Understanding Income Stratification in Post-Apartheid South Africa: A Gini Decomposition Approach

Ratjomose P Machema*

August 21, 2015

Very Preliminary. Please do not quote

Abstract

The goal of this paper is to enhance understanding of income stratification using South African example. First, we apply the Analysis of Gini (ANOGI) on data collected in 1993, 2008 and 2012 to measure the extent of income stratification. Second, we use shift-share analysis to examine how changes in a range of income sources interact to account for changes in household income distribution and income stratification. Our results indicate that the stratification of income distribution according to races has fallen between 1993 to 2008. And compared to races, groups by the years of schooling have the highest intra-group variability and overlapping in both 1993 and 2008. Using the shift-share analysis, our findings suggest that government transfers have a tendency to lower stratification while labour income is the most disequalising.

JEL classification: C81; D31; D63; I31

Keywords: Inequality decomposition; stratification; race; South Africa

*Ratjomose Machema is grateful to the financial support from African Economic Research Consortium (AERC). This version of the paper is prepared for the Biennial Conference of the Economic Society of South Africa (ESSA), University of Cape Town (UCT), Cape Town, South Africa, 2-4 Septmeber 2015.

Correspondence: School of Economics, University of Cape Town, Rondebosch 7701, email: ratjomose@yahoo.com; mchrat003@myuct.ac.za;

1 Introduction

One of the least-understood paradoxes of our time is how income disparities between well-defined identity groups continue to persist – even as individuals composing these groups change and social environments evolve. The persistence and/or formation of these inequalities has spurred great interest, among both academics and policy makers. This paper complements past research on these distributional patterns by focusing on South Africa, and what its experience can tell us about the income disparities between-groups. Specifically, we investigate the extent of hierarchical ordering of identity groups according to income (i.e. income stratification), and additionally inspect how changes in a range of income sources interact to account for changes in household income distribution and "income stratification".

Income stratification refers to formation of observable layers, or the state of being comprised of layers (Yitzhaki, 1994). In the context of income, these layers may be segments of the overall income distribution. Therefore, perfect stratification, by race for instance, corresponds to the case in which there is no overlap between different racial groups in their ranges of income (Zhou, 2012). This conceptualization makes income stratification fundamentally different from income inequality. Inequality measures how dissimilar the members of a given group from each other, while stratification shows how different the members of a group from the members of other groups. According to Yitzhaki (1994), these two are actually inversely related, because high within-group inequality is likely to increase overlap of the group, thus reducing stratification (Allanson, 2014; Monti and Santoro, 2011)[1].

Therefore, this paper explores this income stratification concept, using the South African example. Defining the income stratification by the inverse of overlapping of income distributions. Then, the main preposition of our paper is that the differences and distances between income distributions of well-defined identity groups mirror income stratification. Embedded within this preposition are three objectives, first we examine the patterns of both between-group and within-group inequality as well as the impact of overlapping on both. Second, we investigate whether the observed changes in the South African income distribution suggest the emergence of new strata replacing race as a defining stratum. Finally, we assess how changes in a range of income sources interact to account for changes in household income distribution and income stratification. Thus, the purpose of the study is to both elucidate the extent and correlates of income stratification and, to offer an updated/newer perspective on the transformation the country has undergone in the past 20 years.

There is some literature addressing certain aspects of the South African income distribution. Most recently, Leibbrandt and Levinsohn (2014) applied a variant of DiNardo et al. (1996) technique to figure the effects of endowments, returns to the endowments and Child Support Grant (CSG), Gradín (2013); Gradin (2015) decomposed the differ-

[1] Income stratification should not be confused with economic or income segregation which deals with the extent to which individuals or households of different socio-economic characteristics are unevenly distributed thought out a society. Income stratification rather implies a hierarchical ordering of groups according to income.

ences in the risk of poverty across identifiable subgroups, while in a contemporaneous study, Machema (2014) examined the effects of different income sources on income polarization. Despite this burgeoning literature, the effects of source-levels incomes are inconclusive and remain relatively understudied compared with household characteristics. Thus, the first contribution of this study is to shed light on the effects of income sources on income distribution. Second, we contribute substantively to the scarce income stratification literature by using information from South Africa. This kind of analysis has not previously been attempted using South African data. The study, therefore provides important new insights into the patterns and proximate causes of trends in the distribution of income in South Africa. These insights have wide-ranging policy relevance, including the evaluation of social process and opportunities for mobility, and tackling the economic (dis)advantage of (low) high ranked groups.

To achieve the above objectives, we first classify individuals into identity groups based on their race and years of schooling of their household head. Our statistical analysis combines two empirical strategies - the Analysis of Gini (ANOGI) and the shift-share micro-simulation strategy adapted from Burtless (1999), Daly and Valletta (2006) and Larrimore (2014). This opproach is consistent with Cowell and Fiorio (2010), who argue that additional insights are gained by combining both income source and subgroup decompositions. Our findings, which are consistent with the literature, suggest that White individuals are more stratified (or have low degree of overlapping) while the other racial groups racial groups show a high degree of overlapping with the overall population, as well as with each other. Furthermore, we discovered that compared to races, the years of schooling had the highest intra-group variability as well as overlapping between groups in 1993 and 2008. However, in the recent years, the stratification levels of years of education beginning to emerge. Lastly, using the simulations we show that government transfers have a tendency to lower stratification, while labour income as the major factor accounting for changes in household income.

The rest of the paper proceeds as follows. Section 2 provides a brief review of the related literature. Section 3 describes the data sets, presents and compares summary statistics of selected key variables across identity subgroups. Section 4 describes our analytical strategies. Section 5 discusses our results and elaborates the findings and section 6 concludes.

2 Related Literature

In this section, we provide a selective review of some important studies which are related to our paper. Our primary goal is to summarise general conclusions that will facilitate understanding on inferences given in the paper and highlight our contributions.

Our first contribution is to the income stratification literature which is based on the "traditional" decomposition of the Gini coefficient by population subgroups. Most relevant to our paper is the decomposition developed in (Yitzhaki, 1994; Yitzhaki and Lerman, 1991), rearranged by Frick et al. (2006) and extended by (Allanson, 2014; Monti and Santoro, 2011). Using the Gini coefficient presented in Lerman and Yitzhaki

(1984), Yitzhaki and Lerman (1991) interpret the third residual term (known as the overlap/interaction term) as a measure of the sociological concept of "stratification"[2]. Stratification is defined as the inverse of overlapping, thus it refers to formation of observable layers (Yitzhaki and Lerman, 1991). Frick et al. (2006) when examining whether the existence of several independent samples within one survey represent the same population. They rearranged and reinterpreted the Yitzhaki (1994) decomposition in a manner analogous to Analysis of Variance (ANOVA) but performed with the Gini coefficient. They referred to their approach as an Analysis of Gini (ANOGI). Allanson (2014); Monti and Santoro (2011), on the other hand, showed that the impact of stratification (or overlapping) on inequality is a function of the probability of transvariation. In this paper, we follow the application of ANOGI because it encompasses an indicator of stratification. This approach is discussed in detail in Appendix A.3 on page 23.

Most notable applications of ANOGI are on the cohesive tendencies of incomes across regions/countries (Frick and Goebel, 2008; Milanovic and Yitzhaki, 2002); the quality of classification into population subgroups (Heller and Yitzhaki, 2006; Liberati and Yitzhaki, 2012); the differences in the welfare of various groups (Zacharias and Vakulabharanam, 2011); and the economic assimilation of immigrants or the degree of "melting pot" (Ceccarelli et al., 2014; D'Agostino et al., 2015). For instance, in a study of the relationship between wealth inequality and caste divisions in India between 1991/92 and 2002/03, Zacharias and Vakulabharanam (2011) found that forward Hindus caste are more stratified in terms of their wealth distribution (i.e they have a low degree of overlapping) while the other caste divisions show a high degree of overlapping with the overall population, as well as with each other. Investigating the extent of immigrant's assimilation in Italy, (D'Agostino et al., 2015) grouped the population by immigration status to confirm that "the income ranges of different groups show non-negligible degrees of intersection both among the groups and with reference to the overall population. An interesting study by Heller and Yitzhaki (2006) used the ANOGI approach to examine the quality of the classification of families of prehistoric snails according observable characteristics of the shells.

The discernible conclusion from this literature is that the extent of income stratification (and/or economic integration) depends on the study area/context, type of classifications (or identities) and not so much on the number of population groupings. In this paper we contribute by investigating economic integration (or income stratification) across racial groups in the South African context. Defining integration (stratification) by the degree (inverse) of overlapping of income distributions, our work provides, for the first time in South Africa, the dynamics of income stratification and its effect on within-group and between-group inequalities. However, recognizing that changes in the

[2]Actually it was Mehran (1975) and Pyatt (1976) who showed that the decomposition of Gini concentration ratio "reveals more" information than the generalized entropy measures. Dagum (1998) and Deutsch and Silber (1999) decompositions also give an important contributions to the understanding of the overlapping term. For instance, Dagum (1998) introduces the concept of economic distances between distributions and relative economic affluence (REA) by interpreting the overlap term as the intensity of transvariation between population subgroups. Deutsch and Silber (1999) focus on the correspondence between the Gini transvariation concept and the overlapping of distributions.

overall income distribution, and hence income stratification, can result from source-level incomes from any population subgroup. We complement our analysis by applying a shift-share micro-simulation approach to document how the range of income sources account for changes in the household income distribution and, possibly, income stratification. Our shift-share technique is a rank-preserving income exchange derived from Burtless (1999) and Daly and Valletta (2006). It is similar to the one used by Larrimore (2014); Larrimore et al. (2015) to demonstrate how factors accounting for income inequality trends have evolved over time in the United States. This approach, unlike most micro-simulation approaches[3] such as the DiNardo et al. (1996) technique, has the ability to observe how a range of income sources interact to account for changing income distribution, which is our focus in this paper. Detailed discussion of this approach is in section 4.2 on page 9.

3 Data and Descriptive Statistics

3.1 Data

The data used in this paper come from the 1993 Project for Statistics on Living Standards and Development (PSLSD) and the 2008 and 2012 waves of National Income Dynamics Study (NIDS) data sets. The PSLSD was conducted just before the first democratic elections in 1994 to give a broad picture of state of the individuals at the start of democracy. PSLSD was designed as a self-weighting survey, however, due to violence and under-representation of whites in some clusters, then weighting was essential. NIDS is the first national panel survey intended to capture the distributional dynamics of welfare across the nine provinces in South Africa. It is conducted every two years and the first wave started 2008. However, these waves are treated as independent cross-section surveys by weighting them appropriatetly. We only use the first and third waves because of some questionable quality of wave 2 (Finn and Ranchhod, 2013). These surveys, despite some comparability issues which are discussed in detail in Leibbrandt et al. (2012), have several similarities that have resulted in them being important references for analysis of welfare dynamics in post-apartheid South Africa (see Burger and McAravey (2014); Leibbrandt and Levinsohn (2014); Leibbrandt et al. (2010)).

We use real household income per capita as our indicator. It involves assigning to each individual, the total monthly income of his/her household divided by the corresponding household size. This procedure assumes that intra-household allocation is egalitarian, that is, all members of the household get the same share of income, regardless of their individual incomes, their role in the household and other characteristics. To ensure comparability across surveys, total income is calculated as in Leibbrandt

[3]Micro-simulation approaches basically involve construction of counterfactuals (i.e simulated distribution) and its comparison with the actual distribution (Fortin et al., 2011). The difference between the outcome of interest (in our case income stratification or overlapping) allows the influence of the characteristics to be assessed. Fortin et al. (2011) provides a detailed survey of the common micro-simulation approaches.

et al. (2012). That is, we omitted both imputed rental income and agricultural income because they are not uniformly available/measured across the surveys. Total monthly income is the sum of individual incomes from labour, remittances, capital, government transfers and other incomes (see appendix A.1 on page 23 for a list of variables used in each income source). To arrive at real incomes, the per capita monthly household income is deflated to 2012 prices using the national consumer price index. The individual income per capita (also refereed to as household income per capita) is considered the unit of analysis. All individuals and/or households with zero, missing and negative incomes are removed from each sample. This may cause bias, but the excluded cases account for less than 1% of the observations in each sample. The pooled final sample has 97869 individuals of which 38 780 (7629), 27 073 (6862), and 32 016 (7544) individuals (households) are from PSLSD, NIDS 2008 and NIDS 2012 respectively.

We define identity groups by four racial groups (i.e Black/Africa (77.97%), Coloured (9.03%), Indian/Asian (2.59%)[4] and White (10.42%). While this race grouping may not adequately reflect the complex, dynamic and layered reality of identity in the current South Africa, it however captures the history of the country. To test our hypothesis of the emergence of new strata, we also use three education groups that are measured by the number of years of schooling of the household head (i.e unskilled or less that 7 years (45.09%), semi-skilled or from 7 to 12 years (43.47%) and skilled or more than 12 years. (11.44%)). Thus the latter classification will mainly used to determine which grouping is better (i.e has highest stratification).

3.2 Descriptive Statistics

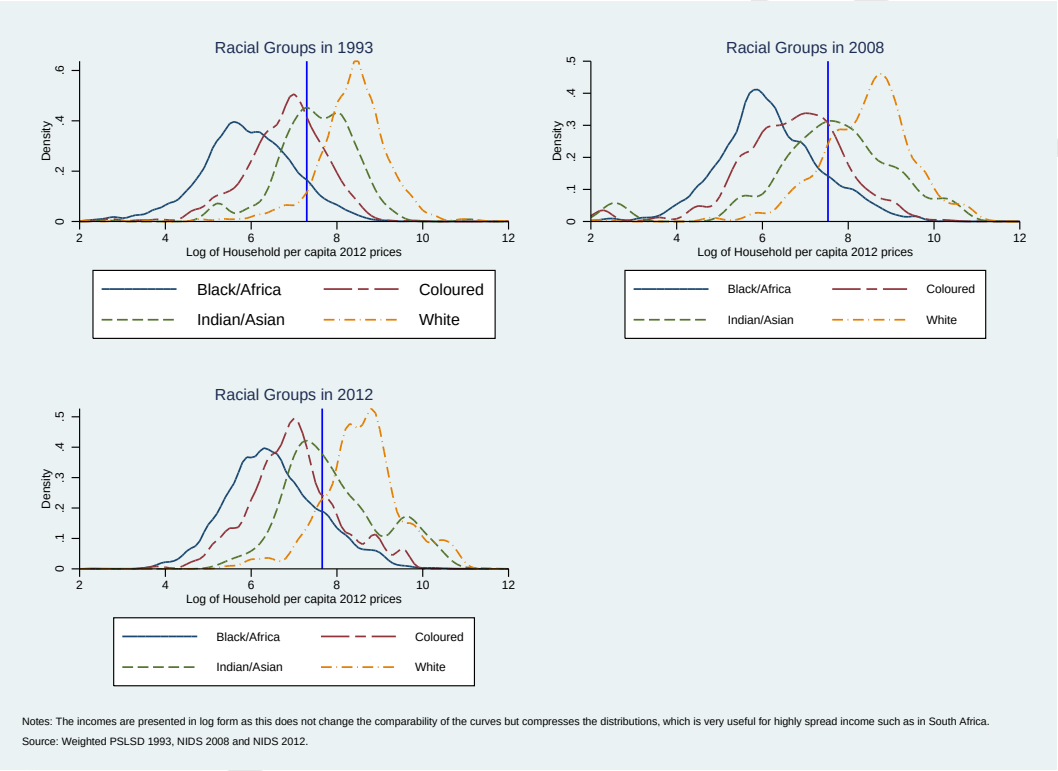
As our preliminary analysis, Table A.1 on page 24 and Figure 3.1 highlight the structure of income profiles across the racial groups and give insightful differences and/or distances by displaying the distributions of real per capita income, in the three years. We notice that between 1993 to 2012, while real incomes increased for all racial groups, the income shares show different patterns. For instance, the Blacks income share increased by 44%, Coloureds share steadied at around 8% and Whites share declined from 52% to 35%. For instance, the Whites average income was more than 8 times, 7 times and 6 times that of Blacks/African in 1993, 2008 and 2012, respectively. The income composition, on the other hand, shows no significant different patterns across the years and racial groups. That is, the average number of income sources within the households has steadied around 1.4. However, the share of the individual income sources shows that government transfers have increased between 1993 to 2008.

Looking at the overlaid distributions -where the blue vertical lines represent the national mean income in that particular year, the following is is evident. The Blacks/African lie behind in terms of real per capita incomes, with the other racial groups, especially the Whites. The higher mode, median and mean of the Whites can be appreciated by the leftward position of their density function. The differences in skewness and kurtosis

[4]Due to smaller proportions of the Indian/Asian race group, our discussion will be limited three races.

in the Black and White curves are immediately obvious, with the Black distribution being more skewed to the left and having thicker tails. While the visual inspection cannot conclusively inform about the extent of income stratification, we however note that between 1993 to 2008, the was an increase in the extent of overlapping between Blacks and other races. This suggests a fall in the extent of income stratification according to races. More details and summary statistics of selected variables can be found in the Appendix.

Figure 3.1: Overlaid kernel densities



Notes: The incomes are presented in log form as this does not change the comparability of the curves but compresses the distributions, which is very useful for highly spread income such as in South Africa.
Source: Weighted PSLSD 1993, NIDS 2008 and NIDS 2012.

The non-parametric Kolmogorov-Smirnov (KS)[5] two-sample test of equality of distributions, does not only confirms that Blacks and Whites do not have the same distribution function, but also fails to reject the null hypothesis that Whites have higher income that Blacks. Repeating the KS tests for the five remaining pair-wise combinations replicates the graphical results, i.e., Whites followed by Indian/Asians, then Coloured have relatively higher per capita incomes than Blacks/Africans. This renders it evident that, these groups are likely to report different levels of overlapping among themselves and with the overall population.

[5] All the KS tests were computed in Stata using the *ksmirnov* command, and the results are available on request.

4 Empirical Strategies

4.1 Analysis of Gini (ANOGI)

The following section presents the two empirical methods used to investigate the extent and proximate causes of income stratification (or economic integration). First, we apply the Analysis of Gini (herein after ANOGI) to effectively measure the extent of income stratification. Let y_i, s_i, $F_i(y)$, $f_i(y)$, μ_i, p_i represent the income, income share of group i, cumulative distribution, the density function, the mean, and the share of group i in the overall population, respectively. And using the Lerman and Yitzhaki (1984)'s Gini coefficient (G) which is expressed as twice the covariance between income and its rank divided by the overall mean income - formally $G = {}^{2cov(y,F(y)}/_{\mu_u}$. And further assuming that the population is composed of identity groups (i.e race and household head's education) such that $F_u(y) = \sum_i p_i F_i(y)$ (i.e cumulative distribution of population is the weighted average of distributions of the groups, weighted by the relative size of population in each group). Then, according to ANOGI, the Gini coefficient (G_u) is decomposed as follows:

$$G_u = \underbrace{\sum_{i=1}^{n} s_i G_i O_i}_{G_{wo}} + G_b = \underbrace{\sum_i s_i G_i}_{\text{IG}} + \underbrace{\sum_i s_i G_i \left(O_i - 1\right)}_{\text{IGO}} + \underbrace{G_{bp+} \underbrace{\left(G_p - G_{bp}\right)}_{BGO}}_{BG} \tag{4.1}$$

where G_b is the Gini coefficient of between-inequality; G_{bp} is the between-Gini coefficient of the Pyatt (1976)'s decomposition; and O_i is the overlapping index of the distribution of identity group i with the overall national distribution. Described formally as;

$$O_i = O_{ui} = \frac{cov_i(y, F_u(y))}{cov_i(y, F_i(y))} \tag{4.2}$$

The properties of O_i make it an insightful index of stratification. The minimum value of O_i is given by the share of identity group i in the population and its maximum value is equal to 2. When the index equals the minimum possible value, it suggests that the identity group in question is a perfect stratum, that is, it occupies an exclusive segment of the overall income distribution. If a particular group has a range of income that coincides with the range of income of the population, then the index will be equal to 1. Finally, if $O_i > 1$ the distribution of income within the group is much more polarized than in the overall distribution. This can happen if the members of the group constitute two strata, one that has much higher and the other that has much lower income than μ_u, the average income of all individuals in all groups (Zacharias and Vakulabharanam, 2011). For a detailed description of this method see Appendix A.3 on page 23.

The IGO component is a weighted sum of the within group Ginis with the weights defined as the income shares. It provides either a negative or a positive revision of intra-group variability (IG)[6] for $O_i < 1$ and $O_i > 1$ respectively. This explains how

[6] This is the standard within-group inequality in the classical Gini decomposition where the over-

overlapping may affect within-group inequality. The BG component is the weighted covariance between the various groups' mean income and their mean rank. BGO represents the loss of between-group inequality (BG) due to overlapping. It is always non-positive, and reaches the upper limit (zero) if the ranges occupied by the different groups do not overlap. To clarify this, consider the difference between G_b and G_{bp}. The Pyatt's (1976) decomposition defines G_{bp} by covariance between mean incomes and the rank of mean income in the distribution of mean incomes - $\bar{F}_i(y)$. Alternatively, Yitzhaki and Lerman (1991) decomposition defines G_b as the covariance between mean income of each group and its mean rank in the overall population -$\bar{F}_{ui}(y)$. Thus, the main difference between them is the way groups ranks are established. Therefore, when there is no overlapping between groups, the two ranks are equal hence all the methods yield the same results. However, incomes generally overlap which makes the correlation between rank of mean incomes and the mean rank of incomes less than 1. Thus a reduction in G_b must be associated with an increase in overlapping. It can be seen that changes in income distribution can affect any one or all of the components in equation 4.1.

4.2 Shift-Share Analysis

This section presents the method used to examine association between income sources and income stratification. To obtain the importance of each factor, each is added sequentially and the resulting stratification (or overlapping) are compared to that would have occurred had the specified source remained unchanged. Consider a general income formation model for individual/household i at time t to be expressed as $y_{it} = g\{x_{it}, w_u, e_{it}\}$, where $g(\bullet)$ is ununknown function of a set of individual or household characteristics or some policy variables (e.g. taxes), x, a set of sampling weights, w, and some unobservable characteristics or error term, e. The function $g(\bullet)$ is not constrained to be linear nor parametric, and can represent a single income-generating equation or a system of income-generating functions.

In this paper, we assume $g(\bullet)$ to be a system of income-generating functions. We use a rank-dependent transformation that was proposed by Burtless (1999), Daly and Valletta (2006) and Larrimore (2014) to perform our counterfactual analysis. It does not involve estimation of a regression model and does not consider $g(\bullet)$ as the income-generating function but more simply as the function that aggregates different individual incomes (such as labour earnings, government transfers, etc.) into total individual income and, eventually, into household income. It incorporates the fact that the income distribution within each sub-population group is changing as well. These changes can result from any income source. Therefore, each individual's income, y_{it}^k can be represented as the sum of their incomes from each income source, f_{1it}^k through f_{Nit}^k:

$$y_{it}^k = f_{1it}^k + f_{2it}^k + \ldots + f_{Nit}^k \tag{4.3}$$

lapping is contained in the residual term.

We assign individuals a percentile rank, p_{fik}, for each income source based on the rank of their source-level income within their sub population group k. To estimate the impact that changes to the distribution of source f_1 have on income inequality, each individual's income from the source f_1 in year t is replaced with the income of the individual at the same percentile rank of the source f_1 income distribution in year t^t.

$$y^k_{it^t}(p_{fik}) = f^k_{1it^t}(p_{fik}) + f^k_{2it} + \ldots + f^k_{Nit} \tag{4.4}$$

This preserves the conditional earnings rank of each individual from source f_1 and the rank correlation of earnings from source f_1 with other income sources, while capturing changes in the source-level income distribution of source f_1 within each population group. As in any micro-simulation analysis a base year had to be picked and 2008 was chosen in our case.

5 Results

5.1 ANOGI results

This section formally examines the extent of income stratification (or economic integration) using the ANOGI approach. Given the inverse relationship between between stratification and inequality, we first consider the evidence regarding the within and between-inequalities. Tables (5.1) and (5.2) give the outcomes of the ANOGI decomposition by racial groups in the three survey years. Each row of table (5.1) reports a snapshort of each component of Gini decomposition by race groups. The first two rows correspond to equation 4.1 while the rest of the rows correspond to equation 4.1 on page 8. From row 1, which gives the standard within-group inequality (G_{wo}), we note that racial income distributions tend to become internally more heterogenous. That is, the contribution of income disparities within the races increased from 42.25% in 1993 to 60.75% and 62.57% in 2008 and 2012, respectively.

Considered in conjunction with the columns headed "sG" in table (5.2) - which gives the size of intra-group component. We note that in 2008 and 2012, the largest contribution to intra-group variability comes from Blacks and the lowest from Coloureds. For instance, in 1993 Blacks were responsible for 19.11 Gini points of inequality, which meant 40% of intra-group inequality, whereas Whites were responsible for 22.2 Gini points and 46% of intra-group inequality. However, in 2008 and 2012 Blacks had 28.7 and 29.7 Gini points, which meant 49% and 53% of intra-group inequality. Thus, one can argue that the poorer income distributions in 2008 and 2012 (as compared with 1993) are mostly driven by widening income disparities within the races. These findings mirrors the household income inequality trends observed by Leibbrandt et al. (2012). Given the inverse relationship between within-inequality and income stratification, we can at this point, argue that the extent of income stratification by races has declined.

Rows 3 and 4 of table (5.1) reports the decomposition of within-inequality to isolate within-inequality without overlapping (IG) and the impact of overlapping (IGO).

Table 5.1: Within-group and Between-group inequality, 1993, 2008 and 2012

ANOGI components	1993	2008	2012
Within-group Gini (G_{wo})	0.2852	0.4278	0.4149
Between-group Gini (G_b)	0.3898	0.2768	0.2482
Within-inequality without Overlapping (IG)	0.4779	0.5824	0.5563
Impact of Overlapping on within-inequality (IGO)	-0.1928	-0.1546	-0.1414
Pyatt (1976) Between-group Gini (G_{bp})	0.4571	0.3704	0.3273
Difference ($G_b - G_{bp}$) BGO	-0.0674	-0.0935	-0.0791

Notes: All rows correspond to the components of equation 4.1 on page 8

Source: Own calculations from weighted PSLSD 1993, NIDS 2008 and 2012

Given that when IGO is zero, it implies that each group perfectly overlaps with the entire population. Then, the decrease in IGO (of almost 25%), between 1993 and 2008, is also suggestive of a reduction in the extent of income stratification during that period. Again, looked in conjuction with the terms under the column "$sG(O-1)$", which instead, identify the revision of the contribution of each race to intra-group variability determined by overlapping component. And recalling that a negative revision of this contribution occurs when $O_i < 1$, that is, when a subgroup forms a stratum in the national distribution. Then we note that, this situation systematically occurs in all races except for Blacks in 2012. This means that the scatter of ranks for Blacks in 2012 is larger than the scatter of ranks of individuals in the national income distribution.

Further information is captured by the between-group components. Using the between-group Gini (G_b) - where each race is represented by its mean income and the mean rank of its members - we notice that its contribution to total inequality is only larger in 1993. The between-Gini of Pyatt (G_{bp}) - where each race is represented by mean income and its rank (instead of mean ranks of the members) gives a slightly stronger picture. The BGO component which identifies the effect of overlapping on between-group component, tends to increase between 1993 to 2008 and to surprisingly decrease in 2012. For instance, in 1993, overlapping of incomes has caused a loss of between-race component by 6.7 Gini points while in 2008 and 2012, the loss was 9.4 and 7.9 Gini points, respectively. Given that when BGO is zero the groups are perfectly stratified, then the implication here is that income stratification declined in 2008, and remarkably increased in 2012. Overall, we note that while the between-group inequality shows evidence of cohesive tendencies towards equality of incomes across races, there is also an increase in the loss of between-inequality due to overlapping.

Contrasting the above results with the racial Gini coefficients reported in columns headed by "Ginis (G_i)" in table (5.2a), we note the following. The path of the racial Ginis mimics the upward (2008) and downward (2012) tendency observed at national level. That is, while the Blacks have relatively poorer income distributions in all the years, it is both the Whites and the Coloureds whose inequality (or Gini coefficients) rise the most between 1993 to 2008. It is worth noting that, our results are united with previous literature on the direction of inequality, if not the magnitudes. That is, between 1993 to 2008, income disparities between racial groups have declined while they have

grown larger within the racial groups. The domination of within-inequality in 2008 and 2012 is, to some extent, indicative of wider variations in the characteristics of individuals (or households) that contribute to income disparities within races. This feature begs the question of whether the changes in the South African income distributions suggest emergence of new strata replacing race as the defining stratum. We attempt to answer this question by applying ANOGI to our years of education identity groups.

Another way of looking at the trends of the distances between incomes is by looking at the columns reporting mean ranks (in table (5.2a)). The means ranks give information on the average position (or rank) of individuals in each racial group calculated on the basis of national income distribution. The average rank of a Black individual in 1993 was located at the 41.1^{st} percentile of the national distribution, whereas by 2008 and 2012, the rank of that individual had risen to 43.9^{th} and 44.2^{nd} percentile, respectively. Individuals in the other racial groups have an average position that is always above 50^{th}percentile. Interestingly, their mean ranks have dropped between 1993 and 2008, and slightly risen in 2012. For example, a Coloured and White individual in 1993 were positioned at the 61.9^{th} and 88.6^{th} percentiles, respectively, but in 2008, their positions had declined to 56.3^{th} and 86.2^{th} percentiles, respectively. Thus, only the Blacks average ranks has increased over the period, which means that the distance between the average positions in the different racial groups has narrowed from 1993 to 2008, but increased or showed no significant change in 2012. In order words, stratification of the distribution according to races has declined between 1993 to 2008, and slightly increased increased in 2012. This distributional characteristic is not surprising and it gives a partial explanation of why Gini coefficients are normally greater for Blacks. In fact, one can argue that this result suggest emergence of a "nouveau riche" and/or growing income polarization within the Black group.

More evidence on stratification is given by the overlapping indices (in columnsheaded by "Overlapping (O_i)" in table (5.2a)). Note that if $O_i = p_i$, then the racial group forms a perfect stratum while a value of 1 indicates that the distribution of each racial group mimics the national income distribution. We notice that all coefficients for Blacks are close to 1, while for Whites are close to the population share of Whites. This suggests that the income distribution of Blacks has a high degree of similarity with the national distribution. Put differently, Blacks form less of a stratum (or have high degree of overlapping) than do others. Coloureds, likewise have high degrees of overlap with the overall national distribution. Whites, on the other hand, have the most minimal degree of overlapping (or high degree of income stratification) in every one of the years. In 2012, $O_i > 1$ for Blacks which means that this income distribution is characterized by presence of extreme observations, some richer and some poorer than the rest of the country.

Adjusting the overlapping indices by $^{O_i - p_i}/_{2-p_i}$ to account for the population share of each group we note the following. The White group, unsurprisingly, still has the lowest amount of overlapping suggesting a near- perfect stratification, while the Coloured group has generally the most astounding degree of overlapping. In a nutshell, the overlap

Table 5.2: Decomposition of Gini coefficient by Racial groups, 1993, 2008 and 2012

(a) Within-group inequality and Overlapping, racial groups

	Ginis (G_i)			Overlapping (O_i)			s_iG_i			$s_iG_i\,(O_i-1)$			Mean rank ($F_{ui}(y)$)		
	1993	2008	2012	1993	2008	2012	1993	2008	2012	1993	2008	2012	1993	2008	2012
Black/African	0.5540	0.6391	0.5976	0.9365	0.9990	1.0017	0.1911	0.2871	0.2974	-0.0121	-0.0003	0.0005	0.4114	0.4394	0.4423
Coloured	0.4990	0.5880	0.5608	0.7018	0.8458	0.8348	0.0418	0.0484	0.0465	-0.0125	-0.0075	-0.0077	0.6192	0.5633	0.5646
Indian/Asian	0.4692	0.6291	0.5985	0.5010	0.6492	0.5620	0.0225	0.0431	0.0401	-0.0112	-0.0151	-0.0176	0.7645	0.7271	0.7549
White	0.4252	0.5095	0.4888	0.2947	0.3539	0.3230	0.2225	0.2039	0.1723	-0.1569	-0.1317	-0.1166	0.8875	0.8620	0.8701
Total	**0.6749**	**0.7046**	**0.6632**				**0.4779**	**0.5824**	**0.5563**	**-0.1928**	**-0.1546**	**-0.1414**			

Notes: O_i is the overlapping index of the distribution of identity group i with the overall national distribution, and the its minimum value is population share of the group (p_i). Thus a large group is more likely to have a greater overlapping index than a smaller group. See Appendix A.3 for detailed clarification. s_iG_i gives the size of intra-group component while $s_iG_i\,(O_i-1)$ identify the revision of the contribution of each subgroup to intra-group variability determined by overlapping.
Source: Own calculations from weighted PSLSD 1993, NIDS 2008 and 2012

(b) Matrices of Overlapping and Ranks

	1993				2008				2012			
	Black	Coloured	Indian	White	Black	Coloured	Indian	White	Black	Coloured	Indian	White
A. Overlapping												
Black/Africa	1	1.118	0.884	0.455	1	1.134	1.013	0.858	1	1.156	1.090	0.838
Coloured	0.643	1	1.043	0.780	0.817	1	0.974	0.889	0.792	1	1.067	0.978
Indian/Asian	0.373	0.725	1	1.005	0.553	0.749	1	1.228	0.454	0.654	1	1.297
White	0.160	0.326	0.625	1	0.256	0.399	0.693	1	0.234	0.343	0.630	1
B. Ranks												
Black/Africa	0.5	0.2666	0.1301	0.0442	0.5	0.3671	0.2116	0.0863	0.5	0.3688	0.1830	0.0823
Coloured	0.7334	0.5	0.2809	0.0967	0.6329	0.5	0.2990	0.1388	0.6312	0.5	0.2681	0.1287
Indian/Asian	0.8699	0.7191	0.5	0.2296	0.7884	0.7010	0.5	0.3233	0.8170	0.7319	0.5	0.3042
White	0.9558	0.9033	0.7704	0.5	0.9137	0.8612	0.6767	0.5	0.9177	0.8713	0.6958	0.5

Notes: Each number of the overlap matrix (panel A) shows the overlapping of the column group (reference) by the row group. And each number of the rank matrix (panel B) shows the average rank of the group named in the row label in the distribution of the group named in the column label.
Source: Own calculations from weighted PSLSD 1993, NIDS 2008 and 2012

indices suggest that the country was less stratified in 2008 than 1993, while in 2012 income stratification seems to have accentuated. Following this findings, an interesting question is how changes in a range of income sources have interacted to account for changes in the overall income distribution, and hence the income stratification. We return to this question in section 5.2 on page 16.

To further illustrate on the extent of income stratification in the country table (5.2b) gives information on the differences among races. In particular, panel A shows the decomposition of the overlap index (as in equation A.4 on page 25) to estimate pair-wise indices of overlapping among the groups. Each term of the overlap matrix (O_{ji}) reports the extent to which the distribution of group j (rows) is included in the range of reference distribution of group i (column). For instance, the overlapping of Blacks by Whites (the reference group) was 0.16 in 1993, while by 2008 and 2012 it was 0.256 and 0.234, respectively. This suggests that the share of the White population group intersecting the range of income distribution of Black group increased between 1993 and 2008, but declined in 2012. Notwithstanding, the overlapping of Whites by Blacks - which demonstrates the extent to which the Black group income distribution is incorporated in the White group distribution - was 0.455, 0.858 and 0.832 in 1993, 2008 and 2012, respectively. Overall finding is that the overlap of each racial group by the White group is lower than the overlapping of Whites by that group. This affirms that, by and large,, Whites still have generally higher incomes, and subsequently, a low share of other racial groups is likely to be found in the upper half of the Whites income distribution. This element is in favour of our speculation that not all racial groups undergo the same integration process nor are they uniformly integrated.

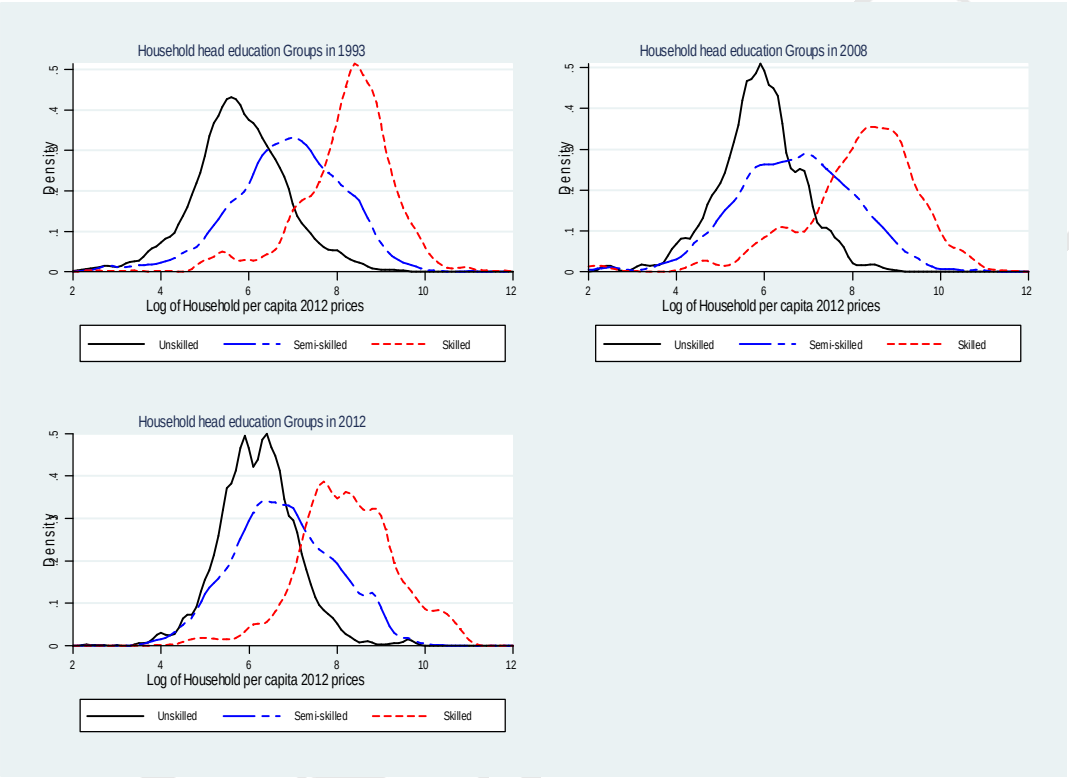
Panel B of table 5.2b finally presents the average rankings of one group in terms of another. An average rank greater than 0.5 means individuals in race/group j (rows) have, on average, higher ranks in the income distribution of the race/group i (columns) than in their own distribution, i.e., they form a richer group relative to the other race group. As expected, White individuals have this characteristic in all sample periods compared to other racial groups. This implies that the average rank of all the other groups are at their lowest levels when they are placed in the distribution of Whites. Black individuals, on the other hand, have, on average, ranks beneath the 10^{th} and 40^{th} percentile in the distribution of White and Coloured groups, respectively.

5.1.1 Do these distributional patterns suggest the emergence of new strata replacing race as the defining stratum?

The second objectiveof the paper is to determine whether these distributional patterns suggest the emergence of new strata replacing race as a defining stratum. To this purpose, we defined three groups based on the number of years of schooling of the household head. We used this education groups to also measure the extent of stratification in the country. and compare the results from racial decomposition. Before we discuss the decomposition results, we first hightilight the income profiles of the three groups. Figure 5.1 displays the overlaid income distributions across these groups in the

three years. As expected, individuals living under an unskilled household head have low levels of incomes in the three years. Visually, we can notice that the distances between the distributions have slightly narrowed between 1993 and 2008, whereas between 2008 and 2012, the distributions have shifted to the right which means overall incomes have increased.

Figure 5.1: Overlaid densities for Household head education groups



The ANOGI outcomes for the years of schooling are presented in table (5.3). The results indicate that the within-inequality increased by 8.4% between 1993 and 2008, and declined by 7.8% 2012. The impact of overlapping on within-inequality remained some how unchanged between 1993 and 2008. The loss of between-group inequality due to overlapping of incomes increased by 3.7 (from 13.9 to 17.6) Gini points between 1993 to 2008, before declining by 3.6 in 2012. This also suggests that stratification was lower in 2008 than in 2012. The mean ranks columns indicate that the average position of individulas living in household with unskilled heads also remained relatively constant in the three survey years. However, the average percentile ranks for individuals with living in households headed by both semi-skilled and skilled individual decreased by 13.3% and 6.5%, respectively between 1993 to 2008. Regarding the extent of income stratification, the overlapping indices forunskilled heads with the overall population fell in 2008, and increased slightly in 2012. That is, income distribution of individuals living households where the head less that 7 years of schooling was more stratified in 2008 than in 2012. Concerning the distributions of individuals where the head between 7 to 12

years and more than 12 years, we note that their overlapping indices increased, in general between 1993 to 2012. Overall, individulas having more educated head more stratified while living in a household with relatively uneducated head lowers the stratification.

To determine whether the years of schooling are a better classifier than race, we follow Heller and Yitzhaki (2006) who argued that a that perfect classification into groups is accomplished if members of each group are similar among themselves (low intra-group variability) and different from others (stratified - i.e. low overlapping and high between-inequality). Comparing the tables (5.3) and (5.1), we notice that, in 1993 as well as in 2008, both the intra-group inequality and overlapping among racial groups are lower than their counterparts in the years of schooling groupings. Therefore, racial groups were better classifier of the society in 1993 and 2008. However, in 2008, the years of schooling groups had tolerably similar patterns of intra-group variability and extent of stratification. Thus, the years of schooling could have easily had the lowest overlapping. This may be evidence illustrating the emergence of strata, in particular years of education of household head. In 2012, years of schooling had the lowest intra-group inequality, while the overlapping indices show some contrasting patterns. This is unmistakably another element in support of the of the hypothesis that the changes in the South African income distribution suggest the emergence of new strata replacing race as the defining stratum.

5.2 Shift-Share results

This section presents the results of the micro-simulation analysis approach outlined in section 4.2. Our objective in this section is to examine the association between changes in a range of income sources and changes in the overall income income distribution and income stratification. In table 5.4, the columns report the simulated outcomes of the ANOGI (equation 4.1). That is, under the labour columns we ask, what would have happend to between-inequality, within-inequality as well as the impact of overlapping on both if all other income components remained at their 2008 levels, while labour income had not changed, i.e., remained at its 1993 distributions. The other columns provide outcomes to the similar "what if" question but using the indicated sources. The results are presented as percentages change from the base year - 2008, thus a negative value means that the component is question would have fallen.

According to table 5.4, the simulated effects indicate that, changes in both labour income and government transfers had the largest effect on all the components in the ANOGI equation. However, the effects would have been negative for labour income and positive for governement transfers. This suggests that the income distribution would be better than the observed 2008 patterns if labour income had not changed while it would be worse if government transfers remained at their 2008 levels. For instance, looking at the column headed labour, which means changing only the labour income distribution to its 1993 levels, while holding all other incomes constant at their 2008 levels. Then all components except for the effect of overlapping on between-race inequality (BGO)

Table 5.3: ANOGI by household head years of schooling in 1993, 2008 and 2012

(a) Within-group inequality and Overlapping, by household head years of schooling

	Overlapping (O_i)			s_iG_i			$s_iG_i\,(O_i-1)$			Mean rank ($\bar{F}_{ui}(y)$)		
	1993	2008	2012	1993	2008	2012	1993	2008	2012	1993	2008	2012
Unskilled	0.9707	0.8779	0.9379	0.1354	0.0690	0.0665	-0.0040	-0.0084	-0.0041	0.3794	0.3715	0.3612
Semi-skilled	0.7945	0.9070	0.9510	0.2765	0.2878	0.2395	-0.0568	-0.0268	-0.0117	0.6242	0.5507	0.5094
Skilled	0.3877	0.5153	0.4641	0.1415	0.2299	0.2543	-0.0866	-0.1114	-0.1363	0.8587	0.8061	0.8085
Total				**0.5534**	**0.5866**	**0.5602**	**-0.1474**	**-0.1466**	**-0.1521**			
G_{wo}	0.4059	0.4400	0.4081									
G_b	0.2700	0.2650	0.2551									
IG	0.5534	0.5866	0.5602									
IGO	-0.1474	-0.1466	-0.1521									
G_{bp}	0.4089	0.4410	0.3951									
BGO	-0.1389	-0.1759	-0.1399									

Notes: O_i is the overlapping index of the distribution of identity group i with the overall national distribution, and the its minimum value is population share of the group (p_i). Unskilled refers to households heads with up to 7 years of schooling, semi-skilled those with between 7 to 12 years of schooling and skilled those with more than 12 years of schooling.

Source: Own calculations from weighted PSLSD 1993, NIDS 2008 and 2012

would have fallen. Specifically, national income inequality would have been reduced by 0.0488. an amount equivalent to 6.95%, while the standard within-race inequality (G_{wo}) would have been reduced by an amount equivalent to 5.95%. The impact of overlaping on within-race inequality (IGO) would have had the largest simulated effect.

Table 5.4: Simulated Within and Between inequalities, 2008 and 2012

ANOGI components	2008			
	Labour	Govt	Capital	Remitt
G_u	-6.95	6.58	-1.00	-0.60
G_{wo}	-5.95	6.43	0.79	2.09
G_b	-8.49	6.83	-3.79	-5.00
IG	-6.93	6.42	-0.65	0.52
IGO	-9.75	6.40	-4.94	-4.43
G_{bp}	-5.38	6.68	-2.57	-3.16
$G_b - G_{bp} = BGO$	3.97	6.31	0.67	1.51

Notes: *Each entry reports what would have happend to the row component had the indicated income remained at its 1993 distribution*

Source: *Own calculations from weighted PSLSD 1993, NIDS 2008 and 2012*

These effects of labour income simulations are not surprising given its largest share in the total income. And our resutls matches the findings in the previous literature that path of national inequality mimics the path of labour income inequality. However, Looking at the terms under the column "Govt" which reports the simulated effects of governement transfers[7], we notice that they are all positive. This, on the other hand, suggest that, varying the government transfers alone, while holding distributions of other income sources constant to the 2008 levels, would have produced positive effect on inequality. That is, national inequality would have incresed by 0.0430 or by about 6.58%. This is effects are also unsurpringly given the concerted government efforts to equalize the opportnity sets since the demise of aprtheid regime in 1994.

Overall, these results suggest that labour income is a disequalising income component while government transfers can be seen an equalising income component. To further see roles of these two income sources, we repeated the simulations with 1993 as the base year. Our results in this case indicated that, if only labour changed to its 2008 distribution while holding other incomes constant at their 1993 levels, then national Gini coefficient increases to 0.7288, which is equivalent to 8.4% increment in inequality. However, only changing government transfers to its 2008 distribution would

[7]Since the incomes are added sequentially, then these government transfers effects are conditional on labour income remaining at its 1993 level whle the distributions of other sources are their 2008 levels. When we performed the simulation with government changing first to its 1993 level then labour income the direction of the simulated effects remained unchaged but the magnitudes changed.

have reduced national inequality by 0.0207 or by about 3.18%. Thus, at this point we can conclude that, based on our datasets, changes in labour income account for most, if not all, of the inequality growth in the post-apartheid era. The columns headed "Capital" and "Remit" report simulated effects of capital income and remiitances. As we can see, the direction of their effects are somewhat similar. That is, under both income sources, national inequality and between-race inequality would have fallen. However, both income sources tend to highten income disparities within the races.

Tables **??** and 5.6, and the discussion that follows, considers the simulated effects of the labour incomes and government transfers on the extent of income stratification. Likewise, all the entries in the tables correspond to percentage changes from the base year - 2008, if the income source in questions does not change i.e., keeps its 1993 distribution. Therefore, a positive value means an increment from the base year, suggesting either a betterment or worsening depending the component in question. Looking the labour income simulations in table **??**, we see that Whites would have had the largest reduction in their incomes while Blacks and Coloureds would have seen an increase in their income shares. The implication of this would have been a decrease in income inequality for all races. Blacks would have had the largest reduction in Gini coefficient, i.e., Gini points equivalent to 11.09%. This income changes would have increased the stratification of Whites with the overall population by about 8.76%. Interestingly, the overlap matrix (panel A) indicates that; 1) the overlap indices of Blacks and Coloureds by Whites would have been reduced by 14% and 6.25%, respectively, which means the the income distances between the races would have been accentuated; 2) the overlap of all the races by Blacks would have risen, also suggesting that there would have been more stratification between races. Overall finding is that, if labour had not changed, the country would be less unequal but more stratified.

Table 5.5: Labour income simulations

	μ_i	s_i	G_i	O_i	F_{ui}	s_iG_i	$s_iG_i(O_i-1)$
Black/African	-8.95	3.46	-11.09	-1.96	0.15	-8.02	419.06
Coloured	-2.72	10.54	-4.50	0.05	0.58	5.56	5.28
Indian/Asian	-8.01	4.53	1.63	4.14	-1.68	6.23	-2.06
White	-18.00	-6.83	-4.85	8.76	-0.58	-11.34	-15.69
Total	**-11.99**		**-6.95**			**-6.93**	**-9.75**

A. Overlapping matrix

	Black/	Coloured	Indian	White
Black/Africa	0	2.12	10.57	17.84
Coloured	0.67	0	-2.32	-4.86
Indian/Asian	-5.95	-0.12	0	-0.93
White	-14.04	-6.25	-1.37	0

B. Ranks matrix

Black/Africa	0	0.98	-6.60	-7.47
Coloured	-0.57	0	-2.06	-2.18
Indian/Asian	1.78	0.89	0	2.09
White	0.69	0.34	-0.97	0

Notes: *All entries show the percentage changes of the column components when labour income does not changes - remains at its 1993 level.*

Source: Own calculations from weighted PSLSD 1993, NIDS 2008 and 2012

A different picture of what would have happend is presented by the government transfers simulations in table 5.6. While all races would have expericend a fall in their average incomes, only the income share of the Blacks would have fallen if the government transfers were changed to their 1993 levels. This is not unexpected given that the share of government transfers in total income has increased considerably in the past 20 years. Furthermore, keeping government transfers unchanged at their 1993 levels would have

Table 5.6: Government income simulations

	μ_i	s_i	G_i	O_i	F_{ui}	s_iG_i	$s_iG_i(O_i-1)$
Black/African	-9.48	-4.60	11.94	0.16	-0.15	6.79	-0.11
Coloured	-4.54	0.60	6.13	-1.59	0.83	6.77	15.93
Indian/Asian	-0.93	4.41	1.72	-0.72	0.23	6.20	7.84
White	-0.61	4.74	1.04	-0.07	0.02	5.83	5.87
Total	**-5.11**		**6.58**			**6.42**	**6.40**

A. Overlapping matrix

	Black/	Coloured	Indian	White
Black/Africa	0	-1.23	-1.11	0.03
Coloured	2.05	0	0.69	0.25
Indian/Asian	1.15	-1.37	0	0.49
White	0.39	-1.09	-1.27	0

B. Ranks matrix

Black/Africa	0	1.58	0.95	0.54
Coloured	-0.91	0	-1.14	-0.83
Indian/Asian	-0.28	0.51	0	-1.09
White	-0.05	0.13	0.49	0

Notes: *All entries show the percentage changes of the column components when labour income does not changes - remains at its 1993 level.*

Source: Own calculations from weighted PSLSD 1993, NIDS 2008 and 2012

increased Gini coefficient for Blacks by a largest amout, equivalent to 11.9%. The effects on the overlapping with the overall population and/or among the races are less pronounced. Nevertheless, the overlapping of all races except Blacks with the overall population would have fallen while that of Blacks would have increased. Looking the differences between the races, we note that the extent to Blacks distribution is included in the Coloured distribution would have fallen while the overlapping Blacks by Whites would have increased. The simulated ranks matrix suggests that average ranks of Blacks in the distribution of other races would have fallen whereas the the ranks of those races in Black distribution would have increased. In general, we can say that government transfers have a tendency to lower stratification of Blacks.

6 Conclusion

Numerous studies have documented the increases in income inequality and the importance of labour market earnings as the major driver. However, there are other aspects of the income distribution (like income polarization and income stratification) that remain relatively unexplored especially in developing countries. While realtively similar to inequality and most cases derived from inequality measures, they capture different dimensions of the overall income distribution. For instance, income stratification - which is our main focus in this paper- focuses on the extent to which different subgroups occupy distinct strata of the income stratification. This may be related to socio-economic mobility of different subgroups in the society. Thus, our main goal in this paper was to ascertain whether the diffrences in the income distributions of well-defined identity groups suggest stratified South Afica. In particular, we sought to examine the income stratification concept in order to understand its patterns, trends and the differential effects of various income sources to its measures. To this purpose, we combined the Analysis of Gini (ANOGI) and a shift-share approach derived from Larrimore (2014), Daly and Valletta (2006) and Burtless (1999) to enhance of our understanding of income stratification in the South African context, The main reason behind this combination was examine how changes in a range of income sources account for changes in distances and differences between groups. While this analysis does not provide any causal relationship, it describes the impact of the considered income sources on income stratification. We define the extent to which income distribution of different racial groups

overlap with each other and overall population as the measure of income stratification.

The main goal of the paper is to investigate the extent of income stratification using South African example. Embedded within this goal are three distinct objectives. First, apply the Analysis of Gini (ANOGI) approach to a harmonized household income per capita, to examine the patterns of both between-group and within-group inequality as well as the impact of overlapping on both. To this effect, we categorised all individuals income into four groups based on their race. Our results from this racial decomposition reveal that, on average, the stratification of income distribution according to races has fallen. This is indicated by the fact that mean ranks of other races besides Blacks have fallen between 1993 and 2008 and remained relatively constant in 2012. Lastly, we found that Whites are more stratified (i.e. they overlap less with the overall population and with other races) and the other racial groups, in general, were less stratified in 2008 than in 2012.

Second, we sought to investigate whether the observed changes in the South African income distribution suggest the emergence of new strata replacing as a defining stratum. Inspired by Heller and Yitzhaki (2006), who contended the the grouping with the highest stratification is the best. Thus, we used three groups based on the number of years of schooling of the household head to measure the extent of stratification in the country. Applying ANOGI to this grouping we discovered that, compared to races, the years of schooling had the highest intra-group variability as well as overlapping between groups. Thus races were still the better classifier. However, in 2012, the years importance of years of education starting to emerge.

Lastly, we a shift-share micro-simulation technique to assess how changes in a range of income sources interact to account for changes in household income distribution and income stratification. Our simulated effects complement the past research on the importance of labour income as the major factor accounting for changes in household income in South Africa. The increasing equalizing effects of government transfers also underscore their importance in total income in the country. As expected the effects on income stratification are mixed, which highlights the differential impact on these income sources on the racial groups. Nonetheless, the simulated effects of government transfers, in general, increased the extent overlapping of the groups. Hence government transfers have a tendency to lower stratification in the country. Defining stratification exclusively along income, and subsequently constructing

References

Allanson, P. (2014). Income stratification and between-group inequality. *Economics Letters*.

Burger, R. and McAravey, C. (2014). What does the middle class mean in a polarised, developing country such as south africa?

Burtless, G. (1999). Effects of growing wage disparities and changing family composition on the us income distribution. *European Economic Review*, 43(4):853–865.

Ceccarelli, C., Giorgi, G. M., and Guandalini, A. (2014). Is Italy a melting pot? *RIEDS*

- *Rivista Italiana di Economia, Demografia e Statistica - Italian Review of Economics, Demography and Statistics*, 0(3-4):23–30.

Cowell, F. A. and Fiorio, C. V. (2010). Inequality decompositions. *GINI Discussion Papers*, 4.

D'Agostino, A., Regoli, A., Cornelio, G., and Berti, F. (2015). Studying income inequality of immigrant communities in italy. *Social Indicators Research*, pages 1–18.

Dagum, C. (1998). A new approach to the decomposition of the gini income inequality ratio. In *Income Inequality, Poverty, and Economic Welfare*, pages 47–63. Springer.

Daly, M. C. and Valletta, R. G. (2006). Inequality and poverty in united states: the effects of rising dispersion of men's earnings and changing family behaviour. *Economica*, 73(289):75–98.

Deutsch, J. and Silber, J. (1999). Inequality decomposition by population subgroups and the analysis of interdistributional inequality. In *Handbook of income inequality measurement*, pages 363–403. Springer.

DiNardo, J., Fortin, N. M., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica: Journal of the Econometric Society*, pages 1001–1044.

Finn, A. and Ranchhod, V. (2013). Genuine fakes: The prevalence and implications of fieldworker fraud in a large south african survey.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. *Handbook of labor economics*, 4:1–102.

Frick, J. R. and Goebel, J. (2008). Regional income stratification in unified germany using a gini decomposition approach. *Regional Studies*, 42(4):555–577.

Frick, J. R., Goebel, J., Schechtman, E., Wagner, G. G., and Yitzhaki, S. (2006). Using analysis of gini (anogi) for detecting whether two subsamples represent the same universe the german socio-economic panel study (soep) experience. *Sociological Methods & Research*, 34(4):427–468.

Gradín, C. (2013). Race, poverty and deprivation in south africa. *Journal of African Economies*, 22(2):187–238.

Gradin, C. (2015). Poverty and ethnicity among black south africans. *European Journal of Development Research*.

Heller, J. and Yitzhaki, S. (2006). Assigning fossil specimens to a given recent classification when the distribution of character variation is not normal. *Systematics and Biodiversity*, 4(2):161–172.

Larrimore, J. (2014). Accounting for united states household income inequality trends: The changing importance of household structure and male and female labor earnings inequality. *Review of Income and Wealth*, 60(4):683–701.

Larrimore, J., Burkhauser, R. V., and Armour, P. (2015). Accounting for income changes over the great recession relative to previous recessions: The impact of taxes and transfers. *National Tax Journal*, 68(2):281–318.

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

Leibbrandt, M. and Levinsohn, J. (2014). Fifteen Years On: Household Incomes in South Africa. In *African Successes: Government and Institutions*, NBER Chapters. National Bureau of Economic Research, Inc.

Leibbrandt, M., Woolard, I., Finn, A., and Argent, J. (2010). *Trends in South African income distribution and poverty since the fall of apartheid.* OECD Paris.

Lerman, R. I. and Yitzhaki, S. (1984). A note on the calculation and interpretation of

the gini index. *Economics Letters*, 15(3):363–368.
Liberati, P. (2015). The world distribution of income and its inequality, 1970-2009. *Review of Income and Wealth*, 61(2):248–273. cited By 0.
Liberati, P. and Yitzhaki, S. (2012). Welfare rankings of the distribution of earnings in italy. *Journal of Income Distribution*, 21(2):38–76.
Machema, R. P. (2014). Income polarization dynamics in south africa. *AERC Biannual 2014, Lusaka*.
Mehran, F. (1975). *A statistical analysis of income inequality based on a decomposition of the Gini index*.
Milanovic, B. and Yitzhaki, S. (2002). Decomposing world income distribution: Does the world have a middle class? *Review of Income and Wealth*, 48(2):155–178.
Monti, M. and Santoro, A. (2009). A note on between-group inequality with an application to households. *Journal of Income Distribution*, 18(3-4):49–62.
Monti, M. and Santoro, A. (2011). Stratification and between-group inequality: A new interpretation. *Review of Income and Wealth*, 57(3):412–427.
Pyatt, G. (1976). On the interpretation and disaggregation of gini coefficients. *The Economic Journal*, pages 243–255.
Yitzhaki, S. (1994). Economic distance and overlapping of distributions. *Journal of Econometrics*, 61(1):147–159.
Yitzhaki, S. and Lerman, R. I. (1991). Income stratification and income inequality. *Review of income and wealth*, 37(3):313–329.
Zacharias, A. and Vakulabharanam, V. (2011). Caste stratification and wealth inequality in india. *World Development*, 39(10):1820–1833.
Zhou, X. (2012). A nonparametric index of stratification. *Sociological Methodology*, 42(1):365–389.

A Appendix

A.1 Variables included in computing income sources

The five major components of total household monthly income are:

1. **Labour market income** includes earnings from main, secondary and part-time work, self-employment, profits, bonuses and benefits and cash allowances.
2. **Capital income** includes interest or dividends, rental and/or property earnings, private pensions and annuities as well as royalties.
3. **Remittances** include value of food, clothing and housing received, alimony, maintenance and similar allowances and all other inter-household remittances received.
4. **Government transfers** include old age and/or social pensions, grants for disability, children, foster care and care dependency, UIF income as well as workman's compensation.
5. **Other income** includes all income not specified elsewhere.

A.2 Descriptive Statistics

A.2.1 Table of summary statistics

A.3 Analysis of Gini

ANOGI is derived from the decomposition of Gini coefficient proposed by Yitzhaki (1994); Yitzhaki and Lerman (1991). We use ANOGI because on top of the traditional within- and between-group inequalities, it also reports

Table A.1: Summary statistics of selected variables across racial groups

	Average income (μ_i)			Income Share (s_i)			Population Share (p_i)			Income composition		
	1993	2008	2012	1993	2008	2012	1993	2008	2012	1993	2008	2012
Black/African	670	1069	1326	0.3449	0.4492	0.4977	0.7587	0.7819	0.7938	1.336	1.409	1.416
Coloured	1435	1628	1949	0.0838	0.0822	0.0829	0.0861	0.0940	0.0900	1.338	1.429	1.569
Indian/Asian	2773	4875	5472	0.0479	0.0685	0.0670	0.0255	0.0261	0.0259	1.410	1.531	1.496
White	5952	7598	8258	0.5234	0.4001	0.3524	0.1297	0.098	0.0903	1.252	1.417	1.327
Total	**1475**	**1860**	**2115**							**1.321**	**1.414**	**1.419**

	Household size			Number of Children			Adult years of Education		
	1993	2008	2012	1993	2008	2012	1993	2008	2012
Black/African	4.68	3.62	3.44	2.05	1.48	1.35	6.90	8.87	9.51
Coloured	4.74	3.86	3.97	1.87	1.41	1.46	8.33	9.25	9.52
Indian/Asian	4.26	3.68	3.85	1.43	1.14	1.08	9.91	10.98	11.24
White	3.02	2.67	2.64	0.93	0.67	0.71	11.15	12.41	12.64
Total	**4.33**	**3.52**	**3.40**	**1.79**	**1.36**	**1.28**	**7.79**	**9.36**	**9.87**

Notes: All incomes were deflated to 2012 prices. All income components calculations were conditional on receiving (i.e positive non-missing values). Income composition is the average number of income sources in the households. All individuals aged less 18 years are classified as children while adults are individuals aged 18 to 65 years.
Source: Own calculations from weighted PSLSD 1993, NIDS 2008 and 2012

the degree of overlapping (or stratification) of the population subgroups and its role in affecting both the within- and between-group inequalities. ANOGI is "the equivalent of ANOVA (analysis of variance) performed with the Gini concentration ratio", but ANOGI offers more than ANOVA because it encompasses an indicator of stratification - the overlap term. Stratification captures the extent to which population subgroups occupy distinct strata within an overall distribution. We follow the exposition presented in Liberati (2015) and Monti and Santoro (2011). Letting y_i, s_i, $F_i(y)$, $f_i(y)$, μ_i, p_i represent the income, income share of group i, cumulative distribution, the density function, the mean, and the share of group i in the overall population, respectively. Yitzhaki and Lerman (1991) defined the stratification index, Q_i as a measure of overlap between members of group i and the rest of the population. Formally;

$$Q_i = \frac{cov_i\left[y, (F_i - F_{ui})\right]}{cov_i\left[y, F_i\right]} \tag{A.1}$$

Q_i is a ratio of two terms, such that on the numerator, we have the covariance between the income of the individuals in group i and these individuals' difference in ranking in their own group and in the rest of the overall population. The denominator, which can be treated here as a normalizing factor, is the covariance for individuals in group i between the incomes and the rankings in their own group.

Using the Lerman and Yitzhaki (1984)'s Gini coefficient (G) which is expressed as twice the covariance between income and its rank divided by the overall mean income - formally expressed as $G = {}^{2cov(y,F(y)}/_{\mu_u}$, and the above definition of Q_i Yitzhaki (1994) proved that Gini coefficient decomposition has an indicator of stratification. Assuming that the population is composed of identity groups (i.e race and household head's education) such that $F_u(y) = \sum_i p_i F_i(y)$ (i.e cumulative distribution of population is the weighted average of distributions of the groups, weighted by the relative size of population in each group).

Then, according to ANOGI, the Gini coefficient (G_u) is decomposed as follows:

$$G_u = \sum_{i=1}^{n} s_i G_i O_i + G_b \tag{A.2}$$

$$= \underbrace{\sum_i s_i G_i}_{\text{Intra-group (IG)}} + \underbrace{\sum_i s_i G_i \left(O_i - 1\right)}_{\text{Impacit of stratification on IG (IGO)}} + G_{bp} + \underbrace{\left(G_p - G_{bp}\right)}_{BGO} \tag{A.3}$$

where $G_b = {}^{2cov(\mu_i, \bar{F}_{ui})}/_{\mu_u}$ is the Gini coefficient of between-inequality; $G_{bp} = {}^{2cov(\mu_i, \bar{F}_i)}/_{\mu_u}$ is the between-Gini coefficient of the Pyatt (1976)'s decomposition; and $O_i = O_{ui} = \frac{cov_i(y, F_u(y))}{cov_i(y, F_i(y))}$ is the overlapping index of the distribution of identity group i with the overall national distribution. The numerator of O_i is the covariance between incomes in group i and their ranking in the national distribution - $F_u(y)$, while the denominator, which can be treated here as a normalizing factor, is the covariance between the same incomes and their ranking within each identity group. This means that $O_i = 1$ if the incomes of a given group i have the same ranking as the national income distribution, i.e. the two distributions perfectly overlap. More generally, $O_i < 1$ when the scatter of the ranks of a given group is narrower that that of total population, $O_i > 1$ when the scatter of ranks of the individuals in a given group i is larger than that of the total population.

The overlapping term/index can further be written as the weighted sum of overlapping of group i with all

other groups j - in symbols:

$$O_i = \sum_j p_j O_{ji} = p_i O_{ji} = p_i + \sum_{j \neq i} p_j O_{ji} \tag{A.4}$$

where p_i is the share of group i in the total population and $O_{ji} = {}^{cov_i(y,F_j(y))}/_{cov_i(y,F_i(y))}$ is the overlapping of group j by group i. The denominator is as before and the numerator is the covariance between incomes of group i and their ranking in the distribution formed by the union of group i and j. From this it follows that $O_{ji} \geq 0$, and if no member of the distribution j lies within the range of distribution i (i.e., group j is a perfect stratum) then $O_{ji} = 0$. In this case, $O_i = p_i$, i.e. total overlapping is equal to the share of group i. However, if over the range of distribution i, the shape of the distribution of group j is similar to the shape of distribution i, then $O_{ji} = 1$ and $O_i = p_i + \sum_{j \neq i} p_j = 1$. And if all observations belonging to distribution j that are located in the range of i and are concentrated at the mean of distribution i, then $O_{ji} \leq 2$, which means it is bounded from above by 2. In general, the higher the overlapping index O_{ji}, the lower O_{ij} will be. That is, the more group j is included in the range of distribution i, the less distribution i is expected to be included in the range of j. Therefore, overlapping index (O_{ji}) describes the extent to which different groups are stratified. Given this properties, it becomes clearer to note that IGO (see equation A.2 on the previous page) provides either a negative or a positive revision of intra-group variability (IG) for $O_i < 1$ and $O_i > 1$ respectively. This explains how overlapping may affect within-component.

To clarify the impact of overlapping on between-inequality, we use the distinction between G_b and G_{bp}. The substantial difference between them is the way groups ranks are established. The Pyatt's (1976) decomposition defines G_{bp} by covariance between mean incomes and the rank of mean income in the distribution of mean incomes - $\bar{F}_i(y)$. Alternatively, Yitzhaki and Lerman (1991) decomposition defines $G_b$as the covariance between mean income of each group and its mean rank in the overall population -$\bar{F}_{ui}(y)$. It can be seen that, when identity are perfectly stratified (i.e. no overlapping) the mean rank of individual incomes id equal to the rank of its mean income in the total distribution, meaning $\bar{F_{ui}} = \bar{F}_i$ and $G_b = G_{bp}$. Therefore, if some incomes overlap, the two ranks differ and $BGO = G_b - G_{bp} < 0$, or alternatively, $\frac{G_b}{G_{bp}} < 1$. It is always non-positive, because it reduces the ability to distinguish between groups. It reaches the upper limit (zero) if the ranges occupied by the different groups do not overlap. Thus, the ratio ($\frac{G_b}{G_{bp}}$) is used as indicator the loss of between-inequality due to overlapping. Monti and Santoro (2011) extending the work of Monti and Santoro (2009), for the case of two groups (a, p such that $\mu_a > \mu_p$ and $n_a + n_p = n$), Monti and Santoro (2011) showed that the ratio $\frac{G_b}{G_{bp}}$ depends on the number of transvariations occurring between the two groups. Hence it is a function of the probability of transvariation. Formally

$$I = \frac{G_b}{G_B} = 1 - \frac{2\sum_{h=1}^{n_p}(R_{ph} - r_{ph})}{n_a n_p} = 1 - 2\int_{\infty}^{\infty} F(y_p)dG(y_p) \tag{A.5}$$

In equation A.5 R_{ph} denotes the rank of y_{ph} in the overall distribution and r_{ph} the rank in the distribution p. The difference $R_{ph} - r_{ph}$ is the number of transvariations in which y_{ph} is involved (or the number of members of affluent group with income lower than y_{ph}) and the term $\sum_{h=1}^{n_p}(R_{ph} - r_{ph})$ is the total number of transvariations (or the number of instances where an income of group p is greater than.an income of group a). Therefore, the ratio $\sum_{h=1}^{n_p}(R_{ph} - r_{ph})/n_a n_p$ is the probability that a random member of the poorer group on average is richer than a random member drawn from the (on average) richer group. The last part says that the measure of impact of overlapping on inequality is equal to one minus the area under the cumulative distribution function of the richer group, evaluated within the range of values of the cumulative distribution function of the poorer group. Allanson (2014) extended the above result to the case of more than two groups.