

WAGES AND WAGE INEQUALITY IN SOUTH AFRICA 1994–2011: PART 2 – INEQUALITY MEASUREMENT AND TRENDS

MARTIN WITTENBERG*

Abstract

We analyse the long-term trends in wage inequality in South Africa, using household survey data. We show that the trends in household income inequality are largely driven by changes in wage inequality. Given the detailed nature of our series we show that measurement issues and breaks in the series need to be dealt with in order to draw robust conclusions from the data. Most standard inequality measures show that wage inequality has increased over the period. Nevertheless the choice of measure matters, because there are different trends in different parts of the distribution. It appears that the distribution below the median has become more compressed, while the top of the wage distribution has moved away from the median. The inequality in the labour market translates into even higher inequality in society given that high earners tend to live together with other high earners while low wage individuals often end up sharing their incomes with the unemployed. Furthermore there are many South Africans with access to no wage income. Given the trends analysed here it is not surprising that overall inequality in South Africa has not come down or has even increased since the end of apartheid.

JEL Classification: C42, D31, J31

Keywords: Gini coefficient, race, survey data, imputation, household inequality

1. INTRODUCTION

South Africa is notorious for its level of inequality. Indeed it is a well-known fact that South Africa has one of the highest Gini coefficients in the world (see for instance figure 1.3, p. 22 in Atkinson, 2015). The prominence of the Gini coefficient in these discussions is such that South Africa's National Development Plan lists reducing the Gini coefficient as one of its key objectives (National Planning Commission, 2012, p. 34). In contrast to the hoped for reduction in inequality in the post-apartheid period, the accompanying article on wage trends (Wittenberg, in press) suggests that the Gini

* Corresponding author: DataFirst, SALDRU and School of Economics, University of Cape Town. E-mail: martin.wittenberg@uct.ac.za

Some of the original data work was done with the support of an infrastructure grant to DataFirst from the Redi3x3 project on "Employment/Unemployment, Income Distribution and Inclusive Growth." The Vice-Chancellor's Strategic Fund of the University of Cape Town paid for the initial construction of the PALMS dataset without which this research would not have been possible. The ILO commissioned some of the initial analyses of wage trends, which are reported in Wittenberg (2014a). I would like to acknowledge the helpful comments and support from Patrick Belser and Kristen Sobek in that work. Andrew Kerr and Arden Finn read the original paper carefully and made useful suggestions. I would also like to thank the editor of the SAJE Steve Koch and two anonymous referees for helpful comments.

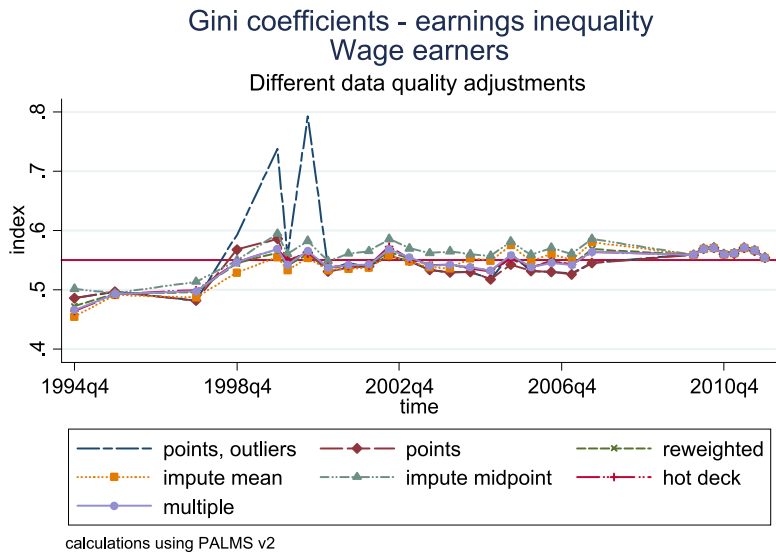


Figure 1. Different ways of dealing with missing data or the outliers have a marked impact on the measure of inequality [Colour figure can be viewed at wileyonlinelibrary.com]

coefficient of wage inequality in fact increased over the period 1994–2011. Other authors (e.g. Leibbrandt *et al.*, 2010) have also commented how stubbornly high the Gini coefficient has remained since the advent of democracy.

The persistence of high inequality in a context where the institutional underpinnings of discrimination are being removed serves as one of the key puzzles and challenges of the post-apartheid era. There are three questions in particular:

- How well measured are the inequality trends? Given the measurement issues raised in the accompanying paper (Wittenberg, in press), what are the implications for the measurement of inequality? What are the key “stylised facts” about wage inequality in South Africa?
- If the increases in wage inequality are real, how do they relate to inherited inequalities, in particular racial differences and differences in access to human capital?
- How do the dynamics in the labour market transmit themselves to income inequality more generally, given that the latter has traditionally been the primary concern?

In this paper we examine these issues. Our main contribution is threefold. Firstly we highlight the importance of measurement issues, as we do in the accompanying paper. Our discussion extends that work in several ways: we show that the wage measurement issues discussed in that paper matter for the measurement of inequality; in addition we note that there appears to be a marked break in the wage inequality series between 1997 and 1998; and when considering income inequality there is a major anomaly in the data for 1999 as there seem to be too many households without any access to income. These breaks bedevil attempts to analyse trends in inequality over time. We also show that how one measures inequality matters for the type of trends that one observes; in particular the “per capita” conversion typically used to “equivalise” incomes within households probably overstates the extent of inequality.

Our second major contribution is to place many of the existing inequality estimates into a much longer time run. We show that estimates produced by Leibbrandt *et al.* (2010) and Bhorat *et al.* (2009) are generally consistent with the wage inequality trends analysed here, although they were estimated from different datasets. However by looking at the entire series we can get a better sense of how robust the “trends” are.

Our third contribution is to point to off-setting trends in the top part and the bottom part of the income distribution. Inequality seems to have increased in the top half of the distribution while being reduced in the bottom half.

We begin our discussion with a brief look at the literature. Section 3 contains a discussion of different measures of inequality. We show that all measures suggest a big shift in the series between October 1997 and October 1998. Different measures however suggest different trajectories after 2000, with some suggesting increases in wage inequality while others suggesting stable or perhaps decreasing levels. Section 4 focuses on the Theil measure and decomposes inequality by race and then by education categories. We show that wage inequality within race groups and within education categories explains more than half of all inequality. In order to make sense of the different patterns emerging from the summary measures we turn to look at the evolution of wages at different points of the distribution in section 5. This suggests that inequality has increased at the top of the distribution but decreased at the bottom. Furthermore the pattern by education categories is heterogeneous. In the following section we amplify on this discussion by focusing on the shares going to different percentiles. It seems that the top 10% of wage earners have increased their share of the total wage bill, while the bottom 50% have lost ground. In the final section we show that the way wage earners sort into households increases overall inequality. It turns out that the trends in overall inequality are driven by the trends in wage inequality.

2. LITERATURE REVIEW

There is a voluminous literature on inequality in South Africa. Leibbrandt *et al.* (2010) and Leibbrandt *et al.* (2012) are among the most comprehensive reviews published recently. Van der Berg (2011) situates the contemporary inequality measures in the broader sweep of South African history. Bhorat *et al.* (2009) discuss both income inequality and non-income inequality. Overall there seems to be agreement that inequality has remained high, or has even increased in the post-Apartheid period (Bhorat *et al.* 2009; Leibbrandt *et al.* 2010). Van der Berg (2011, Table 5, p. 132) suggests, however, that a notable increase in inequality occurred between 1994 and 2000, with stable levels of inequality thereafter. We will reflect on those findings in the light of our empirical analysis below.

These studies rely on estimates at a few points in time. Leibbrandt *et al.* (2012), for example, compare the data from the 1993 Project for Statistics on Living Standards and Development to those from the 2008 National Income Dynamics Survey. Leibbrandt *et al.* (2010) use the same two datasets as well as the 2000 Income and Expenditure Survey. Bhorat *et al.* (2009) compare the 1995 and the 2005 Income and Expenditure Surveys. Van der Berg (2011) compares the 1995, 2000 and 2005 Income and Expenditure Surveys as well as the 1996 and 2001 censuses and the 2007 Community Survey.

The reliance on data sources which have strong income and expenditure modules makes sense in a context where the primary objective is the estimation of overall income

inequality from all sources. The discussion of wage inequality in this context is restricted to a consideration of how inequality in household access to wages translates into overall inequality. By contrast our analysis is based primarily on wage information. This allows us to use 29 separate surveys and as a result we achieve a much more fine-grained picture of the evolution over time. Indeed it turns out that the estimates from the income and expenditure surveys align quite neatly with our series as we show in section 7 below. In the context of the longer time run and with the more granular time pattern it becomes evident that some of the documented upward shifts coincide with what appear to be measurement shifts.

Van der Berg's discussion of the census and community survey information stands apart from the other series. Unfortunately the census income measures are extremely noisy (they are recorded only in bands). The personal income information has to be combined within the household. This is done by point-imputation within the band (see Yu, 2009). Yu imputed for zero and missing incomes using a chained regression imputation approach. Unfortunately, the biggest source of noise in the data is likely to be the single point imputation for people providing the bracket information. One particular issue is that incomes for the "open category" were set at twice the bracket boundary (see Tables 3 and 7 in Yu, 2009). A "spike" that far up the income distribution will have serious effects on inequality measures as we will show when we consider the impact of mid-point imputations on inequality measures on the wage data. A further problem with the census inequality calculations, is that the two sets of brackets are not the same. It is very unclear that any inequality measures calculated on such differently spiky data can be meaningfully compared.

The existing literature has commented on the centrality of labour markets to overall income inequality. As Leibbrandt *et al.* note: "All in all, the labour market is shown to sit centre-stage as the driver of South African income inequality" (2010, p. 19). In their detailed decompositions of inequality by income source Leibbrandt *et al.* (2010, pp. 34-35) provide Gini coefficients for labour income of 0.56 for 1993, 0.65 for 2000 and 0.64 for 2008. These show very high levels of inequality. Indeed they also suggest rising labour market inequality in the post-Apartheid period.

It is important to understand properly the nature of these estimates: they are all labour income *per capita* for people living in households with labour income. Leibbrandt *et al.* (2001) provide a Gini coefficient of 0.53 for household labour income per adult equivalent in 1995. None of these estimates are directly comparable to the Gini coefficients for individual wage income shown in section 3.2, although they are comparable to those reported in section 7, where we aggregate wage income within the household.

There are some other differences. The Leibbrandt *et al.* (2010) figures include all forms of earnings from labour market activities, including self-employment, whereas the Leibbrandt *et al.* (2001) analysis separates wage income from self-employment income. Because of the break in recorded self-employment between the October Household Surveys and Labour Force Surveys noted in the accompanying article (Wittenberg, in press), we are analysing only wage income in this discussion. To that extent the most comparable figures to those reported below are those for 1995 given by Leibbrandt *et al.* (2001). Other comparable figures are those given by Bhorat *et al.* (2009), which also separate out wage and self-employment incomes. Their estimates of *per capita* wage income inequality for 1995 and 2005 are 0.72 and 0.81, respectively, but these estimates seem to include individuals with zero incomes. We consider all these estimates in more detail in section

7. In that section we will provide a new picture of how wage inequality together with “assortative matching” within households contribute to the high levels of income inequality that we observe.

One of the key concerns of the inequality literature has been to document whether race continues to be one of the main drivers of income inequality or whether this is being replaced by differences in human capital. Many studies have decomposed inequality measures (mainly the Theil index) by race or education (*e.g.* Bhorat *et al.* 2009, Table 3). In general the pattern has been that the “within race” inequality measures have increased in importance over time. This is what one would expect if legalised discrimination is becoming less prevalent and salient. Nevertheless these patterns have again been identified by comparing two or three data points over time, whereas on our wage data we can do so on a finer time scale. While the broad trends will be shown to hold up, there are again troubling measurement shifts also.

The Gini coefficient has been central to much of the South African inequality literature, both academic and policy related, except when the Theil index has been used to decompose the inequality measures. Little attention has been paid to how sensitive the trends are to different measures. We will produce some evidence on this below. Indeed we will argue that the single-minded focus on the Gini has obscured different shifts in different parts of the distribution. Furthermore much of the literature has focused on a *per capita* definition of income within the household. The existing consensus seems to be that this is an innocent choice, as argued, for instance by Bhorat *et al.* (2009, p. 3, citing Leibbrandt and Woolard). We will show below that this choice does not affect the **trends** but it certainly affects the measured **level** of income inequality.

3. MEASURING INEQUALITY

3.1 Data and Data Adjustments

The dataset for our analysis is PALMS version 2.1 (Kerr *et al.*, 2013), which combines information from the October Household Surveys, Labour Force Surveys and Quarterly Labour Force Surveys (Statistics South Africa – various years). A more detailed discussion on this dataset is contained in the accompanying article (Wittenberg, in press). The earnings data cannot be used as is, because many individuals only provided bracket information and some did not provide any income at all. Furthermore the information is contaminated by outliers. The accompanying article gives a detailed discussion of some of the possible options for dealing with these issues:

- Use only the existing Rand information ignoring outliers, the bracket information and the missing data. This is not a sensible strategy as the outliers (although only a few in number) have a marked impact on aggregate trends, as shown in Fig. 1. In subsequent figures we will not plot this series.
- Clean out the outliers, but otherwise use only the Rand information ignoring brackets and missing data. This will still produce biased results, since people who respond with bracket information or who refuse outright tend to have higher incomes.
- Impute bracket-means for individuals who provide only bracket information. This provides a reasonable adjustment if one is only concerned about the trends in the means (and is willing to ignore the completely missing information), but creates

“spikes” in the distribution which distort other statistics, such as the dispersion (and hence inequality measures) and estimates of the percentiles.

- Impute mid-points of brackets for individuals with only bracket information. This has the drawbacks of the previous approach but also seems to exaggerate income inequality. This can be seen in Fig. 1, where the Gini coefficients estimated with the mid-point imputations are clearly larger than the other series.
- Use the Rand information (cleaned of outliers) but reweight the observations to account for the missing bracket values. This approach works almost as well as the multiple imputation one, except that it doesn't deal with the cases of the completely missing information.
- Replace the missing information (brackets and completely missing) by taking a draw from the empirically observed cases that match the missing ones, on gender, race, education and occupation. This “hot deck” approach does not get the standard errors right on the estimates, because it assumes that the randomly drawn information is actually measured.
- Taking draws from the empirical distribution multiple times and then calculating the covariance matrix of the estimates according to “Rubin's rules.” This is likely to be the best of all the available options. Fortunately a set of ten draws from the empirical distribution was released with the PALMS dataset. This is the data that we prefer to use in our analysis.

3.2 *Inequality Indices*

We will not rehearse the theory of inequality measurement here (for discussions see Atkinson 1970; Deaton 1997, Chapter 3; Cowell 2000). The standard approach requires that the inequality ranking of different distributions satisfy a range of axioms such as anonymity (the identity of who holds what income should not matter); the population principle (combining two populations with identical distributions should not alter the inequality measure); the principle of transfers (redistributing money from a rich person to a poorer while leaving their relative ranking unchanged should reduce measured inequality); and scale independence. All of the standard inequality measures satisfy these axioms. In cases where one distribution “Lorenz dominates” another they will also agree on which of the distributions is more unequal. They will, however, provide different rankings in cases where the Lorenz curves cross.

(i) *The Gini Index* The most well-known inequality index is the Gini coefficient, which can be defined as twice the area between the Lorenz curve and the line of perfect equality. The inequality measures for the South African wage data using this index are shown in Fig. 1 and (in more detail) in the accompanying paper. Although there is some noise in the estimates, the measures made post March 2007 are noticeably higher than those made during the period from 2000 to 2006. Inequality before 1998 is at a much lower level still.

(ii) *The Atkinson Family of Indices* Atkinson (1970) argued that inequality indices could be derived implicitly from social welfare functions. By considering how a “social planner” might trade off improvements in aggregate income against inequality, he defined a set of inequality measures which can be characterised by their “inequality aversion” parameter ϵ . The formulae are

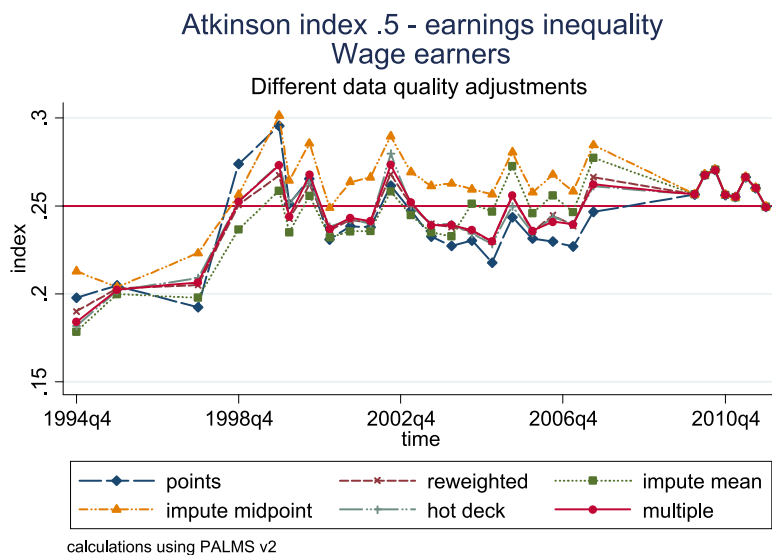


Figure 2. Wage inequality according to the Atkinson index with $\epsilon = 0.5$ [Colour figure can be viewed at wileyonlinelibrary.com]

$$I = 1 - \left(\frac{1}{N} \sum_{i=1}^N \left[\frac{x_i}{\mu} \right]^{1-\epsilon} \right)^{\frac{1}{1-\epsilon}} \quad \text{if } \epsilon \neq 1$$

$$I = 1 - \prod_{i=1}^N \left(\frac{x_i}{\mu} \right)^{\frac{1}{N}} \quad \text{if } \epsilon = 1$$

where x_i is the income of the i -th individual and μ is mean income. The larger ϵ is, the more the inequality measure pays attention to the distribution of income at the bottom.

The empirical estimates on our wage data for $\epsilon = 0.5$, 1 and 2 are shown in Figs. 2–4, respectively. All estimates were calculated using the DASP package (Araar and Duclos, 2013). Estimates for the multiple imputation case were obtained for each imputed dataset and then averaged.

In each Figure the horizontal line is drawn at the “average” inequality level for the period as a whole, calculated by pooling all datasets. Several points are immediately apparent. Firstly, with a low inequality aversion parameter, the trend points to an unambiguous **increase** in inequality over the period (Fig. 2). However, with high inequality aversion, it now looks as though wage inequality at the end of the period is actually **lower** than around the year 2000 (Fig. 4). Clearly there are shifts in the distribution which pull the indices in different directions. We will return to this issue below.

Secondly, it is clear that with higher inequality aversion the differences between different data quality adjustments become more pronounced. In particular in Fig. 4 we see big differences between the “Rand only” series and our preferred estimates (the multiple imputation one). Clearly the omission of some high income figures makes the “Rand only” series look more equal than it really is. Furthermore across all measures we again see that the “mid-point imputation” series seems to find higher levels of inequality than the “reweighted” or the “multiple imputation” ones. The latter two are again very close together.

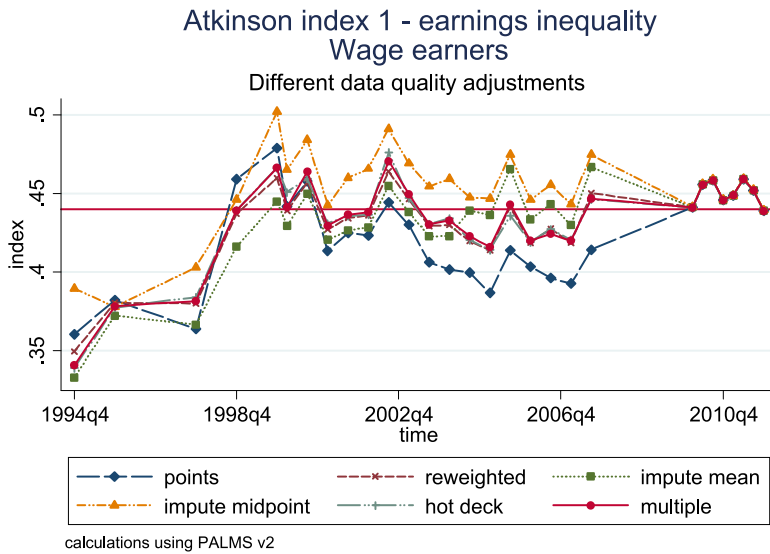


Figure 3. Wage inequality according to the Atkinson index with $\varepsilon = 1$ [Colour figure can be viewed at wileyonlinelibrary.com]

(iii) *The Theil Index* Another family of indices is given by the general entropy class (see Duclos and Araar, 2006, pp. 67-69). We will estimate only the special case where the parameter θ takes on the value 1. This is also known as the Theil index, (sometimes called Theil-T) with formula

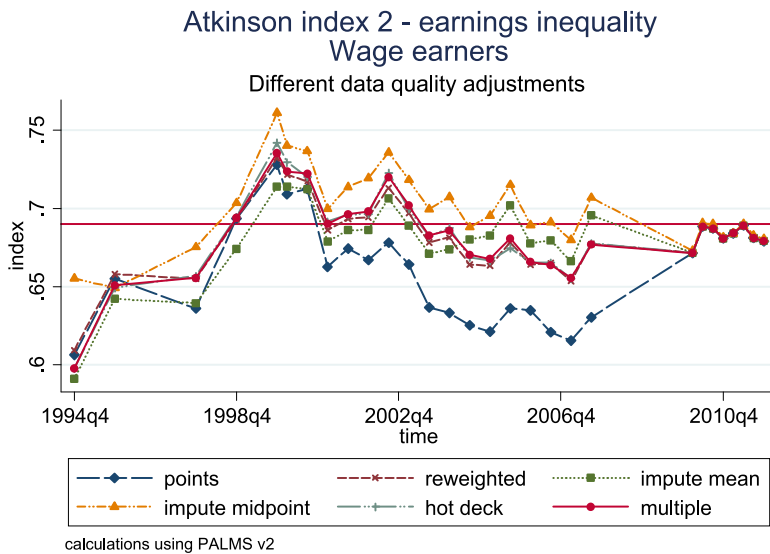


Figure 4. Wage inequality according to the Atkinson index with $\varepsilon = 2$ [Colour figure can be viewed at wileyonlinelibrary.com]

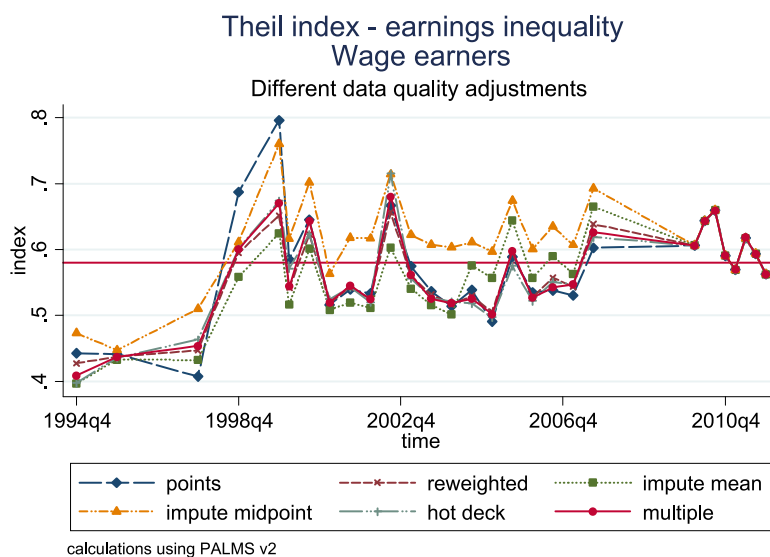


Figure 5. Wage inequality using the Theil T index of inequality [Colour figure can be viewed at wileyonlinelibrary.com]

$$I_T = \frac{1}{N} \sum_{i=1}^N \frac{x_i}{\mu} \log \left(\frac{x_i}{\mu} \right)$$

where μ is again mean income. The estimates (also obtained using DASP – see Araar and Duclos, 2013) are graphed in Fig. 5. In this case the estimates during the early 2000s again show considerable fluctuations, with some evidence of higher inequality in the QLFS period.

(iv) *Comparing Different Measures* Comparing the Theil and the Gini indices to the Atkinson measures, it appears that they are akin to measures with low inequality aversion. In the case of the Gini this corresponds to Atkinson's judgement (1970, p. 261): “two of the conventional measures (the Gini coefficient and the coefficient of variation) tend to give rankings which are similar to those reached with a relatively low degree of inequality aversion – ε of the order of 1.0 or less.” It thus appears that trends in inequality using the Gini may be slightly different to those obtained with an Atkinson index with higher inequality aversion.

It is worth noting that **all** of the inequality measures suggest much lower levels of wage inequality up to 1997. This seems to bear out van der Berg's assessment of a steep rise in inequality between 1994 and 2000. It is troubling, however, that the rise seems to be concentrated largely between 1997 and 1998. This suggests a measurement change rather than a substantive one. Indeed the early OHSs systematically under-sampled small households and under-sampled low income earners (*e.g.* domestic workers), as documented by Kerr and Wittenberg (2015). This would suggest that inequality in the early OHSs may be underestimated and would support the measurement change hypothesis.

Van der Berg's evidence for a dramatic increase in inequality between 1994 and 2000 comes from two sources: a comparison between the 1995 IES and the 2000 IES and a comparison between the 1996 census and the 2001 census. The two IES samples were

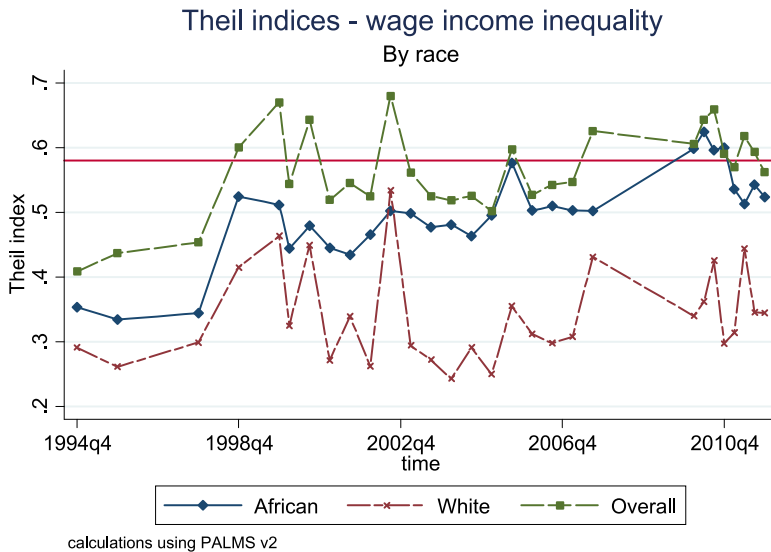


Figure 6. Wage income inequality within the African group has increased over the period [Colour figure can be viewed at wileyonlinelibrary.com]

linked to the respective OHS and LFS samples, so the 1995 IES will also be affected by the sampling problems identified by Kerr and Wittenberg. The census measures, however, would be independent of these. As we noted in the literature review, however, the census measures are derived from incompatible income bands which are collapsed by single-point imputation to Rand values, then aggregated up to household level, supplemented by imputation for zero incomes and missing data. This information is unlikely to be all that accurate.

Nevertheless Leibbrandt *et al.* (2010) also identify an increase in inequality between 1993 and 2000. Their baseline is the 1993 PSLSD, which has much higher quality information than the census. This suggests that even if wage inequality is underestimated in the early OHSs the increase between those surveys and the measures recorded in the 2000s may nonetheless be real. We leave the question of estimating the size of the increase (if any) to future research.

4. DECOMPOSING WAGE INEQUALITY BY RACE AND EDUCATION

One of the advantages of the Theil index is that it can be decomposed by sub-groups. Leibbrandt *et al.* (2001) and Leibbrandt *et al.* (2010) perform such decompositions for overall personal income by race. The pattern they find is very clear: between race inequality is coming down and within race inequality has increased significantly. Bhorat *et al.* (2009) also perform such decompositions for 1995 and 2005 and come to the opposite conclusion. We perform the same decompositions using the wage distribution data.

4.1 Decomposition by Race

In Fig. 6 we show that wage income inequality according to the Theil index has increased markedly within the African group. At the end of the period the index was in excess of 0.5 and close to the overall index of inequality, while it was below these levels earlier. In

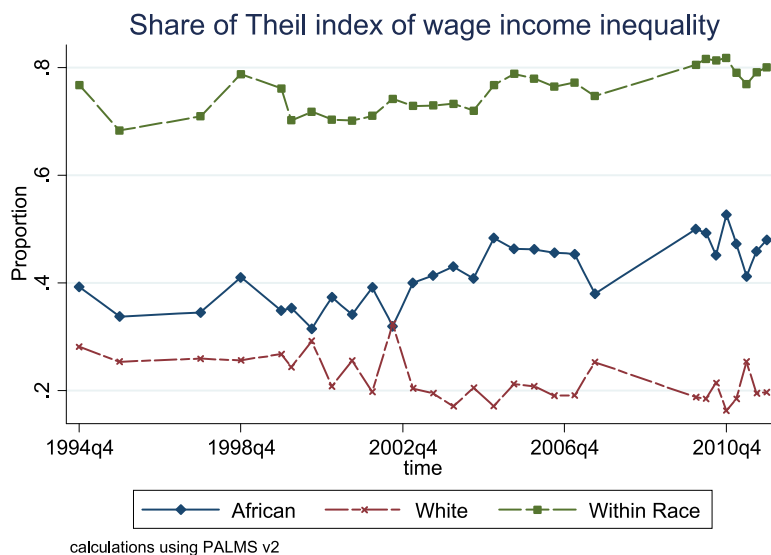


Figure 7. The “within race” share of income inequality has increased at least since 2000. Inequality within the African group contributes an increasing portion to overall inequality [Colour figure can be viewed at wileyonlinelibrary.com]

Fig. 7 we graph out the evolution of the share of overall inequality (as measured by the Theil) which is contributed by the “within race” inequalities. It is striking that the contribution of inequality within the White group is dropping. This is not due to decreasing inequality within this group (Fig. 6 suggests that this is holding steady), but due to the fact that it is becoming a smaller portion of the overall work force. It should be remembered that self-employed professionals are not included in these statistics. Fig. 7 also indicates that “within race” inequality accounts for an increasing share of overall inequality, at least since 2000. These results are obviously not directly comparable to the decompositions of *per capita* household income, they do suggest that the Leibbrandt *et al.* (2010) results are closer to the mark for the long-run trajectory of inequality than those of Bhorat *et al.* (2009).

4.2 Decomposition by Education

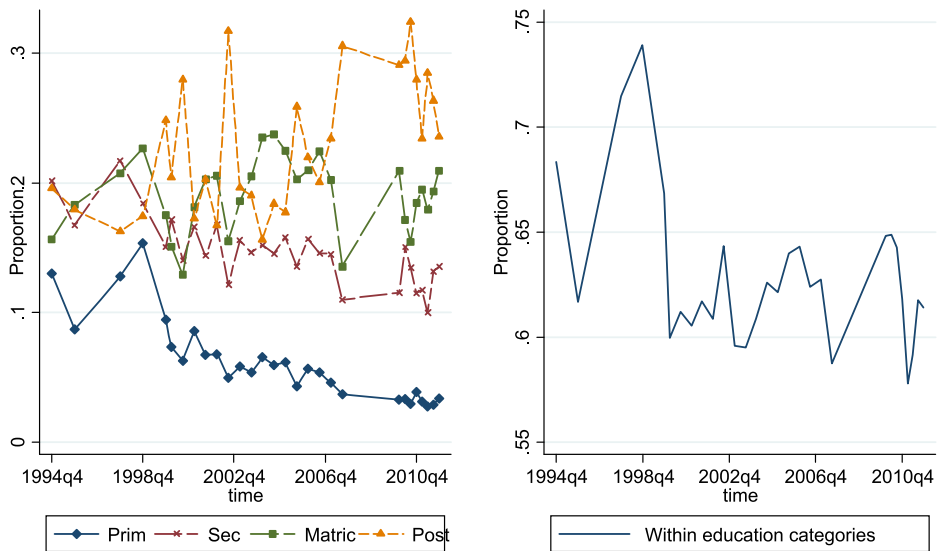
Since we are investigating the distribution of individual wage income, it is natural to investigate whether increasing inequality is being driven by widening differentials between individuals with differing skills and education levels. In Fig. 8 we show a decomposition of overall inequality (as measured by the Theil index) by education categories. We have grouped wage earners into four groups:

- Those with no education or at most primary school.
- Incomplete secondary school education.
- Matric.
- More than matric.

The left-hand panel of Fig. 8 shows what portion of overall inequality can be attributed to inequality within each of the different groups. Strikingly around a quarter of overall inequality is attributable to inequality of earnings among those with a post-matric qualification. Individuals with the least education are making up a shrinking proportion

Decomposing the Theil index of wage income inequality

Share of Index by Education categories



calculations using PALMS v2

Figure 8. Inequality of wages within the post-matric category is accounting for an increasing share of overall inequality [Colour figure can be viewed at wileyonlinelibrary.com]

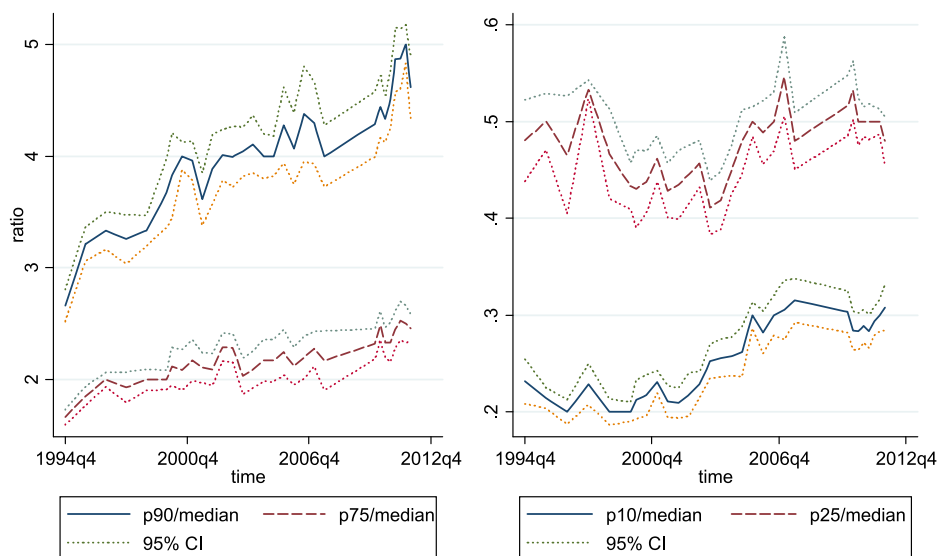
of the overall labour force, so inequality within this group is contributing a decreasing share to overall inequality. The right hand side panel is hard to read, because of the large fluctuations in the estimates, particularly in the early OHSs. It nonetheless shows that over the entire period inequality **within** these broad educational categories accounts for more than 60% of overall inequality, with little evidence that inequality **between** these categories is growing, at least if the OHS part of the series is ignored.

5. LOOKING AT OTHER POINTS IN THE DISTRIBUTION

One of the problems with inequality measures like the Gini coefficient and the Atkinson indices is that they summarise the character of the entire distribution in one statistic. An alternative is to explore changes in the distribution of earnings at different points in the distribution. Fig. 9 does this for the wage data. In this figure we graph various ratios: earnings of the 90th and the 75th percentiles relative to the median (left panel) and the 25th and 10th percentiles relative to the median (right panel). It is immediately apparent that earnings at the top end have increased strongly relative to the median – with much bigger gains at the 90th percentile than the 75th. Judging by the confidence intervals shown in those figures it is very unlikely that these trends reflect sampling noise. This, of course, explains the fact noted in the accompanying paper (Wittenberg, in press) that the mean has increased more rapidly over time than the median.

Looking at the bottom of the distribution, the right-hand panel in Fig. 9 suggests that the gap between the 25th percentile and the median has stayed constant over time, while

Wage inequality in PALMS



Rand earnings for bracket responses and outliers imputed.
Standard errors computed by clustered bootstrap

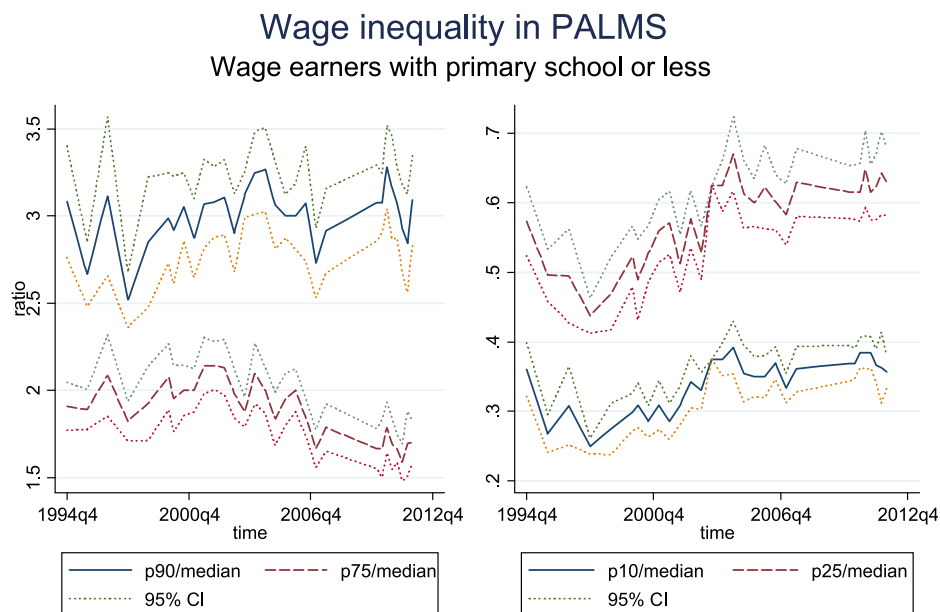
Figure 9. Wage increases of employees at different percentiles relative to the median [Colour figure can be viewed at wileyonlinelibrary.com]

the 10th percentile has shifted up. This suggests some compression of the earnings distribution right at the bottom. Presumably labour legislation, including minimum wages for domestics and agriculture workers, has had the effect of boosting wages among the lowest earners.

The contrasting trends at the top and the bottom of the distribution helps us to reconcile the divergent pictures provided by the Atkinson indices. With low or moderate levels of inequality aversion the increasing spread at the top of the distribution leads to the conclusion that overall inequality has increased. With high levels of inequality aversion the Atkinson index pays more attention to the bottom of the distribution and the compression of the distribution leads to the conclusion that inequality is decreasing. This shows why it is useful to rely not only on inequality measures in judging trends.

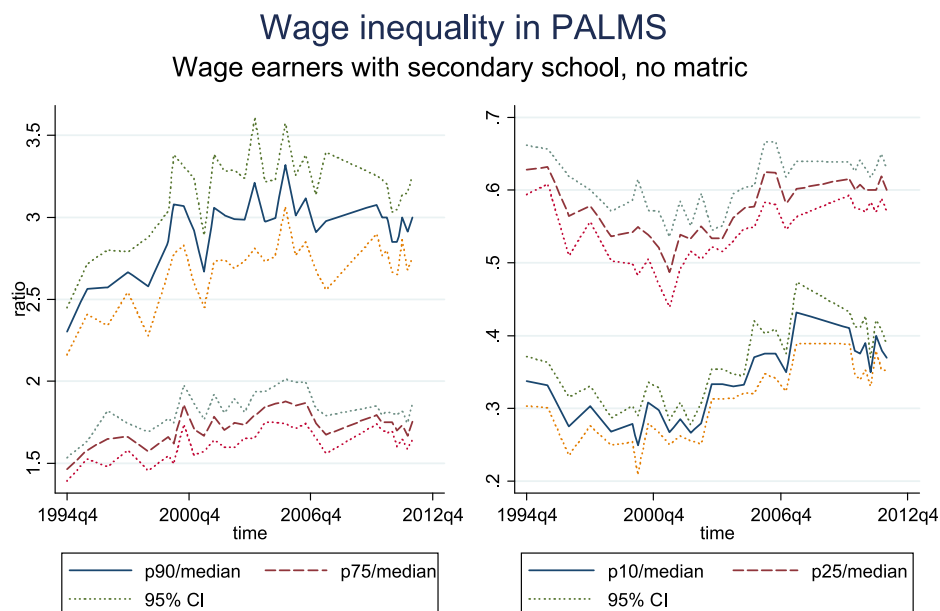
In Figs. 10–13 we extend this approach by considering the p90/median, p75/median, p25/median and p10/median ratios within the four education categories we considered earlier. Interestingly enough the patterns are quite divergent. Fig. 10 suggests that the ratio of top earnings to the median among the least educated workers has not increased, while the bottom percentiles seem to have gained on the median. This suggests an overall compression of the distribution among these workers. As noted earlier, however, this is a shrinking group within the South African labour market.

Fig. 11 suggests that high earners have gained relative to the median among wage earners with at least some secondary schooling. There is also evidence in this category of wage compression at the bottom of the distribution. The tendency of the top tail to move away from the median is strongest among individuals with a matric, as shown in Fig. 12. Unlike the previous two cases, inequality also seems to be increasing at the



Rand earnings for bracket responses and outliers imputed.
Standard errors computed by clustered bootstrap

Figure 10. Wage inequality among the least educated seems to have decreased [Colour figure can be viewed at wileyonlinelibrary.com]

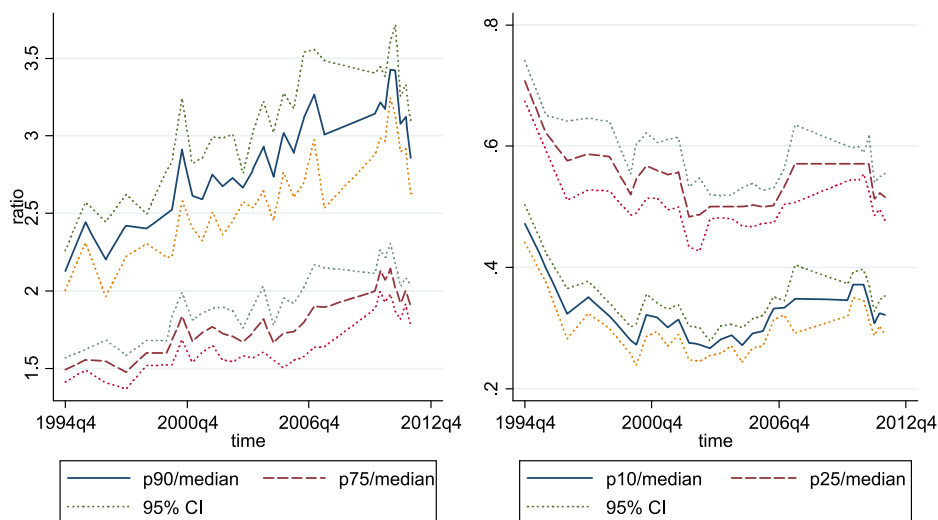


Rand earnings for bracket responses and outliers imputed.
Standard errors computed by clustered bootstrap

Figure 11. Wage trends relative to median among individuals with secondary schooling [Colour figure can be viewed at wileyonlinelibrary.com]

Wage inequality in PALMS

Wage earners with matric

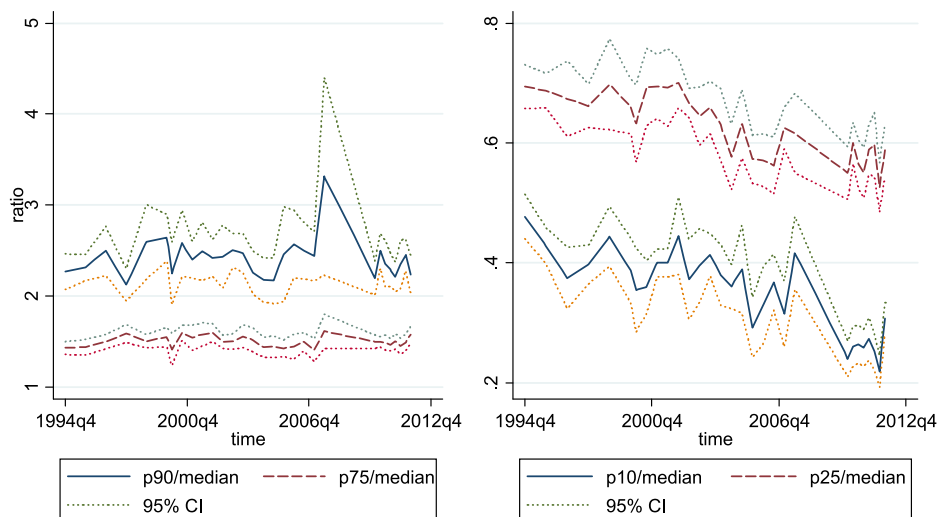


Rand earnings for bracket responses and outliers imputed.
Standard errors computed by clustered bootstrap

Figure 12. Wage trends relative to median among wage earners with a matric [Colour figure can be viewed at wileyonlinelibrary.com]

Wage inequality in PALMS

Wage earners with more than matric



Rand earnings for bracket responses and outliers imputed.
Standard errors computed by clustered bootstrap

Figure 13. Wage trends relative to median among earners with some post-matric education [Colour figure can be viewed at wileyonlinelibrary.com]

bottom of the distribution, with the lower percentiles seeming to lose out to the median. Overall this would suggest increasing inequality within this education category.

An increasing gap between the median and lower percentiles is also evident in the right hand panel of Fig. 13, which graphs the estimates for individuals with some post-matric education. In this case, however, the median is keeping pace with the wage increases higher up the income distribution.

Overall the picture is somewhat heterogeneous: among workers with lower education levels (*i.e.* the least skilled), the wage distribution has become more compressed at the bottom, with some “fanning out” of wages at the top, at least for workers with some secondary school. For matrics the entire distribution seems to have widened over time, whereas for people with some post-matric education the median has kept up with higher earnings, but there are bigger wage disparities at the bottom. This heterogeneous picture again highlights why a single summary measure of inequality may conceal as much as it may reveal.

6. SHARE OF THE TOTAL WAGE BILL GOING TO THE RICH

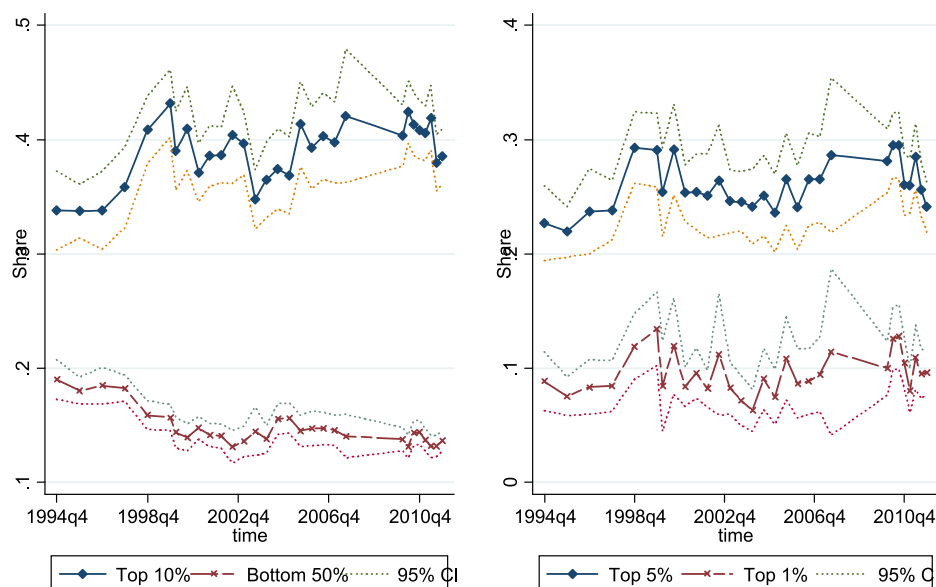
Piketty has made the same point about “synthetic indices” of inequality: “it is impossible to summarise a multidimensional reality with a unidimensional index without unduly simplifying matters and mixing up things that should not be treated together” (2014, p. 266). Indeed he also cautions against placing too much stock on ratios of percentiles (p. 267), such as those we reported in the previous figures. For instance the p90/median ratio can conceal what happens to incomes above the threshold, in this case the ninetieth percentile. Piketty prefers to analyse the shares going to various portions of the distribution. Analytically this is equivalent to focusing on what happens to various points on the Lorenz curve.

In Fig. 14 we show the evolution of the shares of the top 10%, top 5%, top 1% as well as the bottom 50% of the wage distribution. Again there are fluctuations in the series, but it appears that the top percentiles have gained share, while the bottom 50% has unambiguously lost ground. To calibrate the size of the concentration of wages at the top, it is worth noting that in France the share of the top 10% has been steadily in the range 25 to 28% for most of the last 100 years (Piketty 2014, figure 8.1, p. 272), while the share of the top 1% has been between 6 and 8% (Piketty 2014, figure 8.2, p. 273). By those standards the South African wage distribution shows considerably more inequality.

7. WAGE INEQUALITY AND OVERALL INEQUALITY

Nonetheless the shares going to the top decile of wage earners are not as large as the shares going to the top decile of the *per capita* household income distribution as reported, for instance, in Leibbrandt *et al.* (2010, p. 26 and Table A.3.3). Those figures show the top decile taking 54, 57 and 58% of all income in 1993, 2000 and 2008, respectively. There are two reasons why the overall income distribution is even more unequal than the wage distribution. Firstly, there are other types of income which exacerbate inequality, such as self-employment income and capital income. While the latter is likely to be extremely important (as argued by Piketty, 2014), it is measured rather poorly in most household surveys. Secondly, however, the way in which wage earners group into

Shares of total wage bill



calculations using PALMS v2

Figure 14. Wage inequality in South Africa 1994–2011 according to the shares accruing to high income earners [Colour figure can be viewed at wileyonlinelibrary.com]

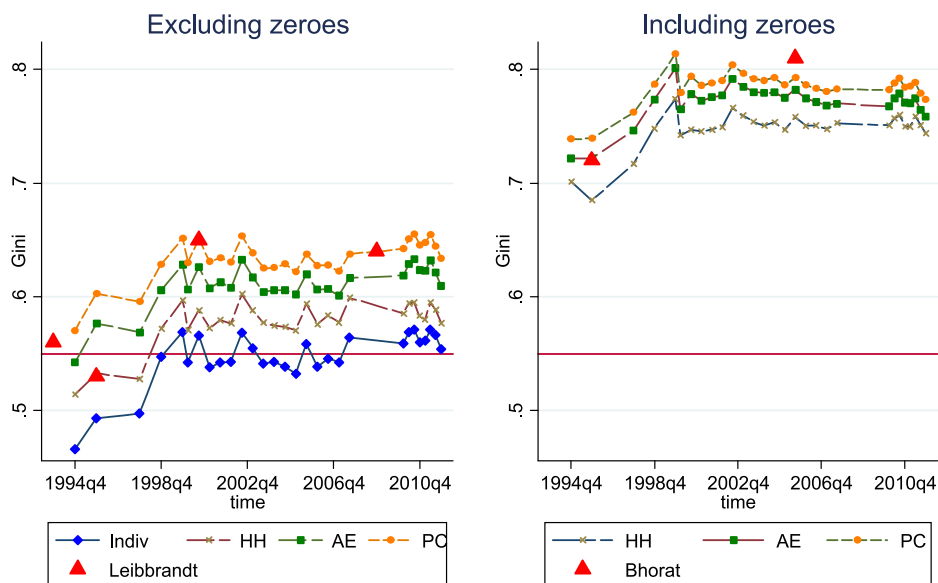
households concentrates the income. High earners are frequently in households with other high earners while low earners often share their income with the unemployed. Furthermore high earners typically live in smaller households, *i.e.* share their incomes with fewer dependents.

With our data we cannot explore the first avenue towards increasing inequality (other income sources), but we can comment on the second (the nature of households). In Fig. 15 we show how households intersect with wage inequality.

The left-hand panel shows four series, as well as some results from the literature for comparison. The bottom-most line shows the Gini coefficient of individual wage inequality according to the multiple imputation method. This is the same series shown in Fig. 1 and in the accompanying article (Wittenberg, *in press*). The line immediately above it represents total **household** wage inequality. Here the individual incomes are aggregated up within the household, but each household counts once only and no account is taken of how many people it contains nor is any attempt made to “redistribute” the household income to the individuals within it. It is evident that household wage inequality is more unequal than individual wage inequality. This will be due mainly to the fact alluded to earlier that high earners tend to live with other high earners.

The remaining two series in the left-hand panel of Fig. 15 show the distribution of individual income inequality if household income is “equivalised” according to different rules and distributed to its members. The top line (labelled PC) is a *per capita* allocation, while the line below it (labelled AE) is per “adult equivalent” according to the formula used in Leibbrandt *et al.* (2001):

Wage Income Inequality 1994-2011



calculations using PALMS v2 plus reported estimates

Figure 15. Household vs. individual wage inequality 1994–2011 according to the Gini coefficient. The reported estimates are: (Left-hand panel) Leibbrandt et al. (2010) for 1993, 2000, and 2008 and Leibbrandt et al. (2001) for 1995; (Right-hand panel) from Bhorati et al. (2009) [Colour figure can be viewed at wileyonlinelibrary.com]

$$AE = (Adults + 0.5 * Children)^{0.9}$$

This formula assumes that children consume only half as much as adults, but that there are no strong economies of scale within households.

It is evident that all these adjustments matter. Inequality per adult equivalent is more pronounced than household inequality, since poorer households tend to be larger, so once the income is redistributed to its members the gap between the top of the distribution and the bottom has widened. The highest measured inequality is recorded with the *per capita* adjustments. This raises the question as to whether this may not be exaggerating inequality. Effectively one assumes that there are **no** economies of scale within households and that all individuals (including children) have the same consumption needs.

The left-hand panel however understates inequality because it does not take into consideration members of households that have no wage income. The right hand panel, by contrast includes individuals in households recording zero wage income. It is evident that wage income inequality is much higher once these individuals are included. Of course some of these individuals will have labour market incomes (the self-employed) and some of them will have incomes from previous wage earnings (retirees or other people living from saved income). So the right hand panel overstates labour market related inequality. Nevertheless it is a useful corrective to the left hand panel particularly in South Africa where there is a high unemployment problem.



Figure 16. The proportions of individuals living in individuals with no wage income according to PALMS [Colour figure can be viewed at wileyonlinelibrary.com]

We have superimposed on both panels some of the previous estimates of labour market inequality reported in the literature. As we noted in section 2 the Leibbrandt *et al.* (2010) estimates include self-employment income, whereas the Leibbrandt *et al.* (2001) and Bhorat *et al.* (2009) numbers exclude them. It is interesting to note that although these estimates come from other data sources (the income and expenditure surveys, the PSLSD and NIDS) and they seem to make different assumptions, they do on the whole agree with the estimates shown in Fig. 15. There are some differences. The Leibbrandt *et al.* (2001) estimate for 1995 seems to be for total household income rather than equivalised household income and the figure from the 2005 IES reported by Bhorat *et al.* (2009) seems too high.

Looking at the left-hand panel of Fig. 15 it is clear that the pattern of individual wage inequality (bottom series) as discussed in the previous sections of this paper is the overall driver of all these household wage income inequality trends. Since the literature agrees that labour market inequality is the dominant driver of overall inequality, it is clear that the increase in measured wage inequality which occurred between the early OHSs and the post-1998 surveys (as discussed in section (iv) of 3.2) must be driving the increase in overall inequality measures reported in the literature.

Before concluding our discussion, it is worth drawing attention to the noticeable drop in inequality shown in the right hand panel of Fig. 15 between 1999 and 2000. It is possible to explain this if we note that there is a simple formula derived by Lerman and Yitzhaki (cited in Leibbrandt *et al.* 2001) for calculating the Gini for any distribution which has a point mass at zero:

$$G_x = p_{x=0} + p_{x>0} G_{x|x>0}$$

here G_x is the Gini coefficient for the entire distribution (including the zeroes), $G_{x|x>0}$ is the Gini coefficient for the distribution excluding the zeroes and $p_{x=0}$ and $p_{x>0}$ are the proportions of zero and non-zero observations, respectively.

Obviously the proportion of individuals with zero income will be a key determinant of the evolution of the overall Gini G_x . We show the relevant numbers in Fig. 16. Several

trends are noteworthy in that graph. Firstly we see in the early OHSs that there seem to be fewer individuals living in households with no wage income. The 1995 OHS seems particularly anomalous, which adds to the problems already noted in relation to that dataset (Wittenberg, 2014b). The 1999 OHS, by contrast, seems to find too little wage employment. This also seems to be a measurement problem.

8. CONCLUSION

Our results can be summarised as follows:

- Measured wage inequality shows a big increase between the surveys before 1998 and those thereafter. The early OHSs are known to have under-sampled small households (Kerr and Wittenberg, 2015) so this suggests that at least part of the increase is an artefact of changes in measurement. Nevertheless other surveys also suggest increases in inequality, which suggests that the trend may be real, but the magnitude of the increase may be overstated.
- The top of the earnings distribution has moved away from the median while the bottom has moved up. Depending on the measure of inequality used this will suggest either that overall inequality has increased, remained stable or decreased over the period 2000–2011. Nonetheless it seems clear that the rich have gotten richer.
- The way in which wage earners sort into households exacerbates inequality overall. How household income is equivalised affects the size of the measured Gini coefficient. Per capita income shows the highest level of inequality.
- The trends in wage inequality are sufficient to explain the trends in overall inequality reported in the literature. Understanding the patterns in the data is therefore of considerable interest.

We have also shown that paying attention to the quality of the data is important in analysing these trends, whether it is in noting the anomalies shown in Fig. 16 or the impact of different adjustments evident in Fig. 1. One of the big advantages of using all the available data in PALMS is that it puts the individual estimates reported in the literature into a more detailed context, as seen particularly in Fig. 15. It makes it easier to avoid misleading baselines, like the OHS 1995. Clearly more work is necessary both in unpicking the quality of the data and in explaining the shifts we have documented. Overall, however, there seems little doubt that wage inequality has ticked up over the 1994–2011 period.

9. DATASETS

Kerr, Andrew, David Lam and Martin Wittenberg (2013), Post-Apartheid Labour Market Series [dataset], Version 2.1, Cape Town: DataFirst [producer and distributor], 2013. [zaf-datafirst-palms-1994-2012-v2.1]

Statistics South Africa, Labour Force Surveys 2000.1–2007.2 [datasets], Pretoria: Statistics South Africa [producer], Cape Town: DataFirst [distributor], 2013.

Statistics South Africa, Labour Market Dynamics in South Africa 2010–2011 [datasets], Pretoria: Statistics South Africa [producer], Cape Town: DataFirst [distributor], 2012.

Statistics South Africa, October Household Surveys 1994-1999 [datasets], Pretoria: Statistics South Africa [producer], Cape Town: DataFirst [distributor], 2013.

Statistics South Africa, Quarterly Labour Force Surveys 2010.1–2011.4 [datasets], Pretoria: Statistics South Africa [producer], Cape Town: DataFirst [distributor], 2013.

REFERENCES

- ARAAR, A. and DUCLOS, J.-Y. (2013), DASP: Distributive Analysis Stata Package, PEP, World Bank, UNDP and Université Laval. Version 2.3.
- ATKINSON, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2: 244-263.
- (2015). *Inequality: What Can Be Done?* Cambridge, MA: Harvard University Press.
- BHORAT, H., Van Der WESTHUIZEN, C. and JACOBS, T. (2009), Income and Non-Income Inequality in Post-Apartheid South Africa: What are the Drivers and Possible Policy Interventions? DPRU Working Paper 09/138. Available at: <http://ssrn.com/abstract=1474271>
- COWELL, F. A. (2000). Measurement of inequality. In A. B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution*, Vol 1. Elsevier, 87-166.
- DEATON, A. (1997). *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore: Johns Hopkins University Press for the World Bank.
- DUCLOS, J.-Y. and ARAAR, A. (2006). *Poverty and Equity: Measurement, Policy and Estimation with DAD*. Ottawa: Springer.
- KERR, A. and WITTENBERG, M. (2015). Sampling methodology and field work changes in the October Household Surveys and Labour Force Surveys. *Development Southern Africa*, 32: 603-612.
- LEIBBRANDT, M., WOOLARD, I., BHORAT, H. (2001). Understanding contemporary household inequality in South Africa, Chapter 1. In Bhorat, H., et al., *Fighting Poverty: Labour Markets and Inequality in South Africa*. Cape Town: UCT Press, 21-40.
- , WOOLARD, I., FINN, A. and ARGENT, J. (2010). Trends in South African Income Distribution and Poverty since the Fall of Apartheid, OECD Social, Employment and Migration Working Papers, No. 101, OECD Publishing. Available at: <http://dx.doi.org/10.1787/5kmms0t7p1ms-en>
- , FINN, A. and WOOLARD, I. (2012). Describing and decomposing post-apartheid income inequality in South Africa. *Development Southern Africa*, 29: 19-34.
- NATIONAL PLANNING COMMISSION (2012). National Development Plan 2030: Our Future – make it work. Available at: <http://www.poa.gov.za/news/Documents/NPC%20National%20Development%20Plan%20Vision%202030%20lo-res.pdf>
- PIKETTY, T. (2014). *Capital in the Twenty-First Century*. Cambridge, MA: Belknap Press of Harvard University Press.
- VAN DER BERG, S. (2011). Current poverty and income distribution in the context of South African history. *Economic History of Developing Regions*, 26: 120-140.
- WITTENBERG, M. (2014a). Analysis of employment, real wage, and productivity trends in South Africa since 1994, ILO Conditions of Work and Employment Series, Working paper No. 45, March 2014. Available at: http://www.ilo.org/public/libdoc/ilo/2014/114B09_23_engl.pdf
- (2014b). Data issues in South Africa. In H. Bhorat, A. Hirsch, R. Kanbur, and M. Ncube (eds), *Oxford Companion to the Economics of South Africa*. Chapter 7, pp.79–83. Oxford: Oxford University Press.
- Wages and wage inequality in South Africa 1994–2011: Part 1 – wage measurement and trends. *South African Journal of Economics* (in press).
- YU, D. (2009). The comparability of Census 1996, Census 2001 and Community Survey 2007, Stellenbosch Economic Working Papers, 21/09. Available at: <http://www.ekon.sun.ac.za/wpapers/2009/wp212009/wp-21-2009.pdf>