0.15	0.60	0.33
	(0.19)	(0.15)	(0.18)	(0.21)	(0.16)
Race	6.03	17.62	8.03	–0.68	3.41
	(2.79)	(2.94)	(4.20)	(2.59)	(2.48)
Gender	0.66	0.48		–0.01	–0.12
	(0.55)	(0.39)		(0.61)	(0.47)
Urban	–0.06	0.36	–0.34	–0.30	–0.21
	(0.21)	(0.25)	(0.31)	(0.30)	(0.23)
Sector	–0.79	–2.95	1.43	1.60	0.07
	(2.62)	(1.77)	(3.51)	(2.65)	(2.46)
Constant	–3.20	–19.89	–2.31	7.89	0.55
	(5.00)	(4.81)	(7.01)	(5.53)	(5.08)
			Total		
Education	12.79	21.00	16.36	8.86	11.27
	(3.85)	(2.71)	(5.99)	(3.10)	(3.36)
Potential experience	–11.24	–13.14	–18.83	–14.57	–11.75

	(3.72)	(2.31)	(5.67)	(3.88)	(3.72)
Union	–2.19	–1.76	–2.54	–2.41	–2.15
	(1.10)	(0.89)	(1.25)	(1.34)	(0.96)
Formal status	0.18	0.07	0.72	0.96	0.60
	(0.20)	(0.15)	(0.20)	(0.23)	(0.17)
Race	6.07	17.97	8.16	–0.63	3.47
	(2.79)	(2.93)	(4.19)	(2.59)	(2.47)
Gender	0.53	0.36		–0.15	–0.26
	(0.56)	(0.40)		(0.63)	(0.48)
Urban	–0.50	–0.15	–0.72	–0.79	–0.71
	(0.20)	(0.23)	(0.28)	(0.27)	(0.21)
Sector	–1.35	–3.84	1.13	1.33	–0.38
	(2.61)	(1.76)	(3.50)	(2.64)	(2.45)
N	89,459	89,459	47,388	87,733	77,552

Note: robust standard errors in parentheses.
Source: authors' calculations using PALMS V3.1 dataset.

Table A3: Robustness checks—RIF regression decomposition of the Gini coefficient: 2011–14

	Monthly earnings	Without weights	Males only	Hourly earnings	Hourly for full-time
			Overall		
Post	63.43	64.35	63.22	63.94	63.04
	(0.37)	(0.37)	(0.52)	(0.38)	(0.40)
Pre	56.30	56.42	56.12	57.64	56.21
	(0.20)	(0.17)	(0.29)	(0.29)	(0.22)
Difference	7.13	7.93	7.10	6.31	6.84
	(0.42)	(0.40)	(0.60)	(0.47)	(0.45)
Endowments	–0.32	–0.64	–0.07	–0.24	–0.30
	(0.07)	(0.09)	(0.10)	(0.07)	(0.07)
Structure	7.46	8.57	7.17	6.54	7.14
	(0.43)	(0.44)	(0.60)	(0.48)	(0.46)
			Endowments		
Education	0.01	–0.12	0.07	0.04	0.08
	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)
Potential experience	0.01	0.02	0.01	0.02	0.00
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Union	0.05	0.11	0.02	0.07	0.02
	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)
Formal status	–0.18	–0.21	–0.10	–0.20	–0.22
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Race	–0.18	–0.47	–0.08	–0.15	–0.10
	(0.05)	(0.07)	(0.06)	(0.05)	(0.05)
Gender	–0.06	–0.03		–0.06	–0.07
	(0.02)	(0.01)		(0.02)	(0.02)
Urban	0.05	0.05	0.02	0.06	0.04
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Sector	–0.04	0.02	–0.01	–0.01	–0.04
	(0.04)	(0.03)	(0.05)	(0.03)	(0.03)
			Structure		
Education	–1.21	–3.56	–0.73	–0.75	–1.31
	(1.44)	(1.38)	(2.22)	(1.58)	(1.45)
Potential experience	8.62	7.82	14.19	9.38	7.90
	(2.74)	(2.13)	(4.05)	(3.07)	(3.06)
Union	0.48	0.32	0.61	–0.13	0.26
	(0.80)	(0.75)	(1.03)	(0.92)	(0.77)
Formal status	–0.15	–0.02	0.01	–0.38	–0.22
	(0.19)	(0.19)	(0.22)	(0.24)	(0.16)
Race	–2.59	–8.85	2.28	–1.32	–1.91
	(1.89)	(2.31)	(2.65)	(2.10)	(1.92)
Gender	–0.05	0.03		0.08	–0.40
	(0.44)	(0.47)		(0.49)	(0.45)
Urban	–0.37	–0.43	–0.43	–0.37	–0.40

	(0.12)	(0.14)	(0.16)	(0.14)	(0.12)
Sector	0.46	0.29	0.03	–0.58	0.08
	(1.30)	(1.41)	(1.76)	(1.37)	(1.33)
Constant	2.25	12.99	–8.80	0.61	3.13
	(3.87)	(3.73)	(5.66)	(4.14)	(4.24)
			Total		
Education	–1.20	–3.68	–0.66	–0.70	–1.23
	(1.44)	(1.39)	(2.21)	(1.57)	(1.44)
Potential experience	8.63	7.83	14.21	9.40	7.90
	(2.74)	(2.13)	(4.05)	(3.07)	(3.06)
Union	0.54	0.43	0.63	–0.07	0.28
	(0.80)	(0.76)	(1.03)	(0.92)	(0.77)
Formal status	–0.33	–0.23	–0.09	–0.58	–0.44
	(0.17)	(0.17)	(0.21)	(0.22)	(0.15)
Race	–2.77	–9.32	2.20	–1.47	–2.01
	(1.92)	(2.36)	(2.69)	(2.13)	(1.95)
Gender	–0.10	–0.01		0.02	–0.47
	(0.46)	(0.48)		(0.50)	(0.46)
Urban	–0.32	–0.38	–0.41	–0.31	–0.36
	(0.13)	(0.15)	(0.17)	(0.15)	(0.13)
Sector	0.42	0.31	0.02	–0.58	0.04
	(1.30)	(1.41)	(1.75)	(1.37)	(1.32)
N	128,497	128,497	65,881	125,470	109,779

Note: robust standard errors in parentheses.
Source: authors' calculations using PALMS V3.1 dataset.