Decomposing Well-being Measures in South Africa: The Contribution of Residential Segregation to Income Distribution

Florent Dubois[a,b*] and Christophe Muller[a†]

[a] *Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE*

[b] *Le Mans Université*

Abstract

Despite the influential work of Cutler and Glaeser[13], whether ghettos are good or bad is still an open and debatable question. In this paper, we provide evidence that, in South Africa, ghettos can be good or bad for income depending on the studied quantile of the income distribution. Segregation tends to be beneficial for rich Whites while it is detrimental for poor Blacks. Even when we find it to be also detrimental for Whites, it is still more detrimental for Blacks. We further show that the multitude of results fuelling this debate can come from misspecification issues and selecting the appropriate sample for the analysis. Finally, we quantify the importance of segregation in the income gap between Blacks and Whites in the post-Apartheid South Africa. We find that segregation can account for up to 40 percent of the income gap at the median. It is even often a larger contribution than education all across the income distribution.

Keywords: Post-Apartheid South Africa, Generalized Decompositions, Income Distribution, Residential Segregation

1 Introduction

"Are ghettos good or bad?" is the important question that Cutler and Glaeser[13] tackle in their seminal work. They find in the United States that, on average,

*corresponding author: florent.dubois@univ-lemans.fr

†email address: christophe.muller@univ-amu.fr

Blacks in more segregated areas earn less than White and Blacks in less segregated areas,[1] implying a negative correlation between segregation and income. Although influential, their work do not settle at all the debate on whether segregation is a good or bad thing. For instance, Collins and Margo[12] show that the negative effect found is a rather new phenomenon when put in an historical perspective as it starts only in 1970. Moreover, if segregation was bad for income, then it should not be rational for individuals of higher income classes to form new ghettos which implies a positive correlation between income and segregation.[2] This kind of mechanism was demonstrated theoretically by Sethi and Somanathan[36] while its empirical existence was assessed by Bayer et al.[2] Moreover, the relative importance of segregation compared to other determinants of income is yet to be assessed.

In this paper, we provide evidence that segregation can be simultaneously good and bad for income in the post-Apartheid South Africa. We analyse the contribution of segregation on the whole income distribution rather than just the mean for both Blacks and Whites. Segregation has a negative impact only for Blacks at the bottom of the distribution while it has a positive impact for Whites at the top of the distribution only. We further provide decompositions of the whole income distribution in a Oaxaca-Blinder fashion. Even after dealing with selection issues, segregation still contributes positively to the Black-White income gap as one of the main determinant. For instance, in 2010, it represented 17.4% of the income gap at the first decile, 39.6% at the median, and 10.4% at the 90% quantile. Our decomposition results suggest explanations based on different individual behaviors at the top and the bottom of the income distribution. We also demonstrate how the correlation between income and segregation can be manipulated by making sample restrictions in this kind of analysis.

For more than 15 years, a large body of the literature develops around the Moving To Opportunity (MTO) experiment in the United States. But contrary to Cutler and Glaeser's findings, Katz and coauthors in a series of influential papers[3] could not uncover any significant improvement in the economic condition of individuals moving from a lower quality neighborhood to a higher one, contrary to what was previously found in the parent natural experiment, the Gautreaux

[1]They also find that Blacks are worse off, both in absolute terms and compared to Whites, in schooling outcomes, labor market participation, and single parenthood.

[2]Although the causality goes from income to segregation in this case.

[3]See Katz et al.[25], Kling et al.[27], Ludwig et al.[29]

Program.[4] Only very recently,[5] they find that the effect produced by moving to a higher quality neighborhood might depends on the age at which the individuals move. For children younger than 13, the effect is large, positive, and significant. These children will earn in average 31% more than their counterpart in the control group. However, for older children, the effect is negative if there is any. Their argument being that the duration of exposure to a better environment might be crucial for determining the impact of segregation. But the lack of evidence of neighborhood effects might also be due to the restriction to section 8 vouchers only, which induces a smaller change in neighborhoods quality. But when the analysis is conducted on MTO vouchers,[6] then a positive effect on labor market outcomes can be recovered from a move to a higher quality neighborhoods.[7] This would argue in favor of the Cutler and Glaeser's findings. Oreopoulos[33] in another natural experiment in Toronto public housings finds that neighborhood cannot explain large amount of the variance in labor market outcomes, contrary to family differences.

Another branch of the literature using mostly observational data on immigrants residential segregation provides mixed evidence on the impact of segregation on income levels. Cutler et al.[15] show that the negative relationship found in earlier works might be due to endogenous sorting into ghettos. Once this selection is taken out, a positive relation is found on average. However, the effect is not the same for all groups of immigrants, the least educated groups being negatively affected, while the most educated actually benefit from ethnic concentration. Edin et al.[17] find similar effects with a natural experiment in Sweden. These results are in line with those of Borjas[5][6][7] who shows that the accumulation of capital is partly determined by the levels already accumulated in the neighborhoods. This part of the literature is largely inspired by a consequence of the *spatial mismatch* hypothesis developed by Kain[23]. He postulates that Black ghettos are located far away from job opportunities, thus explaining the high unemployment rates in ghettos.

[4] The Gautreaux Program was a court-ordered randomized mobility program after the lawsuit *Dorothy Gautreaux vs Chicago Housing Authority* in 1966. The administration was convicted of violating the 1964 Civil Rights Act by constructing public housing only in neighborhoods predominantly inhabited by poor minorities. See Keels et al.[26]

[5] See Chetty et al.[11]

[6] In the MTO experiment, individuals were assigned to three different groups. The first group was offered MTO vouchers which are section 8 vouchers restricted to low-poverty neighborhoods only during one year, then unrestricted after. Members of this group were also offered counseling and education by a local non-profit. The second group was offered standard section 8 vouchers but no counseling. Finally, the last group was the control group receiving project-based assistance. Section 8 vouchers pay part of the tenant's rent in the private market while project-based assistance is tied to a particular dwelling.

[7] See Aliprantis and Richter for more details.

Wilson[8] goes further by advocating that individuals in poor ghettos are lacking positive role models because the most able individuals moved to better neighborhoods, thus developing bad work habits, and deteriorating their ability to get and keep good jobs. Finally, Bayer et al.[3] demonstrate that the probability for two individuals to work at the same place is significantly greater when they live in the same neighborhood, and the effect is greater when they share sociodemographics.

Our work lies mostly in the second stream of the literature since we use mostly observational data coming from the National Income Dynamics Study and the South African Census. Contrary to all these previous works, we are, to the best of our knowledge, the first to study South Africa on this matter. We also differentiate by not studying only the mean but also other quantiles of interest for the complete South African income distribution. This allows us to take into account the heterogeneous impact of segregation along the income distribution.

The rest of the paper proceeds as follows: the next section exposes the methodology used and discusses the identification problem of having segregation measures in the standard Oaxaca-Blinder framework. The third section demonstrates how sample selection and levels of aggregation can be used to manipulate the correlation between segregation and income. Section 4 examines the potential nonlinear effect of segregation on income through the lens of a nonlinear regression function. Section 5 describes the South African context and the data used. Section 6 presents the results of the regression analyses and the decompositions for the mean. The following section shows the results along all the South African income distribution. Then the final section concludes.

2 Methodology

2.1 How to measure segregation

There is a lot of different way to measure segregation. They are designed to measure segregation as the propensity of individuals to live with similar peers separated from other groups. The most standard measures assume a partition of the city[9] as given and use the information of the subdivision of the city to compute an index for the city.[10] However, they are not conceptually identical and measure different aspect of segregation. Massey and Denton[30] propose to consider five distinct dimensions of segregation: evenness, exposure, concentration,

[8] See Wilson, 1987, The truly disadvantaged: The inner city, the underclass, and public policy.

[9] Segregation measures can also be used at the country level. We will stick to the city term to the remainder of the paper as residential segregation is mostly discussed at the city level.

[10] Two recent measures developed by Echenique and Fryer[16] and Mele[31] propose alternatives that are independent of a given partition of a city.

centralization, and clustering.

Evenness refers to the degree of overlap between the spatial distributions of the two groups.[11] As such, it borrows a lot of similarities and tools to the inequality literature. Thus, you can find evenness indices such as the Gini Index, the Theil Index of entropy, or the Atkinson Index. However, the most prominent index in the empirical literature on segregation remains the Dissimilarity index. This index measures the proportion of the minority group that would have to relocate in order to achieve an equal spatial distribution. Its formula in the case of two groups, say Blacks and Whites, for a partition of the city into I districts, is:

$$Dissimilarity = \frac{1}{2}\sum_{i\in I}|\frac{White_i}{White_{Population}} - \frac{Black_i}{Black_{Population}}| \tag{1}$$

where $Group_i$ is the number of "Group" individuals in the location i, $Group_{Population}$ is the total number of "Group" individuals in the population.

Exposure measures the degree of potential contact between the two groups. The two mostly used measures of exposure are the interaction and isolation indices. They measure respectively the average probability to interact with a member of the other group, or with a member of the same group. Only in the two-group case, the sum of these two indices is equal to one. Their formula is:

$$Interaction = \sum_{i\in I}\frac{White_i}{White_{Population}}\frac{Black_i}{Total_i} \tag{2}$$

and

$$Isolation = \sum_{i\in I}\frac{White_i}{White_{Population}}\frac{White_i}{Total_i} \tag{3}$$

where $Total_i$ is the total population of location i.

Evenness and exposure measures have been widely studied in the literature for the simple reason that the measure of segregation have a lot in common with the measure of inequalities. Most measures of evenness can be derived from the segregation curve the same way inequality indices are derived from the Lorenz curve. Segregation indices are usually evaluated from a minimal set of properties to fulfill. In the context of school segregation (Frankel and Volij[20]), these properties are labelled *Scale Invariance, Symmetry, Independence, School Division Property, Composition Invariance, Group Division Property*, and various aggregation and decomposability properties which are not interesting for our purpose.[12] The *Scale*

[11] There are some generalizations in a multi-group context but for the sake of simplicity, we will consider only the two-groups case for the remainder of this section.

[12] Interested readers can refer to the paper of Frankel and Volij[20] for more details.

Invariance property means that the measure of segregation is not affected by an identical variation in all the ethnic groups. All indices have this property. *Symmetry* implies that the segregation measure is the same for all groups. *Independence* refers to segregation measures in which compositional changes inside a subdistrict are not affected by the other subdistricts. The *School Division Property* states that segregation cannot diminish if two districts are split into two (or more). This property implies a cautious choice of the appropriate aggregation level for applied researcher. *Composition Invariance* suggests that the segregation measure does not depend on the total size of the groups. Finally, the *School Division Property* signifies that segregation cannot diminish if a group is split into two (or more). This last property is essential for multi-group measures of segregation. Actually, the main conclusion of the literature on this question is that there is no perfect index fulfilling all these desirable properties.[13] Then researchers have to choose carefully the most appropriate index with respect to their application.

Since these standard measures are subject to several problems (sensitivity to the level of aggregation, sensitivity to boundaries changes, same measure of segregation for all the individuals living in the same neighborhood...), Echenique and Fryer[16] proposes an alternative measure of exposure that relies on the network structure at the individual level. However, in practice such information is not available and researchers have to make strong assumptions on the precise level of aggregation which approximate accurately enough the individual network structure.

Concentration captures the share of physical area occupied by one group. It is seldom used in empirical studies since it requires detailed geographic information about the areas. The Delta index is the main measure of concentration used in the literature. Its formula derives from the dissimilarity index as:

$$Delta = \frac{1}{2} \sum_{i \in I} \left| \frac{White_i}{White_{Population}} - \frac{Area_i}{Area_{Total}} \right| \tag{4}$$

where $Area_i$ is the area of the location i, and $Area_{Total}$ is the total area of the city or country under study.

Centralization reflects the propensity to live close to the city center. However, it is not clear how a centralization index should be interpreted for the analysis of segregation. The idea comes from the fact that minorities (essentially Blacks) in the U.S. tend to live in the oldest and poorest housing units in the city centers. Thus, a more centralized group would have been more segregated. In different contexts, this is not the case and centralization might not mean the same thing for segregation. For instance, in Europe, the center of the biggest cities is mostly occupied by rich individuals coming from the majority group.[10]

[13] See Hutchens[21], Echenique and Fryer[16], and Frankel and Volij[20] for instance.

Several indices exist to measure centralization. The share of individuals of group j living in the limit of the inner city in a metropolitan area is the simplest. However inner city boundaries reflect essentially the expansion of the city and not necessarily a segregation process. Moreover, this index does not take the spatial distribution of a group into account. The two other measures (Relative Centralization Index and Absolute Centralization Index) rely on the distance between a location and the central business district. Yet, the central business districts might not be good reference points and we might have to make some assumptions about the computation of distances.

Finally, clustering measures the propensity for individuals to live in a location next to a neighborhood with a similar racial mix. Having one big cluster of contiguous neighborhoods inhabited by the same group implies probably more segregation than the same neighborhoods scattered all over the metropolitan area. There are several indices that can capture this feature, but their calculus requires a sufficiently low level of aggregation to be meaningful. Further, even in the case of disaggregated data, and even with approximations of contiguity by some decay functions, the computational burden is still heavy for those indices.

2.2 Oaxaca-Blinder decomposition in linear model

Decomposition techniques answer the question "how much of a quantity of interest can be attributed to specific factors?". Since the seminal work of Oaxaca[32] and Blinder[4], we know how to decompose the mean outcome difference into two parts: an "explained" part attributable to differences in characteristics between the two groups, and an "unexplained" part attributable to differences in the estimated coefficients for the two groups. Take the example of a regression model in which the only regressor is the worker education level. Then, in the first effect, the mean difference in income levels between Blacks and Whites would be attributable to the fact that Blacks and Whites do not have the same level of education for instance. In turn, in the second effect, this difference would be attributable to different returns of education for Blacks and Whites.

More formally, let $\Delta\bar{Y}$ be the mean outcome difference between the average income of Blacks and Whites, such that:

$$\Delta\bar{Y} = \mathbb{E}[Y_W] - \mathbb{E}[Y_B] \tag{5}$$

where $\mathbb{E}[Y_j]$ stands for the unconditional expected income of group $j = \{W; B\}$. When needed, we will hereafter denote group membership by the variable D which will take on one of the j's values.

Note that this difference is explicitly observed but it will not say much more than Blacks and Whites have different income (or not) on average. Because we only observe the income of Whites for Whites, we need to know the counterfactual average income of Whites if they would have had the characteristics of Blacks in order to make meaningful comparisons of the same individuals. Thus, by adding and subtracting this counterfactual average income, we can rewrite our problem as the following:

$$\begin{aligned}\Delta\bar{Y} &= \mathbb{E}[Y_W] - \mathbb{E}[Y_B] \\ &= \mathbb{E}[Y_W|D=W] - \mathbb{E}[Y_B|D=B] \\ &= \mathbb{E}[Y_W|D=W] - \mathbb{E}[Y_W|D=B] + \mathbb{E}[Y_W|D=B] - \mathbb{E}[Y_B|D=B]\end{aligned} \tag{6}$$

where $\mathbb{E}[Y_j|D=j]$ is the observed expected income for member of group j, while $\mathbb{E}[Y_j|D=k]$ with $k \neq j$ is the counterfactual expected income for member of group j if they had the characteristics of group k members.

The first part, $\mathbb{E}[Y_W|D=W] - \mathbb{E}[Y_W|D=B]$, is usually termed the composition effect, reflecting the difference in characteristics. The second part, $\mathbb{E}[Y_W|D=B] - \mathbb{E}[Y_B|D=B]$, is called the structure effect, reflecting the difference in coefficients (or returns).

We now assume that a correct representation of income is a standard linear model of the form:

$$Y_{ij} = \mathbf{X_{ij}}'\boldsymbol{\beta_j} + \epsilon_{ij} \tag{7}$$

where Y_{ij} is the income of individual i of group j, $\mathbf{X_{ij}}$ is a column vector of individual characteristics[14] such as age, experience, higher diploma for instance, $\boldsymbol{\beta_j}$ is the associated vector of coefficients, and ϵ_{ij} is the unobserved heterogeneity specific to individual i of group j, normally and independently distributed with mean 0 and variance σ.

Thus, the final decomposition in equation (6) can be estimated in the following way:

$$\Delta\hat{\bar{Y}} = (\bar{X}_W - \bar{X}_B)\hat{\beta}_B + \bar{X}_W(\hat{\beta}_W - \hat{\beta}_B) \tag{8}$$

where $\bar{X}_j$ is the sample average estimate of the conditional expectation $\mathbb{E}[X|D=j]$, and $\hat{\beta}_j$ is the OLS estimate of β_j. The first term is the estimate of the composition effect, while the second estimates the structure effect. The linearity assumption (equation (7)) is essential to perform a simple detailed decomposition. In this case, each contribution of a specific covariate sums up to reconstitute the total composition and structure effects as:

[14]This vector might include a constant.

$$\Delta\hat{\hat{Y}} = \sum_{h=1}^{H}(\bar{X}_{Wh} - \bar{X}_{Bh})\hat{\beta}_{Bh} + \sum_{h=1}^{H}\bar{X}_{Wh}(\hat{\beta}_{Wh} - \hat{\beta}_{Bh}) \tag{9}$$

for H covariates.

2.3 Why evaluating segregation in the decomposition framework is hard

Our problem is to quantify the relative contribution of segregation to the income distribution. However, if we include directly in our linear framework a segregation measure which is specific to the location (such as the dissimilarity index), and not to the group, it is no longer clear what the composition effect will capture. The risk being that an incorrect part of the gap will be attributed to the composition and structure effects. In fact, every index that is symmetric (Frankel and Volij[20]) will have this feature. To clarify the problem, consider the linear framework as in equation (7). Now, assume that one of the covariates is a measure of segregation S_{ijl} for the individual i of group j who resides in location l with its associated coefficient γ.

$$Y_{ij} = \mathbf{X_{ij}}'\boldsymbol{\beta_j} + S_{ijl} \times \gamma_j + \epsilon_{ij} \tag{10}$$

However, all individuals living in the same location have the same value no matter their group (i.e. $S_{ijl} \equiv S_l$). Thus, when trying to estimate the components of a detailed decomposition, the composition effect will not reflect a difference in the level of segregation experienced by the individuals but instead a difference in the probability to live in a location with a particular level of segregation. The closer this probability for the two groups, the closer to zero the part of the composition effect due to dissimilarity.

On the other hand, exposure and concentration measures do not have this problem of dilution as it would be both group and location specific. We provide a fictitious example of this problem in the Appendix.

As we need an asymmetric measure of segregation,[15] we will use mainly the isolation index which is one of the two most used indices in the literature. However, the isolation index is not exempt from defaults. Compared to other measures, he does not satisfy the *Composition Invariance* property which hinders the prospect

[15]In the words of Frankel and Volij[20](p.6): *"Although [Symmetry] is a standard property which is satisfied by most indices, it may not be suitable for work that focuses on the problems that face a particular ethnic group. For instance, if one is interested in the social isolation of blacks from all other groups, then one may prefer an index that treats blacks differently."* See also Lieberson and Carter[28] for more details on the isolation index.

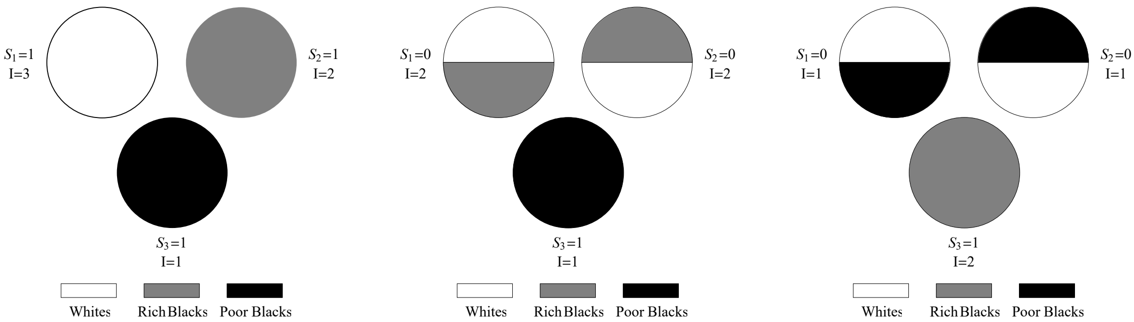
for intertemporal and inter-cities comparisons. We tackle this issue in several manners. First, this property is still debated. Proponents advance that "segregation refers to the effect of ethnic origins on destinations (schools, neighborhoods, etc.)." while adversaries "view segregation as capturing different degrees of exposure of one ethnic group to another, and thus oppose the principle." (Frankel and Volij[20], p.8). Lieberson and Carter[28] also discuss this point. Our second way to address this issue relies on the way we introduce the segregation variable in our analysis. Segregation is measured with the 2001 Census and is fixed for all individuals. It reflects the idea that segregation acts slowly on income. Since we do not use the panel dimension of our dataset, the effect of segregation does not vanish in the fixed effects. Finally, this solves the *Composition Invariance* problem at least for intertemporal comparisons. Finally, we also perform estimations with the dissimilarity index which satisfies this property. Results are very close. They are reported in the Appendix. An alternative would have been to use the Atkinson Index which is both asymmetric and invariant to the composition. However, he is rarely used in practice, and the practioner has to make arbitrary choices on the weights to put on different parts of the segregation curve. It would imply for instance that you suppose that a 60% White and 40% Black neighborhood is more segregated than a 40% White and 60% Black neighborhood, or the reverse according to the choice of the weights.

3 The chemistry behind a positive or negative correlation

The estimation of the relation between income and segregation relies heavily on the correlation between the two variables in the sample used for the regressions. Understanding precisely how these two variables relates are crucial for proper inference afterward. Consider a situation where there are three different groups of individuals who should decide in which of three locations they are going to reside. Rich Whites compose the first group, while rich Blacks and poor Blacks form the two remaining groups. One extreme case would be a situation where each group chooses a different location as in Figure 2a. Now, if the income distribution is such that Whites are all richer than rich Blacks who are themselves all richer than poor Blacks,[16] then we observe a completely segregated society stratified by income. However, as each location is completely segregated the same way - they would have for instance a dissimilarity index of 1 each - the level of segregation does not vary with the level of income and thus the correlation between the two would be exactly 0.

[16]We assume for simplicity that all individuals have the same income in each group.

A negative correlation would arise when the most segregated locations are also the poorest. If this holds strictly and segregation increases exactly in the same proportion as income decreases, then we would observe a perfectly negative correlation as in figure 2b. On the contrary, income and segregation would be perfectly positively correlated if the most segregated locations were also the most affluent. In figure 2c, we look at a case where Whites are as poor as the poor Blacks while rich Blacks are richer than everyone else.



(a) (b) (c)

Figure 1: Different configurations leading to (a) no correlation, (b) a negative correlation, or (c) a positive correlation between income and segregation. S is for segregation while I is for income.

There are two ways to make this correlation changes. Actually, either the change in one variable is no longer proportional to the other, or the sample used to study the covariance includes locations suggesting different relations between segregation and income. The first type of variation can weaken the strength of the relation but cannot reverse the sign of the relation. However, the second type of variation associated with sample selection is much more severe as it can generate the three different types of linear relations. In Figure 2, we are in a configuration where seven locations exhibit different racial mixes and income levels. If we compute the correlation between income and segregation considering the seven locations, we would obtain a correlation of roughly 0.42. Now, if we remove from the sample the locations that do not have a sufficiently high share of Blacks. In practice, several studies (Cutler and Glaeser[13], Cutler et al.[14]) consider a threshold of 10% thus removing metropolitan statistical areas (MSA hereafter) not reaching this share of Blacks. If we apply the same condition, we only use the blue subsample of figure 6, and the correlation between the two variables drops to 0.27, a 36% reduction. If we further restrict our analysis to only younger people, for instance to 20-30 years old individuals as in Cutler and Glaeser[13], we might find that it corresponds only to the red subsample of Figure 2. In this case, the

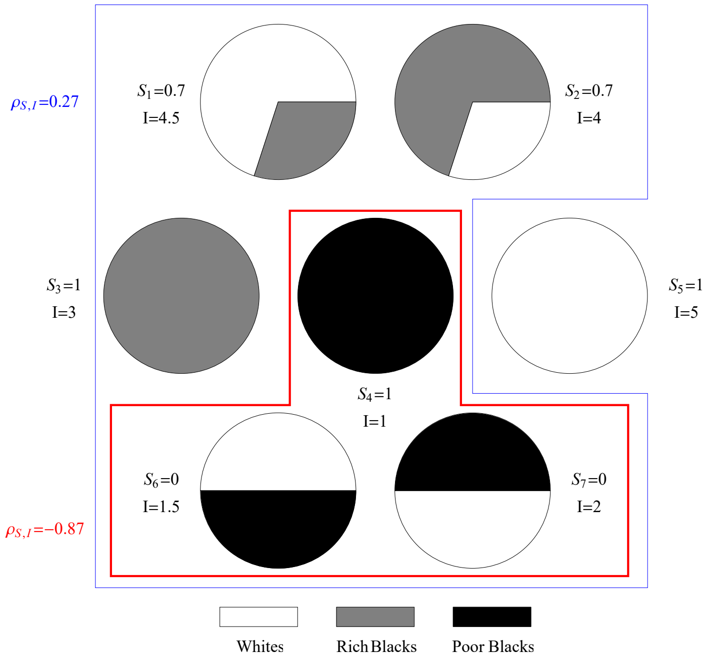
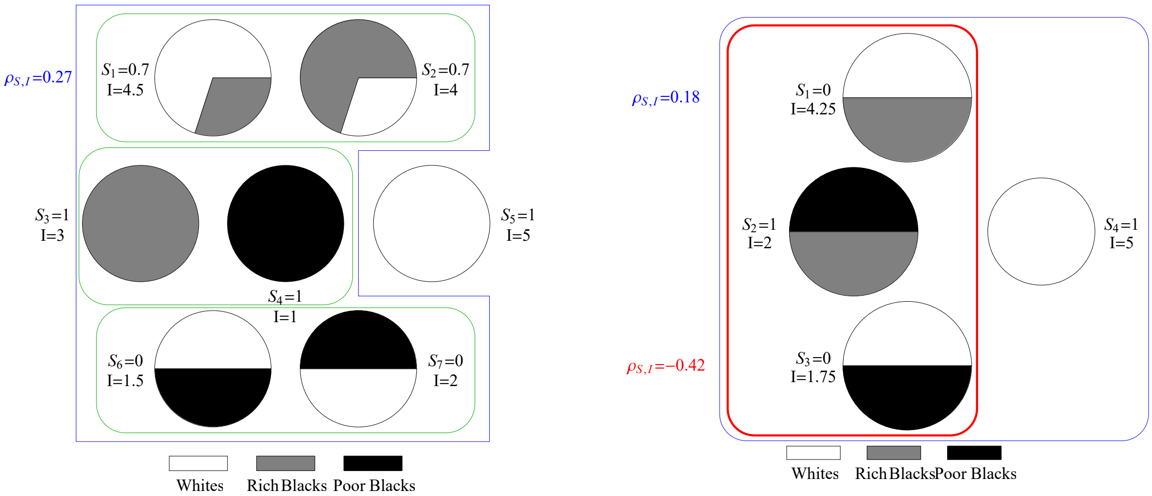


Figure 2: The impact of sampling restriction on the income-segregation correlation.

correlation becomes negative and much stronger, rising by roughly 107% to -0.87.

The choice of a specific geographic scale for segregation is actually as crucial as the sample selection issue. As segregation is sensitive to the boundaries used to compute indexes, it is lowered when the level of aggregation increases since some segregated neighborhoods will be diluted with more integrated neighborhoods. Choosing higher level of aggregation imposes *in fine* another implicit selection issue. In the previous example, if we aggregate locations, except the rich White ghetto, by pairs because they are neighboring areas, we would find a lower positive correlation of 0.18. If we further combine this restriction with a minimal number of Blacks, we would obtain a negative correlation of -0.42. These two situations are depicted in Figure 3.

Then two questions remain. What is the appropriate geographic scale for conducting an analysis of segregation? Then, what should be the relevant sample for analysis? Unfortunately, for the first question, there are no objective evidences in favor of one scale to another. The empirical researcher will have to make normative choices about the relevant scale. However, we might think that a scale which is not related to a competent local authority would not be appropriate since it cannot use this information for public policies. As an example, magisterial districts in South Africa are a division of the judiciary system but is not empowered to manage public housing projects for instance. Similarly, a scale which is too



(a) (b)

Figure 3: Changing the level of aggregation of the data can generate (a) a weaker positive correlation or (b) a stronger negative correlation.

aggregated would not mean anything since it is hardly believable that this scale has any significance for the individuals. The MSAs are probably useful for public policies but individuals do not choose their location according to the characteristics of the MSA. They probably do not interact also with the 100 000 individuals that an MSA comprises. So, we would like to find the relevant definition of a neighborhood for individuals, but this relevant neighborhood might first be too small to have a reliable measure of segregation, and then might overlap with other administratively defined neighborhoods. As we cannot define an explicit neighborhood for each individual, a good measure of the neighborhood should be the name of the neighborhood of the local economic life of the individuals, *i.e.* the neighborhood where they go to school, go to groceries, do sport... In South Africa, subplaces play this role, while in the US census tracts fill this purpose. This kind of neighborhoods averages around 2000 for subplaces and 4000 for census tracts. For South Africa, the main place level which regroup 10000+ individuals might still provide some good information about segregation as well. However, the levels below subplaces (enumeration areas) and census tracts (census blocks) might not constitute a good definition of neighborhood.

Which sample should be used to conduct such analysis of segregation is not an easy question as well. From the previous demonstrations, restrictions such as a minimal size of the minority or concentrating only on urban areas exclude de facto areas like rich white homogeneous ghettos or rural black areas. Then these kinds of restrictions neglect a priori phenomenon such as rural-urban migration or minor the importance of the White segregation. Then, an appropriate sample should contain all the possible locations, excluding only technical areas which are not meant for living such as industrial zones or national parks, and areas where a problem of enumeration might have occurred resulting in missing observations.

In this paper, we will take these two aspects into account for as many subplaces as possible. We are constrained by the data we are using. Ideally, we would like to compute segregation measures at the subplace level but the public data of the NIDS do not allow for this possibility. We only have information about the municipal districts they are living in. However, we are able to compute precise measures of segregation at these municipal districts since we have data on the full count of the census in 1996, 2001 and 2011 aggregated by different local areas, and most importantly by subplaces.

Finally, this discussion on the impact of two important methodological choices in segregation studies, relevant sample and aggregation level, have some resemblances with the literature on the properties of segregation indices. However, we would like to attract your attention on the differences that, in fact, anchor this discussion much more in an applied view of the good statistical practices in empirical segregation studies. First, the object of interest is different. We study the correlation between segregation and income, whereas the literature on segregation indices is interested only in measuring segregation. This literature is actually concerned with a more primitive question which is the question of properly measuring segregation. Our question arrives later in the empirical research process: once we know how to properly measure segregation, what consequences do our methodological choices have on the relation between segregation and income? Then, the choice of the appropriate sample is a pure statistical feature of the empirical analysis which does not correspond to any axiom. The impact of the choice of the appropriate geographic frame does, indeed, look similar to the axiom labelled *School Division Property* in Frankel and Volij[20]. However, all the standard indices (Atkinson, Dissimilarity, Mutual Information, Gini, H-index of entropy, Isolation (and Normalized Exposure)) have this property, reaffirming our point that the correlation between income and segregation can be manipulated independently of the chosen index and its properties.

4 The impact of segregation on income

Estimating the impact of segregation on income is hard for several reasons. First, there exists both highly segregated poor black neighborhoods and highly segregated rich white neighborhoods. As explained above, we can expect the effect to be negative in these black neighborhoods and positive in these white neighborhoods. If there would exist only these kinds of highly segregated neighborhoods,

Table 1: OLS regressions of isolation on income

	All	Whites	Blacks
Isolation	-1.84***	0.72***	-0.00
	(-26.92)	(3.36)	(-0.01)
Constant	9.66***	8.69***	7.90***
	(154.48)	(61.63)	(90.01)
Observations	16387	1153	15234
R^2	0.042	0.010	0.000

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

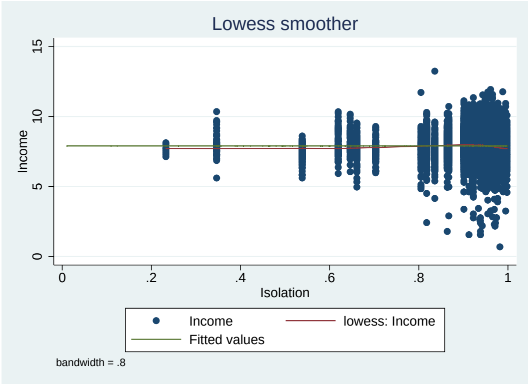
we would simply have to run a regression specific for blacks and another for whites. However, there also exists highly segregated rich black neighborhoods and highly segregated poor white neighborhoods. This fact implies that at both extremes of the income distributions of the two groups, highly segregated neighborhoods occur. In between, relatively more integrated neighborhoods exists that are composed of middle classes.

So, the relation between income and segregation is nonlinear *per se*. Thus estimating this relation by linear econometrics can be misleading, and we might only capture the fact that the poor black segregated neighborhoods are the most common type of neighborhoods in the sample/population of neighborhoods. We estimate a nonparametric local regression to get a sense of the shape of the conditional mean function relating income and segregation[17] for Blacks, Whites, and both together for the pooled sample of all the waves of the NIDS.[18] Both Locally Weighted Scatterplot Smoothing and OLS fitted values are depicted in Figure 4.

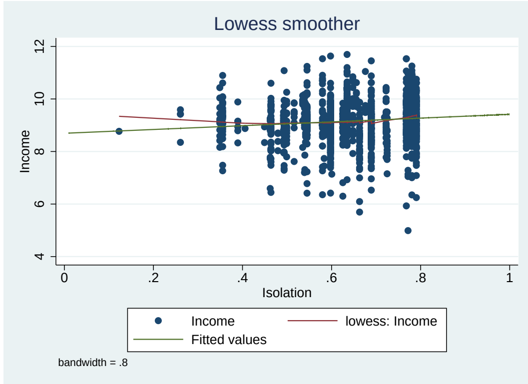
The first thing to note from this experiment is that nonlinear patterns (if any) are not revealed by using a nonparametric approach and using linear econometrics in this case should not be a problem. If there are any nonlinearities in the regression function, there seems to be a quadratic pattern for Whites only, but it seems far from evident. For Blacks, both the linear prediction and the nonparametric fit are almost flat. The same applies for the total sample. However, both for Whites and the total sample, the linear fit estimates poorly (if not incorrectly) the beginning of the relation between income and segregation. Thus, for Whites, up to 0.5

[17]In this experiment, we only present the results with the isolation index. Results with the dissimilarity index are similar and do not change the conclusions drawn. They are nevertheless available on request.

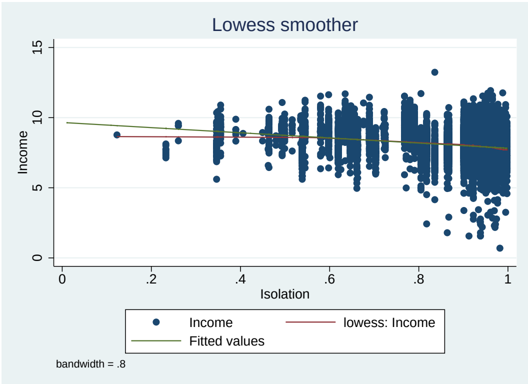
[18]We conduct these regressions on a sample of individuals being between 15 and 65 years old. We describe more precisely the data used in the following section.



(a)



(b)



(c)

Figure 4: Local regression for Blacks (a), Whites (b), and everyone (c) using the isolation index

for isolation, the relation seems negative while the linear fit suggests a positive relation. For the total sample, the relation is flat up to 0.5 for isolation while the linear fit estimates a negative relation. This indicates that misspecification issues lead to overestimate slightly the true coefficient, even though it does not look that severe.[19]

However, if we look at the coefficient estimated, we could be tempted to declare strongly significant effects of isolation on the average income. In Table 1, we report the coefficients estimated by OLS of simple regressions of isolation on income for the three samples. For Blacks, there is no effect at all since the regression function is almost flat. However, for Whites, the effect is 0.72, meaning that if isolation increases by one percentage point, then, assuming that the estimated effects were

[19] In this case, it seems to be due to very influential observations that bend the two fits differently at the beginning of isolation at which data are sparse.

causal, the average income of Whites raises by 0.72% which is quite a large effect.[20] This is even stronger when we regress with both Blacks and Whites pooled. Whites being in average richer and less segregated, this distorts the regression function to be downward sloping. If segregation raises by one percentage point, the income is predicted to decrease by 1.84%. In both cases, the effect is highly significant. However, this needs to be put in perspective with the percentage of the variance of income explained which are quite low. It represents 4.2% for the total sample, 1% for Whites, and almost nothing for Blacks.

Thus, it seems easy to be lured to misleading conclusions with linear estimations. Nevertheless, there is an important feature that appears clearly and has not drawn enough attention in the literature. All samples exhibit heteroscedasticity. For instance, for Whites, log incomes range from 5 to almost 12 when isolation is close to 0.8 while it varies only from roughly 7 to 11 around 0.35. Thus nonlinear patterns might well be present but not necessarily in the conditional mean.

Finally, as the mean is very sensitive to extreme values, linear estimations seem even more inappropriate to assess the nonlinear patterns we think to be at play with neighborhood effects. This argues in favor of a distributional analysis which we undertake in Section 7.

5 Context and Data

5.1 Residential segregation in South Africa

South Africa is a highly ethnically diverse country with no less than 11 official languages for the most represented groups. It has a long legacy of segregation starting with the Dutch colonization in 1652. It was then pioneering its institutionalization during the British period beginning in 1795. During more than 300 years, overt racism, slavery, and white supremacy were common in South Africa.

In the 1790s, manumitted slaves should carry a pass to leave Cape Town. In 1809, a series of laws obliged the Khoikhoi to register where they lived and to obtain a pass from their employer for leaving the registered place. A similar system was enforced in the Afrikaner republics in the mid-nineteenth century. Moreover, if the employer had taken care of a Khoikhoi child for eight years, this child had to serve him as an apprentice for ten years, tying *de facto* the parents too. During the 1860s, the Natal Colonial government created reserves for Africans called *locations*. Later, a dual legal system was put in place, the (revised) customary African law for Africans and the Roman Dutch law for Whites and problems involving both groups.

[20]According to this estimate, moving from an integrated place (isolation=0.26) to a highly segregated place (isolation=0.8) would induce an increase of almost 40% in average.

With the establishment of the Union of South Africa in 1910, the number of segregative laws increased a lot. On one hand, a series of laws aimed at restraining Africans to compete with Whites for skilled occupations. The Mines and Works Act in 1911 forbade African strikes in the mining industry and reserved skilled positions to Whites. A decade later, the Apprenticeship Act of 1922 imposed educational conditions for being apprentice which *de facto* excluded Africans as they had only a limited access to schools[21]. The Industrial Conciliation Act (1924) only gave the right to negotiate with the employer to Whites.

On the other hand, the territorial repartition between Whites and Africans was a thorny question. Thus, a couple of laws sharpened the difference between Whites areas and African reserves. In 1913, the Natives Land Act prohibited Africans from buying lands from non-Africans outside the reserves. The law designated around 7% of the South African territories as reserves. But despite the act planned to increase the size of the reserves, it never exceeded 12% due to the pressure of White landowners. The Natives (Urban Areas) Act of 1923 was a cornerstone in the residential segregation policy conducted. It defined White areas where Blacks could not live in as well as African areas in the South African towns. A legal authority was authorized to expel Africans from the White areas to African locations. As a complementary tool, the pass system was reinforced in order to enforce this measure. In 1930, the law was completed in order to remove the "female surplus" (see Thomson[38]) and later strengthened in 1945. These laws came for controlling the massive inflow of Africans to towns. But they were not as effective in practice as they were supposed to be in theory. Africans responded by establishing squatter camps on the fringe of the cities.

Later, influenced by a national socialist ideology, the White regime tried to limit the interactions between Africans and Whites. For instance, the Immorality Act of 1927 was particularly designed to preserve the purity of the Afrikaner race by forbidding any sexual intercourse between Africans and Whites[22]. During the 1940s, the German influence invigorated. All the Afrikaner associations[23] mobilized to express their desire to reinforce the White domination and ensure the purity of the White race. Consequently, the National Party[24] emerged as a credible opposition in 1943 and won the very tight election of 1948.

[21] This law were reinforced later by the Bantu Education Act of 1953 which prohibited Africans to attend school in the same institutions as Whites.

[22] This legislation was then completed by forbidding relationship between "Europeans" and "non-Europeans" in 1950 (See Thomson[38]).

[23] The Broederbond and the Afrikaner churches and press played a crucial role. The Afrikaner intellectuals as well were ardent advocates of the White supremacy (See Rich[34] and Thomson[38]. Finally, the Sauer report in 1946 concluded that the segregation policy, explicitly labelled *Apartheid*, should be strengthened.

[24] The National Party was the political party supporting the Afrikaner claims while the United Party represented mainly the British interests.

From 1948, the National Party had a firm control over South Africa. Their first decisions were to harden the previous segregative legislations. Thus, all people were categorized among four racial groups: Black, White, Coloured, or Asian[25]. Interracial relations were also prohibited, both between married and unmarried individuals[26]. Then the government defined again some areas to be color specific both for works and residences with the Group Areas Act of 1950. This law generated a lot of tensions between Africans and Whites as it was used to remove by force Africans from areas for allowing Whites to settle in. In 1951, the government decided to amend the reserves system. They were then transformed into ten homelands, each one designed to be the legitimate territory of a specific ethnic group and independent in a near future[27]. But the homelands were scattered into a lot of pieces of lands and represented only a small part of the South African territory. As a consequence, a lot of Africans migrated from these poor and overpopulated areas. All these legislations were easily enacted. Even when the government encountered some resistance, it actually managed to bend the constitution to reach its goal[28]. During the following years, the regime strengthened the pass laws by limiting their duration to seventy-two hours. They also destroyed the squatter camps by relocating the employed ones in township and deporting the others in the homelands. Coloureds and Asians were also displaced to specific townships. According to Thomson[38], more than 3.5 million people were deported between 1960 and 1983.

In the 1970s, the Apartheid regime faced an economic crisis reinforced by the international economic sanctions. This drove the regime to loosen the segregation laws progressively, while trying to resist on some aspects of the Apartheid policy. Finally, in 1990, F.W. De Klerk starts the political transition to a non-racial democracy. The process took four years to see Nelson Mandela elected president. He inherited a divided country in a disastrous economic situation without the means to implement all the policies necessary to improve the living conditions of South Africans. Even during the following year, economic growth was not as strong as it should have been to finance his development policy. Regarding segregation, he proposed two main policies: a program of redistribution of the

[25]This categorization, promulgated by the Population Registration Act in 1950, has settled in the South African society and remains in use for some administrative or policy purposes like the censuses for instance. However, the boundaries between each racial category were fuzzy and led to mistakes with dramatic consequences for families.

[26]As mentioned before, the Immorality Act (1950) extended the prohibition of interracial relations to "Europeans" and "non-Europeans" while the Prohibition of Mixed Marriage (1949) is explicit about its legal effects.

[27]Only four homelands got independency between 1976 and 1981, Transkei, Bophuthatswana, Venda, and Ciskei. They are known as the TBVC states.

[28]For instance, the Reservation of Separate Amenities Act (1953) legalized inequalities in public services.

lands seized by Whites during the Apartheid, and the construction of one million low-cost houses. The first failed completely while only a few hundreds of claims were granted or compensated. The second managed to build 75% of the one million houses promised, although they were mainly destined for poor Africans. Finally, his greatest achievement against segregation was perhaps to have reconciled South Africans with themselves. During the Mbeki and Zuma mandates, efforts against segregation were not continued. They choose to resort to affirmative action but it ended in nepotism. Despite an invigorated growth, criminality is still high. Africans still do not have the same quality and level of education than Whites. They still live in high poverty slums despite the emergence of an African middle-class. The country has a lot of corruption problem, the president Zuma himself being involved. Even racial concerns are regularly present in the public debate.

5.2 Data

We will use two sources of data for our analyses. The first one is a panel of households studying the evolution of their living conditions, the National Income Dynamics Study (NIDS hereafter). It is surveyed every two years for a nationally representative sample. Covering the period 2008-2014, there are four waves available. The advantage is that the same individuals are surveyed every wave, and we believe this survey to be more reliable especially for data on incomes. However, once we disaggregate the sample by racial groups, sample sizes shrink really fast.

Our second source of data will be the South African Censuses conducted in 2001 and 2011, both the 10% count and the community profiles. The 10% count has the advantage to have large sample sizes since it is an individual survey of 10% of the South African population but the number of socio-economic characteristics is rather limited. The community profiles gives the total count of the South African population but aggregated at geographic level ranging from the enumeration areas to the provinces. It has the advantage of being exhaustive but only giving informations about the distributions of characteristics inside a location. However, like the 10% count, it is subject to a lot of measurement error, and we cannot conduct individual analyses.

We will mainly use the NIDS data as our reference dataset and try to confirm the validity of our results on bigger samples with the Censuses. As we are interested in studying income differences in South Africa, we will restrict our sample to adults older than 15. Thus, our base sample is constituted of 11 775 Africans and 828 Whites. Table 2 reports mean and standard deviations of the variables used in the analysis, across ethnic groups and survey rounds. As expected, Whites are, for instance in 2008, more educated, older, and richer than Africans. They are however as segregated as Africans despite having more interactions with the other groups.

Table 2: Descriptive Statistics

	2008				2010				2012			
	White		Black		White		Black		White		Black	
	Mean	Std D.	Mean	Std D.	Mean	Std D.	Mean	Std D.	Mean	Std D.	Mean	Std D.
Income	9.09	0.89	7.70	0.91	9.20	0.86	7.80	1.01	9.19	0.85	7.96	0.90
Isolation	0.65	0.12	0.93	0.09	0.65	0.12	0.93	0.08	0.65	0.12	0.93	0.08
Dissimilarity	0.84	0.08	0.85	0.07	0.85	0.07	0.86	0.07	0.85	0.07	0.86	0.07
Years of schooling	12.82	2.03	8.92	4.36	12.99	2.14	9.68	4.03	13.04	1.98	10.08	3.87
Experience	21.48	11.69	22.68	13.23	24.55	12.08	22.03	13.08	22.95	12.47	20.95	13.01
Male	0.51	0.50	0.57	0.50	0.53	0.50	0.51	0.50	0.52	0.50	0.52	0.50
Age	40.31	11.68	37.60	11.12	43.53	12.20	37.71	11.05	41.98	12.63	37.04	11.13
Observations	440		2922		243		3090		261		3996	

	2014				Pooled				Census			
	White		Black		White		Black		White		Black	
	Mean	Std D.	Mean	Std D.	Mean	Std D.	Mean	Std D.	Mean	Std D.	Mean	Std D.
Income	9.20	0.86	8.01	0.93	9.16	0.87	7.89	0.94	9.14	1.29	7.39	1.26
Isolation	0.64	0.12	0.93	0.08	0.65	0.12	0.93	0.08	0.66	0.08	0.94	0.05
Dissimilarity	0.85	0.07	0.86	0.07	0.85	0.07	0.86	0.07	0.87	0.04	0.88	0.04
Years of schooling	13.28	1.83	10.55	3.66	13.00	2.01	9.94	3.98	12.95	2.76	8.58	4.63
Experience	24.03	13.07	20.19	12.73	22.94	12.27	21.24	13.00	28.57	17.32	27.06	20.48
Male	0.57	0.50	0.50	0.50	0.53	0.50	0.52	0.50	0.52	0.50	0.50	0.50
Age	43.31	13.23	36.74	11.01	41.94	12.38	37.18	11.07	47.51	16.72	41.63	17.83
Observations	229		5291		1173		15299		202056		1132412	

Income is the logarithm of income deflated to November 2014 Rands for the NIDS waves. It is expressed in 2011 Rands for the Census. Male is expressed as a percentage of the population.

6 Results

For the time being, we will assume for simplicity that income expectations are determined by the individuals' education and experience, possibly non-linearly. Moreover, we will assume that the level of segregation an individual is facing also affects the mean income of the individuals. One reason for this might be that individuals leaving in ghettos may be more prone to develop bad work habits (absenteeism, tardiness, low reliability...) which might reduce their income.[29] Note that our measure of segregation is fixed at the 2001 year because we cannot measure segregation from NIDS and have to rely on a measure coming from the 2001 Census. However, since segregation is most likely to take time to effect the income levels, we believe that it is not inconsiderate to proceed that way. Our model takes the following form for each individual i:

$$\begin{aligned} Y_i = \alpha + \beta_1 \times Education_i + \beta_2 \times Education_i^2 + \beta_3 \times Experience_i \\ + \beta_4 \times Experience_i^2 + \beta_5 \times Segregation_i(2001) + \epsilon_i \end{aligned} \tag{11}$$

where $\alpha, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are parameters to estimate, and ϵ_i is a centered error term.

We run this regression for Africans and Whites separately, on the one hand, and with both groups pooled, on the other.[30] Results are displayed in Table 3. They must be seen as preliminary descriptivee examination of the relation between earnings and correlates, before moving to decomposition analyses. However, if correlates can be considered as approximately exogenous, the estimated parameters would reflect causal effects. Thus, it may be useful to dwell a moment on their interpretation.

[29] See Sáez-Martí and Zenou[35] for more details on ghetto/oppositional culture.

[30] Results for the same analysis with the dissimilarity index and without any segregation index are available from the authors, and are qualitatively similar.

Table 3: Linear Regressions

	2008			2010			2012		
	White	Black	Pooled	White	Black	Pooled	White	Black	Pooled
Experience	0.08***	0.04***	0.04***	0.07***	0.04***	0.04***	0.04***	0.03***	0.03***
	(6.84)	(10.82)	(12.26)	(5.36)	(8.99)	(9.99)	(2.70)	(9.74)	(9.90)
Experience (square)	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00*	-0.00***	-0.00***
	(-5.87)	(-5.99)	(-6.87)	(-4.20)	(-4.81)	(-5.34)	(-1.96)	(-3.58)	(-3.46)
Years of schooling	-0.17	-0.05***	-0.05***	0.24	-0.03**	-0.03**	-0.02	-0.03**	-0.03***
	(-1.53)	(-4.03)	(-4.64)	(1.34)	(-2.19)	(-2.29)	(-0.07)	(-2.53)	(-2.87)
Years of schooling (square)	0.01***	0.01***	0.01***	-0.00	0.01***	0.01***	0.01	0.01***	0.01***
	(3.28)	(16.85)	(19.45)	(-0.36)	(12.55)	(13.87)	(0.85)	(15.96)	(17.21)
Isolation	1.08***	-0.40**	-1.32***	0.39	-0.64***	-1.39***	0.02	-0.24*	-1.04***
	(3.70)	(-2.57)	(-12.94)	(1.01)	(-3.29)	(-10.18)	(0.06)	(-1.66)	(-9.58)
Constant	7.32***	6.66***	7.47***	5.34***	6.90***	7.54***	7.60***	6.72***	7.45***
	(10.07)	(42.09)	(62.59)	(4.48)	(34.46)	(47.36)	(5.07)	(43.76)	(59.25)
Observations	440	2922	3362	243	3088	3331	261	3996	4257
R^2	0.355	0.346	0.450	0.359	0.263	0.338	0.227	0.314	0.361

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3 : Linear Regressions (continued)

	2014			Pooled			Census		
	White	Black	Pooled	White	Black	Pooled	White	Black	Pooled
Experience	0.03**	0.02***	0.02***	0.05***	0.03***	0.03***	0.17***	0.11***	0.12***
	(2.47)	(7.46)	(7.76)	(8.79)	(17.50)	(19.08)	(129.27)	(304.73)	(321.60)
Experience (square)	-0.00	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***
	(-1.26)	(-2.90)	(-2.68)	(-6.75)	(-7.63)	(-8.29)	(-106.49)	(-71.58)	(-102.71)
Years of schooling	-0.31	-0.06***	-0.06***	-0.07	-0.04***	-0.04***	-0.02	-0.18***	-0.22***
	(-1.13)	(-5.04)	(-5.06)	(-0.80)	(-6.18)	(-6.74)	(-1.50)	(-89.02)	(-116.69)
Years of schooling (square)	0.02*	0.01***	0.01***	0.01***	0.01***	0.01***	0.02***	0.03***	0.03***
	(1.87)	(17.32)	(18.10)	(3.10)	(31.01)	(33.83)	(35.14)	(234.62)	(272.79)
Isolation	0.55	-0.46***	-1.03***	0.60***	-0.41***	-1.18***	-0.66***	-5.39***	-4.98***
	(1.37)	(-3.22)	(-9.29)	(3.34)	(-5.15)	(-20.81)	(-7.55)	(-120.78)	(-211.16)
Constant	8.83***	7.29***	7.77***	7.22***	6.89***	7.56***	2.57***	5.56***	5.48***
	(5.00)	(47.91)	(60.64)	(13.20)	(83.74)	(115.50)	(27.54)	(124.87)	(209.04)
Observations	229	5291	5520	1173	15297	16470	263334	2248132	2511466
R^2	0.279	0.259	0.296	0.300	0.294	0.354	0.146	0.177	0.218

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In this set of regressions, we find that several expected effects. For experience, we find a positive and significant effect for both groups with decreasing returns as its square is negative. However, as experience is a function of age, it might also capture a life-cycle phenomenon, older people being richer than their younger counterpart. Isolation also behaves as demonstrated in Section 4 as the effect of isolation on income is always positive for Whites, negative for Blacks, and negative for the pooled samples. Moreover, this effect is even larger in absolute value than the effects for both groups separately. Finally, the effect of the number of years of schooling is a bit puzzling since it is negative and significant for Blacks and not significant at all for Whites. However, the negative sign of education comes from the addition of the squared education. Removing this variable or adding a cubic term for education restores the positive relation between income and education.

In the second time of the analysis, we decompose the mean in the Oaxaca-Blinder fashion in order to elicit the magnitude of the role of the different correlates, notably the segregation feature.[31] Results are displayed in Table 4. First, as South Africa is one of the most unequal country in the World, the gap between the average real monthly log incomes of the two group is considerable. It ranges between 6658 Rands and 7457 Rands depending on the year and data used. It is worth noting that this gap is declining over time. In 2010, the gap peaks at 7457 Rands while in the next four years, it decreases to reach a minimum of 6886 Rands in 2014.[32] The reduction in the gap is mainly due to the increase of the real monthly income of Blacks, as it is fairly stable for Whites between 2010 and 2014.

When we look at the decomposition of this gap, the composition and the structure effects are quite comparable in the NIDS while the composition effect is more than three times higher than the structure effect in the Census. The difference between the two samples might be due to the small White sample in the NIDS (around 250 Whites per wave). Despite the emergence of a Black middle-class, Blacks are still lagging behind in many socio-economic characteristics. A sizable and significant composition effect is thus a logical finding. However, that this composition effect is roughly equal to the structure effect is much less anticipated. It means that Blacks with similar socio-economic characteristics as Whites benefit less than Whites from these characteristics, and that this is as important as the difference in socio-economic characteristics. This might be a consequence of racial discrimination in the job or housing markets for instance. Or it might re-

[31] To avoid to transfer some of the structure effect into the composition effect, we add a group dummy for the decomposition. This point is further developed by Jann[22]. Moreover, similar results obtained with the dissimilarity index and without any segregation index have been found.

[32] According to a report on poverty levels by Statistics South Africa, the financial crisis of 2008-2009 had a major impact on the most vulnerable individuals: "The number of people living below the food line increased to 15,8 million in 2009 from 12,6 million in 2006, before dropping to 10,2 million people in 2011". This explains partly the evolution of the income gap observed.

Table 4: Oaxaca decompositions

	2008	2010	2012	2014	Pooled	Census
Differential						
Prediction_1	9.09***	9.20***	9.19***	9.20***	9.16***	9.14***
	(214.76)	(167.89)	(175.69)	(161.92)	(361.67)	(3179.77)
Prediction_2	7.70***	7.80***	7.96***	8.01***	7.89***	7.39***
	(455.18)	(429.09)	(560.12)	(626.79)	(1035.33)	(6246.91)
Difference	1.39***	1.40***	1.23***	1.20***	1.26***	1.75***
	(30.56)	(24.20)	(22.66)	(20.55)	(47.78)	(561.74)
Explained						
Experience	-0.01	0.06***	0.04***	0.06***	0.04***	0.11***
	(-0.57)	(3.68)	(2.73)	(4.13)	(5.37)	(119.91)
Education	0.64***	0.57***	0.51***	0.46***	0.52***	0.78***
	(20.78)	(14.81)	(15.74)	(14.81)	(32.28)	(340.39)
Isolation	0.02	0.13***	0.06	0.11***	0.07***	0.43***
	(0.47)	(3.34)	(1.61)	(3.22)	(3.94)	(93.84)
Total	0.65***	0.76***	0.61***	0.63***	0.63***	1.32***
	(12.75)	(12.70)	(11.76)	(12.82)	(24.28)	(262.84)
Unexplained						
Experience	0.16	0.23	-0.12	0.11	0.13*	-0.61***
	(1.34)	(1.52)	(-0.86)	(0.91)	(1.94)	(-80.73)
Education	-1.13*	1.25	-0.33	-1.76	-0.52	-0.29***
	(-1.93)	(0.98)	(-0.21)	(-0.99)	(-1.05)	(-8.66)
Isolation	1.05***	0.71***	0.19	0.68**	0.70***	1.58***
	(4.80)	(2.78)	(0.58)	(2.53)	(5.40)	(58.16)
Constant	0.66	-1.56	0.88	1.54	0.33	-0.25***
	(1.02)	(-1.17)	(0.52)	(0.85)	(0.63)	(-5.66)
Total	0.75***	0.63***	0.61***	0.57***	0.64***	0.42***
	(12.69)	(9.30)	(9.79)	(9.02)	(20.55)	(77.83)
Observations	3362	3331	4257	5520	16470	1334468

t statistics in parentheses

Prediction_1 is the predicted logarithm of the real monthly income of Whites (in 2014(Nov.) Rands).

Prediction_2 is the predicted logarithm of the real monthly income of Blacks (in 2014(Nov.) Rands).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

flect different behaviors from Blacks and Whites. For instance, if Blacks work mainly in rural areas or work in the industrial sector, having a master degree might give them a lower wage than a White working in the financial sector in an urban area. Thus, this difference in return might just reflect a premium for urban areas and/or industrial sectors. Finally, over time, when excluding the crisis wave 2010, the composition effect increases from 46.4% in 2008 to 52.7% in 2014, while the structure effect decreases from 53.6% in 2008 to 47.5% in 2014.

With a closer look at the detailed decomposition, we first note that all the groups of variables[33] contribute positively to the gap through the composition effect. This confirms what we were discussing previously, Blacks are lagging behind the Whites in terms of education and experience. The educational system still exerts school segregation with a separation in schools for Blacks and schools for Whites.[34] The impact of experience might result from the fact that Whites are on average older than Blacks. Therefore, we cannot interpret this effect as purely a lack of experience resulting from discrimination of Blacks in the job market. However, the effect is rather small.

For segregation, it is a different matter. The effect is always close to zero and at least for two waves, it is not significantly different from zero. The small sample size for the Whites might be part of the problem despite that the statistical significance is the weakest for the largest two samples. However, it can also come from heterogeneity in the isolation variable. Whites are less isolated on average, while there is much more heterogeneity since the standard errors is 1.5 higher than the one for Blacks as shown in Table 2.

The main contributor to the composition effect is education. It accounts for between 73% and 95.5% of the effect in the NIDS, while it represents only 59.1% in the Census. Even if there is a significant difference between the NIDS and the Census, education is still the main contributor. Then, segregation cumulates between 3% and 17.5% of the composition effect. Except for the year 2008, segregation represents at least 9.8% of the composition effect. Finally, experience is much less important since it only accounts for between 1.5% and 9.5%. In the NIDS, the impact of experience is roughly twice smaller than the impact of segregation. In the Census, experience has an even smaller importance as it represents only 8.3% while segregation is responsible for 32.6% of the composition effect. These figures emphasize the detrimental effect of the dual school system in South Africa. However, segregation is also responsible of a significant share of half of the income gap. Differences in experience are less prominent and it is not clear what they represent exactly.

[33]We analyse jointly the experiences variables (experience and its square), and the education variables (years of schooling and its square).

[34]See Spaull[37] for more details about the persisting duality of the school system and the disadvantages faced by Blacks.

When we turn to the detailed decomposition of the structure effect, segregation always contributes positively to the gap while education tends to reduce it, except in 2010. Experience increases the gap, except in 2012, in the NIDS whereas it goes in the opposite direction in the Census. Moreover, in the NIDS, the effect of experience is small and almost always not significant, while in the Census, it is the second largest contributor and strongly significant. The statistical significance might be a matter of sample size, while the change in sign of the contribution might be a sample selection issue. If individuals in the NIDS are different than in the Census, then the correlation between experience and income might be different. As shown in Table 2, Blacks and Whites are older in the Census than in the NIDS and, despite that the age gap is pretty similar, there is much more age heterogeneity in the Census, typically due to measurement errors. The constant of the regression is always positive in the NIDS, except in 2010, but never significant. In the Census, it becomes significant and negative.

In terms of contributions, segregation is the first factor only in the pooled sample and the Census, accounting respectively for 41.7% and 57.9% of the structure effect. When we disaggregate by year in the NIDS, segregation is responsible for between 11.9% and 35%. Education is the second main factor contributing for between 20.7% and 43% in the NIDS waves. In the pooled sample, it reaches 30.9% of the structure effect, and drops to 10.6% only in the Census. Experience, as for the composition effect, seems to play a minor role in the structure effect as it accounts for between 2.7% and 7.7%. However in the Census, experience booms to 22.3% being more important than education. When we look at these figures, we should bear in mind several caveats. First, the constant represents a large amount of the structure effect ranging between 19.6% in the pooled sample and 55.3%. In the Census it is only equal to 9.2%. The second caveat is the level of significance of each contribution. All factors are highly significant in the Census, while in the NIDS, only segregation is significant most of the time. Experience and education are significant in the pooled sample and in 2008 respectively, but only at 10%. The constant is never significant. Finally, in 2012, there is not any statistically significant contribution, despite the total structure effect being significant at the 1% level.

By increasing the income gap, segregation has a different impact on Blacks and Whites. This might be due to three potential scenarios: either Whites benefit from segregation while Blacks suffer from segregation, Whites benefit more from segregation than Blacks, or Whites suffer less from segregation than Blacks. If we go back to the linear regressions in Table 3, we can find different patterns in the NIDS, for wave-specific regressions. In 2008, the coefficient for isolation is positive for Whites, while it is negative for Blacks, suggesting the first scenario. In 2010, this coefficient is still negative for Blacks, while it is not significant for Whites, suggesting the third scenario. This latter pattern occurs during the other waves. In the pooled sample, the first scenario is explicit. Finally, in the Census, we have the third scenario. But the coefficient for Whites is negative and significant whereas the coefficient was positive but not significant in the NIDS when this third scenario occurs. So, what are the lessons from these regressions? First, Blacks seems to always suffer from segregation and, in every case, they suffer more than Whites. Second, the effect for Whites is not clear. According to the Census, they also suffer from segregation, but less. This seems to be counterintuitive since a lot of Whites in South Africa still struggle to maintain residential segregation, so there should perceive some benefits doing it. But we could interpret that the other way around. We can find this intuitive if more integration would produce positive externalities on income. As they are less isolated on average than Blacks (although with a higher variance), and richer, if the most affluent are the most integrated, then there is a negative correlation between segregation and income. When using the NIDS, the effect for the Whites is either positive or statistically not different from zero. In the first case, it is in line with an explanation that Whites are richer and want to maintain residential segregation for protection, schools, and other motives. As such, Whites are stratified by income and by race; thus, the more segregated, the richer the individual. In the second case, it indicates that it has no effect on average contrary to Blacks. But it also indicates that there exists a lot of different situation with both poor and rich Whites being highly segregated. The mechanisms behind these two situations might be different and suggest to go deeper in the analysis.

7 Decomposition of other features of the income distribution

We go beyond the standard Oaxaca procedure decomposing the mean of the income distribution by using the (Recentered Influence Function) RIF-regressions developed by Firpo et al[18]. Rather than focusing on arbitrary quantiles of the distribution, we focus particularly on a set of quantiles that matters particularly for South Africa. As it is one of the most unequal country in the world, the top and bottom deciles are important to describe the very poor and the very rich. Moreover, we will also look at the median and the third quartile. We are also interested on quantiles describing poverty thresholds. The World Bank has estimated two international poverty lines at 1.90\$ and 3.10\$ a day.[35] In South Africa, slightly more than 15% were living with less than 1.90\$ a day in 2008, and around 16.5%

[35]These amounts are expressed in 2011\$ ppp.

in 2011.[36] More than twice these proportions were living with less than 3.10$ in both 2008 and 2011 (33.3% and 34.7% respectively). Finally, national poverty lines were estimated by the national South African statistical institute, Statistics South Africa.[37] An upper-bound poverty line[38] was specified, which represents the poverty headcount ratio of individuals able to consume a basket containing both essential food items and non-food consumptions. In 2009, 56.9% of the population were living with less than 577 ZAR of 2009. This figure decreases in 2011 to 45.5% of the population living with less than 620 ZAR of 2011. These last two measures complete our list of quantiles of interest.

Results of the RIF regressions at each quantile of interest are displayed in Table 17 and 18 for Whites and Africans respectively. We only report the coefficient of isolation for the sake of brevity, since the impact of education and experience throughout the income distribution is already widely studied. These regressions unveils an interesting pattern. To make the discussion suggestive, we report there in terms of potentially causal effects. However, it should must be kept in mind that the results are, *stricto sensu*, about descriptive additive decompositions of distributions. Segregation is harmless for Whites at the bottom of the distribution, while it has a positive effect at the top. For Blacks, the poorest suffers from segregation, while there is no gain at the top of the distribution. Therefore, racial segregation is linked to within-group economic inequality.

For Whites in Table 17, the first quantiles (up to the 45th) do not generally exhibit significant coefficients for isolation. This is true in 2010 (except for the 15th quantile), 2012, 2014 (up to the median), and when all years are pooled together. Above the 45th quantile, the impact of segregation on income becomes positive and significant, and is generally the strongest at the very top. In 2008, the effect is always positive and significant, and it is the strongest at the top of the distribution. An increase of one percentage point of the probability to interact with another White would be associated with an increase of 2.09% of the income of Whites in the last earning decile in 2008. Using the Census yields a distinct pattern: All Whites are weakly gaining from being segregated, except those in the last quartile.[39] This negative effect at the top of the distribution, although significant, is more than six times weaker than in the pooled data of the NIDS.

For Blacks in Table 18, the first quantiles (up to the 57th)have almost always

[36] The measures are poverty headcount ratio at the specified income levels.

[37] See the report published online by Statistics South Africa on the website of the South African government.

[38] Other poverty lines were defined with a smaller basket but are close to the quantiles found with international poverty lines.

[39] The estimates being similar from one quantile to the next one patrly comes from the construction of the income variable which is a categorical variable imperfectly transformed into a numeric one.

negative and significant coefficients of isolation. This is the case in 2008, 2010, 2012 (up to the 35th quantile), 2014 (up to the 46th quantile), and when data are pooled (up to the 56th quantile). In the 2011 Census,[40] the effect is always negative and significant up to the third quartile. The effect of segregation on income appears however positive and significant for the 90th quantile in the census while it is always unsignificant in the NIDS. There is also a general pattern in terms of magnitude. Segregation is generally harsher at the bottom quantiles and the effect fades away as we move toward the top of the distribution. In 2014, when the probability to interact with another Black increased by one percentage point, the associated decrease in income for Blacks earning less than 90% of their peers would have been of 1.69%. The same year, the decrease was only of 0.17% at the 56th quantile.

Several factors might explain these results. First, Blacks suffering from segregation at the bottom of the distribution is consistent with the fact that they live in the poorest areas with less access to basic public services. These areas are thus not attractive for richer individuals and they will likely move to more affluent neighbourhoods as their income increases enough. For instance, there are townships in which piped water is not available and inhabitants have to rely on boreholes, rain-water tanks, or water vendors among others. In this deprivation context, this negative correlation might also be explained by complementarity factors. For instance, Bramoullé and Kranton[9] show that risk sharing across communities can generate welfare gains for those that are directly or indirectly connected, while it can be detrimental for those that are not connected, but overall social welfare might be higher. As isolation is a measure of the probability to interact with someone of the same community, it can be viewed as a measure of connectedness of the two communities. Thus, in these deprived areas very vulnerable to shocks, reducing the mitigation of a random shock would probably have large consequences on the income levels of the remaining inhabitants. Statistics South Africa estimated that, due to the 2008-2009 financial crisis, they were more than 3 million people more living under the Food Poverty Line in 2009.[41] Similarly, the leavers will probably bring with them their economic and social contribution to the community. Hence, in presence of complementarity gains, someone living will reduce the benefits of all the remaining inhabitants. Moreover, it is even more likely that the richest inhabitants will move first from these poor areas, accentuating most probably the loss for the stayers.

[40]Note that too few discriminating observations hamper the estimation of the impact of segregation on income at the 10th quantile. Thus, the zero with no standard error is there to symbolize that estimation was unfeasible in that case.

[41]"The Food Poverty Line is the level of consumption below which individuals are unable to purchase sufficient food to provide them with an adequate diet". It was equivalent to 305 ZAR of 2009 and 321 ZAR of 2011 in respectively 2009 and 2011.

For Whites, the mechanisms at play seem different. Indeed, the positive correlation between segregation and incomes at the top of the distribution can be indicative of a selection mechanism, as previously for the Black townships. The richest individuals being Whites, they prefer to live in luxury neighbourhoods. Hence, the apparent benefit from segregation for Whites at the top of the distribution. But it might also be caused by the need for Whites to protect their economic and social positions with respect to Blacks. Long before the end of the Apartheid, Whites were concerned about keeping their high status in the South African society. When the political transition was seen as ineluctable, they struggled for not being relegated to second-class citizens. Still today, they are concerned about this "threat".[42] Then, being segregated might be a way to keep their higher status, as compared to other racial groups and especially Blacks. Thus, if being segregated was a mean to ensure that businesses were kept exclusively among members of the same group, then effectively increasing segregation may also increase income levels of this group. Bramoullé and Goyal[8] show that favoritism can be easily sustain in small homogeneous group when the different groups are all distinct from another.

Part of the effects in RIF regressions may be contaminated by selection issues for both Blacks and Whites. Nevertheless, we can check that these effects are not driven by a selection bias since, under strong ignorability, the composition effect of an Oaxaca-Blinder decomposition is a selection bias for the treatment effect literature.[43] Indeed, assuming this, when we turn to the decomposition of the income gap along the quantiles of interest, the composition effect coming from segregation is sometimes positive and significant, but always very small. Although the overall composition effect is positive and significant, it is much smaller than the structure effect and it is mainly generated by a differences in education levels. Thus, Whites are richer because they are better educated. However, more interestingly, the structure effect due to segregation is almost always the strongest positive and significant contribution to the income gap, especially for the lower tail of the distribution. For instance, in the pooled data (see Table 11), the structure effect due to segregation is 20 times larger than its composition counterpart, and more than 3 times the structure effect due to experience, at the first decile. All these elements together confirm the previous result that segregation has a different impact along the income distribution.

[42]For instance, a petition for asking to the European Council a right to return to Europe for White South Africans if things were to go wrong was started two years ago on change.org. It gathered more than 56 000 signatures the 17th of February, 2017.

[43]See Fortin et al.[19] for more details on the differences between the decomposition and the treatment effect literatures.

Table 5: Impact of segregation (isolation) on the White income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
2008	0.81	0.86**	0.79*	1.19***	0.68*	0.69*	0.78**	0.85**	0.76**	0.69*	0.68*	0.66*	1.10***	2.09***
	(1.24)	(2.00)	(1.87)	(2.71)	(1.87)	(1.91)	(2.16)	(2.36)	(2.11)	(1.90)	(1.85)	(1.79)	(2.96)	(3.87)
2010	0.01	-1.14*	-1.05	-0.68	0.14	0.31	0.29	1.27**	1.28**	1.15**	1.38***	1.44***	1.11***	-0.06
	(0.02)	(-1.68)	(-1.62)	(-1.03)	(0.25)	(0.57)	(0.54)	(2.54)	(2.57)	(2.40)	(3.05)	(3.22)	(2.67)	(-0.15)
2012	0.21	-0.05	-0.01	0.09	0.00	-0.23	-0.38	0.07	0.09	0.02	0.12	0.12	0.52	0.57
	(0.20)	(-0.08)	(-0.02)	(0.17)	(0.00)	(-0.45)	(-0.74)	(0.13)	(0.17)	(0.03)	(0.24)	(0.24)	(1.04)	(0.57)
2014	0.31	0.16	0.11	0.04	0.65	0.62	0.69	0.51	0.63	0.72*	1.06**	1.06**	1.26**	1.69**
	(0.29)	(0.21)	(0.16)	(0.05)	(1.49)	(1.45)	(1.61)	(1.24)	(1.55)	(1.67)	(2.36)	(2.36)	(2.14)	(2.31)
Pooled	0.55	0.43	0.50	0.47	0.33	0.34	0.36	0.60***	0.59***	0.65***	0.58***	0.61***	0.94***	1.25***
	(1.40)	(1.35)	(1.56)	(1.54)	(1.47)	(1.52)	(1.60)	(2.74)	(2.74)	(3.04)	(2.70)	(2.86)	(4.19)	(3.66)
Census	0.03***	0.20***	0.20***	0.20***	0.10***	0.10***	0.10***	0.02**	0.02**	0.02**	0.02**	0.02**	-0.07***	-0.19***
	(2.67)	(10.15)	(10.15)	(10.15)	(7.90)	(7.90)	(7.90)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(-6.53)	(-11.46)

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Impact of segregation (isolation) on the Black income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
2008	-0.72**	-0.77***	-0.74***	-0.80***	-0.63***	-0.63***	-0.70***	-0.77***	-0.70***	-0.75***	-0.49**	-0.52***	0.09	0.05
	(-2.15)	(-3.57)	(-3.49)	(-3.95)	(-3.32)	(-3.34)	(-3.70)	(-4.09)	(-3.67)	(-3.83)	(-2.47)	(-2.60)	(0.40)	(0.24)
2010	-1.62***	-1.22***	-1.22***	-1.35***	-0.91***	-0.80***	-0.79***	-0.50**	-0.54**	-0.73***	-0.76***	-0.71***	0.03	-0.10
	(-4.52)	(-3.97)	(-3.95)	(-4.80)	(-3.84)	(-3.41)	(-3.42)	(-2.28)	(-2.40)	(-3.18)	(-3.21)	(-2.97)	(0.12)	(-0.38)
2012	-0.87***	-0.59**	-0.59***	-0.41**	-0.80***	-0.73***	-0.65***	-0.24	-0.31	-0.02	0.27	0.27	0.09	-0.10
	(-3.68)	(-2.46)	(-2.86)	(-2.05)	(-3.83)	(-3.44)	(-3.04)	(-1.16)	(-1.50)	(-0.08)	(1.59)	(1.58)	(0.45)	(-0.42)
2014	-1.62***	-1.38***	-1.31***	-1.12***	-0.91***	-0.83***	-0.68***	-0.43***	-0.41***	-0.20	-0.11	-0.15	0.32	0.27
	(-6.34)	(-5.97)	(-6.13)	(-5.65)	(-5.45)	(-5.20)	(-4.18)	(-2.79)	(-2.67)	(-1.26)	(-0.72)	(-0.92)	(1.64)	(1.31)
Pooled	-1.20***	-0.90***	-1.04***	-0.93***	-0.74***	-0.73***	-0.72***	-0.52***	-0.47***	-0.41***	-0.17*	-0.15	0.15	0.02
	(-7.42)	(-7.57)	(-8.96)	(-8.13)	(-7.27)	(-7.21)	(-7.13)	(-5.46)	(-4.92)	(-4.38)	(-1.78)	(-1.62)	(1.43)	(0.17)
Census	0.00	-0.52***	-0.52***	-0.52***	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***	-0.17***	-0.42***	0.06***
	(.)	(-126.07)	(-126.07)	(-126.07)	(-127.64)	(-127.64)	(-127.64)	(-127.64)	(-127.64)	(-127.64)	(-127.64)	(-127.64)	(-122.65)	(11.61)

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Oaxaca decomposition of the 2008 income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
Differential														
Prediction_1	8.00***	8.31***	8.33***	8.36***	8.77***	8.79***	8.82***	9.02***	9.06***	9.16***	9.24***	9.25***	9.63***	10.19***
	(93.25)	(141.21)	(142.93)	(144.71)	(165.31)	(165.42)	(166.74)	(174.67)	(176.88)	(182.59)	(186.83)	(187.48)	(189.55)	(130.52)
Prediction_2	6.67***	6.91***	6.94***	6.97***	7.28***	7.29***	7.30***	7.49***	7.52***	7.60***	7.73***	7.76***	8.32***	8.91***
	(228.20)	(362.30)	(371.56)	(381.90)	(426.30)	(424.80)	(423.92)	(409.08)	(403.92)	(384.91)	(357.33)	(351.69)	(293.61)	(306.94)
Difference	1.33***	1.39***	1.39***	1.39***	1.49***	1.50***	1.52***	1.53***	1.54***	1.56***	1.51***	1.49***	1.31***	1.28***
	(14.66)	(22.50)	(22.69)	(22.97)	(26.71)	(26.81)	(27.24)	(27.91)	(28.31)	(28.92)	(27.95)	(27.61)	(22.55)	(15.35)
Explained														
Experience	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)	(-0.57)
Education	0.64***	0.64***	0.64***	0.64***	0.64***	0.64***	0.64***	0.64***	0.64***	0.64***	0.64***	0.64***	0.64***	0.64***
	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)	(20.70)
Isolation	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)	(0.46)
Total	0.65***	0.65***	0.65***	0.65***	0.65***	0.65***	0.65***	0.65***	0.65***	0.65***	0.65***	0.65***	0.65***	0.65***
	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)	(12.73)
Unexplained														
Experience	0.89***	0.91***	0.93***	0.92***	0.85***	0.84***	0.80***	0.48***	0.42***	0.23*	-0.03	-0.08	-0.66***	-0.59***
	(2.62)	(4.10)	(4.29)	(4.37)	(5.58)	(5.54)	(5.37)	(3.60)	(3.15)	(1.74)	(-0.23)	(-0.63)	(-4.80)	(-3.27)
Education	1.66	0.69	0.57	0.79	1.18**	1.02*	0.84*	-0.43	-0.60	-0.96	-1.49*	-1.56*	-2.67**	-5.32
	(0.69)	(0.44)	(0.38)	(0.50)	(2.09)	(1.93)	(1.70)	(-0.77)	(-1.01)	(-1.48)	(-1.89)	(-1.95)	(-2.10)	(-1.56)
Isolation	1.18**	1.25***	1.18***	1.49***	1.01***	1.01***	1.14***	1.25***	1.12***	1.12***	0.87***	0.89***	0.61*	1.29***
	(2.23)	(3.62)	(3.48)	(4.35)	(3.40)	(3.44)	(3.85)	(4.25)	(3.80)	(3.75)	(2.89)	(2.93)	(1.86)	(3.17)
Constant	-3.05	-2.10	-1.93	-2.46	-2.20***	-2.02***	-1.91***	-0.42	-0.04	0.52	1.51*	1.60*	3.39**	5.25
	(-1.22)	(-1.31)	(-1.25)	(-1.52)	(-3.37)	(-3.27)	(-3.25)	(-0.66)	(-0.06)	(0.72)	(1.76)	(1.85)	(2.57)	(1.53)
Total	0.68***	0.74***	0.74***	0.74***	0.84***	0.85***	0.87***	0.88***	0.90***	0.91***	0.86***	0.85***	0.67***	0.63***
	(7.00)	(10.26)	(10.34)	(10.48)	(12.95)	(13.05)	(13.37)	(13.88)	(14.13)	(14.46)	(13.50)	(13.25)	(9.95)	(7.10)
Observations	3362	3362	3362	3362	3362	3362	3362	3362	3362	3362	3362	3362	3362	3362

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Oaxaca decomposition of the 2010 income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
Differential														
Prediction_1	8.21***	8.35***	8.40***	8.42***	8.94***	8.96***	8.98***	9.17***	9.19***	9.33***	9.43***	9.44***	9.85***	9.99***
	(78.76)	(79.78)	(83.04)	(83.63)	(125.39)	(126.33)	(128.17)	(135.67)	(136.63)	(142.24)	(147.01)	(147.46)	(164.81)	(201.31)
Prediction_2	6.65***	6.92***	6.95***	6.97***	7.41***	7.45***	7.47***	7.68***	7.72***	7.78***	7.92***	7.97***	8.50***	9.07***
	(193.80)	(263.22)	(261.66)	(279.70)	(378.30)	(386.30)	(387.13)	(384.19)	(380.97)	(372.64)	(350.79)	(344.11)	(329.47)	(331.72)
Difference	1.56***	1.42***	1.46***	1.45***	1.53***	1.50***	1.51***	1.49***	1.47***	1.54***	1.52***	1.47***	1.35***	0.92***
	(14.24)	(13.19)	(13.92)	(13.98)	(20.70)	(20.43)	(20.76)	(21.11)	(20.87)	(22.44)	(22.31)	(21.64)	(20.79)	(16.24)
Explained														
Experience	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***
	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)	(3.67)
Education	0.57***	0.57***	0.57***	0.57***	0.57***	0.57***	0.57***	0.57***	0.57***	0.57***	0.57***	0.57***	0.57***	0.57***
	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)	(14.77)
Isolation	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***	0.13***
	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)	(2.66)
Total	0.76***	0.76***	0.76***	0.76***	0.76***	0.76***	0.76***	0.76***	0.76***	0.76***	0.76***	0.76***	0.76***	0.76***
	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)	(11.44)
Unexplained														
Experience	0.34	1.16***	1.10***	1.23***	0.82***	0.74***	0.68***	0.61***	0.60***	0.45***	0.18	0.16	-0.45***	-0.78***
	(0.78)	(2.76)	(2.74)	(3.11)	(4.07)	(3.67)	(3.42)	(3.56)	(3.47)	(2.85)	(1.19)	(1.06)	(-3.21)	(-6.05)
Education	6.65*	7.52**	7.17**	7.79**	3.72***	3.48***	3.00**	0.44	0.39	-0.21	-1.68	-1.94*	-4.46***	-3.20***
	(1.79)	(2.12)	(2.11)	(2.27)	(3.00)	(2.84)	(2.52)	(0.41)	(0.36)	(-0.21)	(-1.62)	(-1.86)	(-3.64)	(-2.96)
Isolation	1.38**	0.26	0.32	0.69	0.80*	0.81*	0.79*	1.16***	1.19***	1.29***	1.47***	1.47***	0.56	-0.08
	(2.47)	(0.50)	(0.63)	(1.37)	(1.92)	(1.94)	(1.91)	(2.99)	(3.09)	(3.39)	(3.96)	(3.96)	(1.61)	(-0.22)
Constant	-7.58**	-8.29**	-7.90**	-9.02***	-4.58***	-4.29***	-3.72***	-1.49	-1.48	-0.75	0.78	1.02	4.94***	4.22***
	(-1.99)	(-2.31)	(-2.30)	(-2.62)	(-3.55)	(-3.38)	(-3.01)	(-1.32)	(-1.32)	(-0.69)	(0.70)	(0.91)	(3.72)	(3.52)
Total	0.80***	0.66***	0.69***	0.69***	0.77***	0.74***	0.74***	0.72***	0.70***	0.78***	0.75***	0.71***	0.59***	0.16*
	(6.79)	(5.77)	(6.21)	(6.25)	(8.98)	(8.66)	(8.79)	(8.72)	(8.49)	(9.58)	(9.27)	(8.73)	(7.20)	(1.96)
Observations	3331	3331	3331	3331	3331	3331	3331	3331	3331	3331	3331	3331	3331	3331

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Oaxaca decomposition of the 2012 income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
Differential														
Prediction_1	8.14***	8.45***	8.47***	8.53***	8.84***	8.84***	8.85***	9.16***	9.17***	9.23***	9.34***	9.36***	9.73***	10.35***
	(72.57)	(110.49)	(113.45)	(124.87)	(136.18)	(135.37)	(134.65)	(142.39)	(142.57)	(145.28)	(151.34)	(151.66)	(154.97)	(85.72)
Prediction_2	6.94***	7.06***	7.14***	7.21***	7.51***	7.54***	7.59***	7.81***	7.83***	7.92***	8.07***	8.10***	8.54***	9.17***
	(336.80)	(373.84)	(431.88)	(470.31)	(431.50)	(425.97)	(423.45)	(436.66)	(437.41)	(444.80)	(464.99)	(466.21)	(372.89)	(367.00)
Difference	1.20***	1.40***	1.33***	1.33***	1.33***	1.30***	1.25***	1.35***	1.34***	1.30***	1.27***	1.25***	1.19***	1.17***
	(10.53)	(17.76)	(17.38)	(18.96)	(19.75)	(19.24)	(18.42)	(20.20)	(20.07)	(19.74)	(19.75)	(19.55)	(17.77)	(9.51)
Explained														
Experience	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***
	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)	(2.73)
Education	0.51***	0.51***	0.51***	0.51***	0.51***	0.51***	0.51***	0.51***	0.51***	0.51***	0.51***	0.51***	0.51***	0.51***
	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)	(15.67)
Isolation	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)	(1.48)
Total	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***	0.61***
	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)	(11.38)
Unexplained														
Experience	-0.41	-0.06	0.10	0.08	0.61***	0.56***	0.50***	0.27*	0.22	0.31**	0.11	0.09	-0.50***	-0.68***
	(-1.12)	(-0.23)	(0.41)	(0.35)	(3.54)	(3.23)	(2.89)	(1.68)	(1.35)	(2.07)	(0.78)	(0.63)	(-3.57)	(-2.73)
Education	9.13**	6.50**	5.70**	4.38**	0.85	0.27	-0.27	-0.53	-0.90	-0.93	-1.58	-1.59	-4.01***	-12.84***
	(1.99)	(2.57)	(2.39)	(2.06)	(0.52)	(0.17)	(-0.16)	(-0.38)	(-0.65)	(-0.67)	(-1.19)	(-1.19)	(-2.80)	(-3.73)
Isolation	0.89	0.45	0.48	0.38	0.68*	0.47	0.30	0.21	0.29	-0.03	-0.23	-0.23	0.20	0.40
	(1.25)	(0.95)	(1.08)	(0.93)	(1.81)	(1.22)	(0.77)	(0.54)	(0.74)	(-0.09)	(-0.63)	(-0.62)	(0.53)	(0.59)
Constant	-9.02*	-6.11**	-5.57**	-4.13*	-1.43	-0.61	0.10	0.79	1.12	1.34	2.35*	2.37*	4.89***	13.68***
	(-1.89)	(-2.33)	(-2.27)	(-1.88)	(-0.85)	(-0.36)	(0.06)	(0.53)	(0.76)	(0.90)	(1.66)	(1.67)	(3.19)	(3.82)
Total	0.59***	0.79***	0.72***	0.71***	0.71***	0.69***	0.64***	0.74***	0.73***	0.69***	0.65***	0.64***	0.57***	0.56***
	(4.87)	(8.86)	(8.27)	(8.80)	(9.68)	(9.25)	(8.55)	(10.11)	(9.98)	(9.56)	(9.10)	(8.92)	(7.63)	(4.50)
Observations	4257	4257	4257	4257	4257	4257	4257	4257	4257	4257	4257	4257	4257	4257

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Oaxaca decomposition of the 2014 income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
Differential														
Prediction_1	8.09***	8.45***	8.50***	8.55***	8.93***	8.96***	8.98***	9.14***	9.17***	9.25***	9.32***	9.34***	9.77***	10.24***
	(54.46)	(83.28)	(89.54)	(95.35)	(140.32)	(142.87)	(144.24)	(150.31)	(150.79)	(149.70)	(148.12)	(148.43)	(126.46)	(105.05)
Prediction_2	6.90***	7.14***	7.19***	7.25***	7.61***	7.67***	7.70***	7.86***	7.88***	7.98***	8.04***	8.08***	8.62***	9.22***
	(295.77)	(341.02)	(366.81)	(405.27)	(558.75)	(586.65)	(598.91)	(624.77)	(624.90)	(608.66)	(580.57)	(561.83)	(407.88)	(453.22)
Difference	1.18***	1.31***	1.31***	1.30***	1.31***	1.29***	1.28***	1.28***	1.29***	1.27***	1.28***	1.26***	1.15***	1.02***
	(7.87)	(12.65)	(13.49)	(14.21)	(20.21)	(20.17)	(20.14)	(20.66)	(20.72)	(20.16)	(19.88)	(19.56)	(14.40)	(10.22)
Explained														
Experience	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***
	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)
Education	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***
	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)	(14.82)
Isolation	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***
	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)	(2.64)
Total	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***
	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)	(11.69)
Unexplained														
Experience	0.25	0.29	0.24	0.16	0.49***	0.47***	0.46***	0.23	0.20	0.25*	0.17	0.14	-0.08	-0.30
	(0.56)	(1.01)	(0.91)	(0.62)	(3.00)	(2.91)	(2.84)	(1.61)	(1.41)	(1.71)	(1.19)	(0.97)	(-0.51)	(-1.58)
Education	3.21	2.18	2.28	1.79	-3.00	-3.23	-3.25	-2.59	-2.66	-0.88	-0.21	-0.24	-2.67	-7.43**
	(0.56)	(0.51)	(0.56)	(0.48)	(-1.30)	(-1.44)	(-1.51)	(-1.16)	(-1.19)	(-0.66)	(-0.13)	(-0.15)	(-1.41)	(-2.41)
Isolation	1.60**	1.28**	1.19**	0.96**	1.15***	1.06***	0.97***	0.62**	0.68**	0.54*	0.68**	0.71**	0.40	0.72
	(2.23)	(2.41)	(2.43)	(2.11)	(3.60)	(3.39)	(3.07)	(2.05)	(2.27)	(1.71)	(2.09)	(2.18)	(0.96)	(1.43)
Constant	-4.50	-3.07	-3.03	-2.24	2.04	2.36	2.48	2.40	2.44	0.74	0.01	0.02	2.88	7.40**
	(-0.76)	(-0.71)	(-0.74)	(-0.59)	(0.88)	(1.04)	(1.13)	(1.04)	(1.06)	(0.53)	(0.01)	(0.01)	(1.40)	(2.33)
Total	0.56***	0.68***	0.68***	0.67***	0.69***	0.67***	0.65***	0.66***	0.66***	0.65***	0.65***	0.64***	0.53***	0.39***
	(3.65)	(6.35)	(6.69)	(6.90)	(9.40)	(9.18)	(8.97)	(9.43)	(9.49)	(9.18)	(9.10)	(8.83)	(6.25)	(3.81)
Observations	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Oaxaca decomposition of the Pooled income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
Differential														
Prediction_1	8.12***	8.35***	8.38***	8.42***	8.84***	8.87***	8.88***	9.11***	9.15***	9.19***	9.32***	9.33***	9.71***	10.17***
	(157.15)	(201.52)	(204.07)	(213.91)	(285.21)	(288.11)	(287.85)	(309.18)	(314.00)	(316.90)	(325.26)	(325.81)	(326.33)	(224.07)
Prediction_2	6.80***	7.03***	7.05***	7.08***	7.48***	7.51***	7.54***	7.74***	7.78***	7.83***	7.99***	8.01***	8.52***	9.12***
	(491.60)	(700.80)	(703.51)	(722.99)	(870.34)	(870.06)	(872.11)	(903.94)	(907.93)	(904.93)	(875.10)	(871.82)	(718.12)	(723.09)
Difference	1.32***	1.32***	1.33***	1.34***	1.36***	1.37***	1.35***	1.37***	1.38***	1.35***	1.33***	1.33***	1.19***	1.06***
	(24.60)	(31.02)	(31.47)	(33.04)	(42.34)	(42.77)	(42.06)	(44.71)	(45.36)	(44.77)	(44.32)	(44.05)	(37.09)	(22.39)
Explained														
Experience	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***	0.04***
	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)	(5.37)
Education	0.52***	0.52***	0.52***	0.52***	0.52***	0.52***	0.52***	0.52***	0.52***	0.52***	0.52***	0.52***	0.52***	0.52***
	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)	(32.22)
Isolation	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***	0.07***
	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)	(3.46)
Total	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***	0.63***
	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)	(22.83)
Unexplained														
Experience	0.34*	0.55***	0.58***	0.59***	0.69***	0.65***	0.62***	0.36***	0.33***	0.28***	0.17***	0.14**	-0.41***	-0.46***
	(1.80)	(3.72)	(4.01)	(4.32)	(8.11)	(7.77)	(7.45)	(4.88)	(4.53)	(3.88)	(2.58)	(2.04)	(-5.84)	(-4.74)
Education	3.50	3.12	2.77	2.71	0.91	0.77	0.71	-0.48	-0.56	-0.87*	-0.97**	-1.11**	-3.14***	-5.67***
	(1.46)	(1.49)	(1.40)	(1.39)	(1.64)	(1.43)	(1.32)	(-1.02)	(-1.18)	(-1.72)	(-2.20)	(-2.40)	(-3.19)	(-2.65)
Isolation	1.40***	1.04***	1.21***	1.10***	0.83***	0.83***	0.83***	0.80***	0.75***	0.73***	0.46***	0.47***	0.40**	0.72***
	(4.71)	(4.41)	(5.22)	(4.90)	(4.73)	(4.74)	(4.75)	(4.76)	(4.49)	(4.43)	(2.78)	(2.83)	(2.28)	(2.90)
Constant	-4.55*	-4.02*	-3.86*	-3.69*	-1.70***	-1.51***	-1.44**	0.07	0.23	0.58	1.04**	1.20**	3.71***	5.84***
	(-1.86)	(-1.90)	(-1.93)	(-1.88)	(-2.87)	(-2.61)	(-2.50)	(0.13)	(0.45)	(1.07)	(2.17)	(2.40)	(3.67)	(2.70)
Total	0.69***	0.69***	0.70***	0.71***	0.73***	0.74***	0.72***	0.74***	0.75***	0.73***	0.70***	0.70***	0.56***	0.43***
	(12.16)	(14.97)	(15.24)	(16.03)	(20.10)	(20.34)	(19.83)	(21.23)	(21.56)	(20.81)	(20.33)	(20.08)	(15.19)	(8.55)
Observations	16470	16470	16470	16470	16470	16470	16470	16470	16470	16470	16470	16470	16470	16470

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Oaxaca decomposition of the Census income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
Differential														
Prediction_1	7.26***	7.87***	7.89***	7.92***	8.64***	8.65***	8.67***	9.25***	9.26***	9.29***	9.34***	9.35***	10.00***	10.70***
	(7611.38)	(4766.68)	(4780.84)	(4795.00)	(7634.20)	(7645.44)	(7656.68)	(10023.79)	(10033.18)	(10070.67)	(10126.93)	(10136.30)	(10280.71)	(7317.82)
Prediction_2	5.36	6.41***	6.42***	6.43***	7.11***	7.11***	7.12***	7.14***	7.14***	7.15***	7.16***	7.16***	7.85***	9.20***
	(.)	(25547.16)	(25578.29)	(25609.42)	(91209.94)	(91235.37)	(91260.81)	(91515.13)	(91540.56)	(91642.07)	(91794.88)	(91820.32)	(41001.18)	(29518.13)
Difference	1.90***	1.46***	1.47***	1.49***	1.53***	1.54***	1.55***	2.11***	2.12***	2.15***	2.19***	2.19***	2.15***	1.49***
	(1995.01)	(873.31)	(882.63)	(891.95)	(1348.95)	(1358.42)	(1367.88)	(2282.72)	(2289.93)	(2318.73)	(2361.93)	(2369.13)	(2170.53)	(999.64)
Explained														
Experience	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***	0.11***
	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)	(120.03)
Education	0.78***	0.78***	0.78***	0.78***	0.78***	0.78***	0.78***	0.78***	0.78***	0.78***	0.78***	0.78***	0.78***	0.78***
	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)	(353.33)
Isolation	0.43***	0.43***	0.43***	0.43***	0.43***	0.43***	0.43***	0.43***	0.43***	0.43***	0.43***	0.43***	0.43***	0.43***
	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)	(99.23)
Total	1.32***	1.32***	1.32***	1.32***	1.32***	1.32***	1.32***	1.32***	1.32***	1.32***	1.32***	1.32***	1.32***	1.32***
	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)	(276.96)
Unexplained														
Experience	-0.06***	-0.34***	-0.34***	-0.34***	-0.10***	-0.10***	-0.10***	-0.06***	-0.06***	-0.06***	-0.06***	-0.06***	-0.03***	-0.01***
	(-16.96)	(-71.18)	(-71.18)	(-71.18)	(-31.76)	(-31.76)	(-31.76)	(-24.45)	(-24.45)	(-24.45)	(-24.45)	(-24.45)	(-11.84)	(-2.99)
Education	-0.58***	0.00	0.00	0.00	-0.24***	-0.24***	-0.24***	-0.49***	-0.49***	-0.49***	-0.49***	-0.49***	-0.72***	-0.82***
	(-35.57)	(0.15)	(0.15)	(0.15)	(-15.63)	(-15.63)	(-15.63)	(-47.34)	(-47.34)	(-47.34)	(-47.34)	(-47.34)	(-81.67)	(-74.45)
Isolation	-0.41***	0.19***	0.19***	0.19***	-0.20***	-0.20***	-0.20***	-0.25***	-0.25***	-0.25***	-0.25***	-0.25***	-0.09***	-0.60***
	(-45.19)	(13.50)	(13.50)	(13.50)	(-20.39)	(-20.39)	(-20.39)	(-30.97)	(-30.97)	(-30.97)	(-30.97)	(-30.97)	(-9.63)	(-48.69)
Constant	1.63***	0.27***	0.29***	0.31***	0.75***	0.76***	0.78***	1.60***	1.60***	1.63***	1.67***	1.68***	1.66***	1.61***
	(87.37)	(9.01)	(9.52)	(10.03)	(42.11)	(42.71)	(43.31)	(128.06)	(128.60)	(130.74)	(133.95)	(134.49)	(141.02)	(98.69)
Total	0.58***	0.14***	0.15***	0.17***	0.21***	0.22***	0.23***	0.79***	0.80***	0.83***	0.87***	0.87***	0.83***	0.17***
	(120.06)	(27.61)	(30.76)	(33.91)	(43.53)	(45.77)	(48.01)	(166.79)	(168.19)	(173.81)	(182.23)	(183.64)	(175.30)	(35.68)
Observations	1334468	1334468	1334468	1334468	1334468	1334468	1334468	1334468	1334468	1334468	1334468	1334468	1334468	1334468

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

8 Addressing endogeneity concerns

Endogenous sorting

We might be concerned by the potential contamination of our estimations by an endogeneity problem. Individuals with similar unobservables might chose to live in similar places generating a correlation between segregation and unobservables. This endogenous sorting problem is present in almost all the studies about neighborhood effects. Cutler et al.[15] discuss this issue at length. Usually, individuals are believed to sort endogenously within a city but not across city. Hence, Cutler et al.[15] propose to instrument the neighborhood measure of segregation by a metropolitan/city measure. In our case, this problem is likely absent from our analysis since our measure of segregation is already at the District Council level which is the second most aggregated level below the provinces and above the municipalities. However, we still instrument the segregation variable by its value in 1996 to strengthen our OLS esstimations.

We also provide an instrumentation strategy for the education variable since it is common in mincerian regressions to have endogeneity issue. Skills and talents are unobserved and believed to be correlated to the human capital accumulation process. The standard approach proposed by Angrist and Krueger[**?**] suggests to instrument education levels by their quarter of birth due to the discontinuity generated by age limits at a the beginning of a new school year. However, in our case, this kind of instruments is weak. An alternative would be to exploit variations of the institutions as instrument (Angrist and Imbens[**?**]). However, our test with a dummy for being born before the end of the Apartheid proved to be weak also. We finally turned to the average education level in the cohort of birth. If particular skills and talents are responsible for the level of education of certain individuals, it is likely to be rare events. First, not all talented individuals are converting this advantage into a better outcome since human capital accumulation is a long process likely to be perturbed at some point. Then, the impact of these few talented people is diluted in the mean when pooled with untalented people. On the other side, individuals of the same birth cohort face the same context and are very likely to have a similar level of education. This instrument performs better than our two previous tentatives.

We also compare the IV estimates with the OLS estimates using the Hausman test. This reveals that for Blacks, there are no significant differences between the two types of estimates. With respect of the segregation measure, this is reassuring and confirms that endogenous sorting is already treated by our aggregated measure. For education, the most likely explanation of this phenomenon is that Blacks were constrained in their education levels during the Apartheid, and still today with the duality of the South African school system. Then skills and talents are not likely to play any role in their education choices. For Whites, there are significant differences between the IV and the OLS. Since segregation is exogenous for Blacks, we believe that it is also exogenous to Whites since the endogeneity issue relates to endogenous sorting at the local level which should affect segregation the same way as for Blacks. Thus, education is endogenous for Whites. Related to the explanation for Blacks, Whites were, on the contrary, never constrained in their education choices. Thus, skills and talents might well have influenced their education levels. We report the OLS and IV estimates for the Blacks, Whites, and pooled samples in 2014 in Table 13. The Oaxaca decomposition of the mean for the same year is reported in Table 14. Conclusions are unchanged.

Table 13: OLS and IV regressions (2014)

	Black (OLS)	Black (IV)	White (OLS)	White (IV)	Pooled (OLS)	Pooled (IV)
Experience	0.02***	0.02***	0.03**	0.03**	0.02***	0.03***
	(7.46)	(6.18)	(2.47)	(2.52)	(7.76)	(6.47)
Experience (square)	-0.00***	-0.00	-0.00	-0.00	-0.00***	-0.00**
	(-2.90)	(-1.21)	(-1.26)	(-1.31)	(-2.68)	(-2.09)
Years of schooling	-0.06***	-0.01	-0.31	0.49	-0.06***	-0.09*
	(-5.04)	(-0.35)	(-1.13)	(0.19)	(-5.06)	(-1.81)
Years of schooling (square)	0.01***	0.01***	0.02*	-0.01	0.01***	0.01***
	(17.32)	(6.86)	(1.87)	(-0.11)	(18.10)	(5.19)
Isolation	-0.46***	-0.46***	0.55	1.07*	-1.03***	-0.97***
	(-3.22)	(-3.17)	(1.37)	(1.96)	(-9.29)	(-8.57)
Constant	7.29***	7.05***	8.83***	3.48	7.77***	7.83***
	(47.91)	(38.74)	(5.00)	(0.22)	(60.64)	(30.33)
Observations	5291	5291	229	229	5520	5520
R^2	0.259	0.256	0.279	0.248	0.296	0.295

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Oaxaca decomposition with endogeneity (2014)

	Income
Differential	
Prediction_1	9.20***
	(157.93)
Prediction_2	8.01***
	(626.57)
Difference	1.20***
	(20.07)
Explained	
Education	0.48***
	(14.53)
Isolation	0.28***
	(8.33)
Experience	0.06***
	(3.87)
Total	0.83***
	(16.81)
Unexplained	
Education	3.23
	(0.20)
Isolation	0.83**
	(2.22)
Experience	0.11
	(0.80)
Constant	-3.80
	(-0.24)
Total	0.37***
	(5.85)
Observations	5520

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

City specialization

Another potential threat to our estimation is that segregation at the district council level might reflect specialization and/or special characteristics of the area which have nothing to do with segregation. To adress this concern, we introduce in our quantile regressions metropolitan fixed effects in the form of dummy variables for the six main South African metropolitan areas (Cape Town, Port-Elizabeth, Durban, Johannesburg, Tshwane (previously named Pretoria), and Eastrand). These metropolitan areas are the main economic centers of South Africa and constitute a district council on their own.

However, as segregation is measured at the same level as the metropolitan dummies, there is a potential concern with multicollinearity. This is the reason why we limit these metropolitan fixed effects to only six. As we perform regression by group, the segregation measure is almost a linear combination of the metropolitan dummies. In the whole sample, the collinearity would be perfect if we either restrict our sample to a particular metropolitan area or add a dummy for every district council. As we remove more and more dummies, the collinearity problem weakens. But still adding this six metropolitan dummies dilutes the effect of segregation. In decompositions, this effect implies that the structure effect of segregation is no longer significant in the majority of the cases. To illustrate this effect, we provide estimates of the contribution of segregation to income for the year 2014, first with the six metropolitan dummies (Table 15), then with only two metropolitan dummies (Cape Town and Johannesburg) in Table 16.

Table 15: Oaxaca decomposition of the 2014 income distribution (quantile reference)

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
Differential														
Prediction_1	8.09***	8.45***	8.50***	8.55***	8.93***	8.96***	8.98***	9.14***	9.17***	9.25***	9.32***	9.34***	9.77***	10.24***
	(53.82)	(82.31)	(88.48)	(94.22)	(138.85)	(141.37)	(142.70)	(148.84)	(149.33)	(148.23)	(146.62)	(146.92)	(125.14)	(103.92)
Prediction_2	6.90***	7.14***	7.19***	7.25***	7.61***	7.67***	7.70***	7.86***	7.88***	7.98***	8.04***	8.08***	8.62***	9.22***
	(295.61)	(340.85)	(366.63)	(405.06)	(558.48)	(586.37)	(598.63)	(624.49)	(624.61)	(608.38)	(580.32)	(561.58)	(407.71)	(453.02)
Difference	1.18***	1.31***	1.31***	1.30***	1.31***	1.29***	1.28***	1.28***	1.29***	1.27***	1.28***	1.26***	1.15***	1.02***
	(7.78)	(12.51)	(13.34)	(14.04)	(20.00)	(19.96)	(19.94)	(20.47)	(20.53)	(19.97)	(19.69)	(19.37)	(14.26)	(10.12)
Explained														
Experience	-0.01	0.01	0.01	0.01*	0.03***	0.03***	0.03***	0.05***	0.05***	0.06***	0.08***	0.09***	0.12***	0.10***
	(-1.18)	(0.77)	(0.98)	(1.79)	(3.61)	(3.70)	(3.74)	(4.09)	(4.08)	(4.13)	(4.22)	(4.23)	(4.23)	(4.02)
Education	0.29***	0.34***	0.30***	0.31***	0.32***	0.33***	0.33***	0.38***	0.38***	0.41***	0.50***	0.52***	0.75***	0.62***
	(10.90)	(13.10)	(13.06)	(13.47)	(14.90)	(14.97)	(14.99)	(15.10)	(15.05)	(14.97)	(15.09)	(15.14)	(14.40)	(12.58)
Isolation	0.36***	0.29***	0.26***	0.23***	0.21***	0.19***	0.20***	0.16***	0.17***	0.14***	0.18***	0.18***	0.35***	0.34***
	(7.42)	(6.88)	(6.67)	(5.90)	(6.51)	(6.08)	(6.25)	(5.08)	(5.29)	(4.18)	(4.86)	(4.67)	(6.33)	(4.84)
tshwane	0.02***	0.02**	0.02**	0.02**	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02**	0.02**
	(2.62)	(2.54)	(2.49)	(2.56)	(2.80)	(2.79)	(2.81)	(2.75)	(2.75)	(2.72)	(2.69)	(2.63)	(2.09)	(2.10)
cape	0.01*	0.01*	0.01*	0.01*	0.01*	0.01*	0.01	0.00	0.00	0.00	0.00	0.00	-0.01*	-0.01
	(1.95)	(1.78)	(1.90)	(1.90)	(1.80)	(1.83)	(1.64)	(0.94)	(0.92)	(1.22)	(0.76)	(0.37)	(-1.73)	(-1.21)
portelizabeth	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(-0.22)	(-0.36)	(-0.36)	(-0.36)	(-0.35)	(-0.35)	(-0.35)	(-0.35)	(-0.35)	(-0.36)	(-0.36)	(-0.36)	(-0.36)	(-0.36)
durban	-0.01**	-0.01**	-0.01**	-0.00*	-0.00*	-0.00*	-0.00	-0.00	0.00	0.00	0.00	0.00	0.01**	0.01**
	(-2.03)	(-2.20)	(-1.99)	(-1.73)	(-1.67)	(-1.65)	(-1.20)	(-0.26)	(0.05)	(1.19)	(1.31)	(1.46)	(2.31)	(2.33)
eastrand	0.01	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01
	(1.30)	(1.03)	(0.81)	(0.83)	(1.49)	(1.51)	(1.50)	(1.51)	(1.46)	(1.48)	(1.45)	(1.39)	(1.03)	(1.22)
johannesburg	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00
	(-1.12)	(-1.13)	(-1.11)	(-1.10)	(-1.16)	(-1.17)	(-1.16)	(-1.17)	(-1.17)	(-1.16)	(-1.13)	(-1.12)	(0.69)	(1.03)
Total	0.66***	0.65***	0.58***	0.57***	0.58***	0.58***	0.59***	0.61***	0.62***	0.63***	0.79***	0.81***	1.24***	1.09***
	(11.94)	(13.21)	(13.11)	(12.74)	(15.35)	(15.11)	(15.38)	(15.01)	(15.12)	(14.33)	(15.12)	(14.99)	(15.39)	(11.88)
Unexplained														
Experience	0.48	0.43	0.33	0.24	0.53***	0.51***	0.50***	0.24*	0.21	0.23	0.14	0.11	-0.15	-0.36*
	(1.11)	(1.45)	(1.20)	(0.92)	(3.24)	(3.15)	(3.07)	(1.73)	(1.52)	(1.58)	(0.98)	(0.72)	(-0.91)	(-1.82)
Education	4.90	3.26	2.89	2.34	-2.60	-2.84	-2.89	-2.42	-2.47	-0.92	-0.29	-0.34	-3.38*	-8.00***
	(0.89)	(0.81)	(0.74)	(0.64)	(-1.12)	(-1.25)	(-1.31)	(-1.10)	(-1.12)	(-0.62)	(-0.17)	(-0.20)	(-1.69)	(-2.62)
Isolation	0.94	1.12	0.90	0.59	0.69	0.53	0.32	0.20	0.30	0.12	0.18	0.21	0.08	0.64
	(0.95)	(1.53)	(1.33)	(0.92)	(1.51)	(1.17)	(0.70)	(0.48)	(0.75)	(0.30)	(0.41)	(0.50)	(0.14)	(0.89)
tshwane	-0.00	-0.03	-0.02	-0.01	-0.02	-0.02	-0.01	-0.01	-0.01	-0.02	-0.01	-0.01	-0.00	-0.00
	(-0.06)	(-0.77)	(-0.51)	(-0.28)	(-0.94)	(-0.69)	(-0.41)	(-0.53)	(-0.68)	(-0.90)	(-0.60)	(-0.60)	(-0.07)	(-0.03)
cape	-0.00	-0.02	-0.02	-0.01	-0.00	0.00	0.01	-0.01	-0.02	-0.01	-0.00	-0.00	0.03	0.01
	(-0.05)	(-0.76)	(-0.63)	(-0.47)	(-0.09)	(0.14)	(0.48)	(-0.78)	(-0.87)	(-0.61)	(-0.17)	(-0.07)	(1.03)	(0.30)
portelizabeth	-0.05	-0.01	-0.01	-0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01	-0.00	0.01
	(-1.14)	(-0.64)	(-0.55)	(-0.46)	(0.02)	(0.14)	(0.25)	(0.58)	(0.54)	(1.02)	(1.08)	(1.11)	(-0.14)	(0.31)
durban	0.01	0.01	0.00	0.01	0.00	0.00	0.01	-0.01	-0.01*	-0.01	-0.01	-0.00	0.02	0.03
	(1.03)	(0.60)	(0.50)	(0.70)	(0.46)	(0.53)	(0.64)	(-1.58)	(-1.69)	(-1.26)	(-0.68)	(-0.51)	(1.31)	(1.21)
eastrand	0.04	-0.00	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.02	0.01	0.01	-0.00	-0.05*
	(0.87)	(-0.01)	(0.22)	(0.42)	(0.26)	(0.46)	(0.60)	(0.71)	(0.71)	(1.10)	(0.52)	(0.54)	(-0.20)	(-1.89)
johannesburg	0.06**	0.03	0.01	0.01	0.00	0.00	0.00	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01	-0.01
	(2.24)	(1.43)	(0.48)	(0.54)	(0.03)	(0.13)	(0.10)	(-0.45)	(-0.44)	(-1.31)	(-0.69)	(-0.74)	(-0.95)	(-0.57)
Constant	-5.86	-4.11	-3.38	-2.44	2.13	2.52	2.74	2.68	2.65	1.24	0.47	0.48	3.34	7.66**
	(-1.03)	(-1.00)	(-0.85)	(-0.66)	(0.89)	(1.08)	(1.22)	(1.16)	(1.15)	(0.78)	(0.27)	(0.27)	(1.51)	(2.38)

Table 16: Oaxaca decomposition of the 2014 income distribution (quantile reference)

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
Differential														
Prediction_1	8.09***	8.45***	8.50***	8.55***	8.93***	8.96***	8.98***	9.14***	9.17***	9.25***	9.32***	9.34***	9.77***	10.24***
	(54.24)	(82.96)	(89.19)	(94.97)	(139.83)	(142.36)	(143.72)	(149.81)	(150.29)	(149.20)	(147.62)	(147.92)	(126.01)	(104.66)
Prediction_2	6.90***	7.14***	7.19***	7.25***	7.61***	7.67***	7.70***	7.86***	7.88***	7.98***	8.04***	8.08***	8.62***	9.22***
	(295.72)	(340.96)	(366.75)	(405.20)	(558.66)	(586.56)	(598.82)	(624.68)	(624.81)	(608.56)	(580.49)	(561.74)	(407.82)	(453.15)
Difference	1.18***	1.31***	1.31***	1.30***	1.31***	1.29***	1.28***	1.28***	1.29***	1.27***	1.28***	1.26***	1.15***	1.02***
	(7.84)	(12.60)	(13.44)	(14.15)	(20.14)	(20.10)	(20.07)	(20.59)	(20.65)	(20.10)	(19.81)	(19.49)	(14.36)	(10.19)
Explained														
Experience	-0.01	0.00	0.00	0.01	0.03***	0.03***	0.03***	0.05***	0.05***	0.06***	0.08***	0.09***	0.12***	0.10***
	(-1.37)	(0.42)	(0.70)	(1.59)	(3.56)	(3.65)	(3.70)	(4.08)	(4.09)	(4.13)	(4.23)	(4.24)	(4.24)	(4.04)
Education	0.30***	0.34***	0.31***	0.31***	0.33***	0.33***	0.34***	0.39***	0.39***	0.42***	0.51***	0.53***	0.76***	0.63***
	(11.05)	(13.18)	(13.13)	(13.54)	(15.03)	(15.10)	(15.11)	(15.20)	(15.14)	(15.04)	(15.15)	(15.20)	(14.40)	(12.63)
Isolation	0.36***	0.29***	0.26***	0.23***	0.21***	0.20***	0.20***	0.16***	0.17***	0.14***	0.18***	0.18***	0.35***	0.34***
	(7.50)	(7.00)	(6.77)	(5.98)	(6.62)	(6.18)	(6.34)	(5.15)	(5.34)	(4.20)	(4.87)	(4.68)	(6.27)	(4.82)
cape	0.01*	0.01*	0.01*	0.01*	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	-0.01*	-0.01
	(1.81)	(1.65)	(1.82)	(1.83)	(1.60)	(1.63)	(1.40)	(0.58)	(0.62)	(1.03)	(0.54)	(0.17)	(-1.70)	(-1.24)
johannesburg	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00
	(-1.09)	(-1.11)	(-1.10)	(-1.08)	(-1.15)	(-1.16)	(-1.15)	(-1.16)	(-1.16)	(-1.15)	(-1.11)	(-1.11)	(0.60)	(1.04)
Total	0.66***	0.65***	0.58***	0.56***	0.57***	0.56***	0.57***	0.60***	0.61***	0.61***	0.77***	0.79***	1.22***	1.06***
	(11.82)	(13.29)	(13.17)	(12.75)	(15.13)	(14.87)	(15.11)	(14.71)	(14.84)	(14.00)	(14.79)	(14.68)	(15.08)	(11.62)
Unexplained														
Experience	0.46	0.43	0.34	0.24	0.54***	0.51***	0.50***	0.25*	0.22	0.23	0.15	0.11	-0.17	-0.36*
	(1.02)	(1.45)	(1.20)	(0.91)	(3.30)	(3.21)	(3.12)	(1.76)	(1.56)	(1.60)	(0.99)	(0.73)	(-1.03)	(-1.83)
Education	4.97	3.27	2.92	2.39	-2.65	-2.88	-2.90	-2.37	-2.43	-0.96	-0.26	-0.31	-3.22	-7.74**
	(0.91)	(0.82)	(0.76)	(0.67)	(-1.18)	(-1.32)	(-1.38)	(-1.10)	(-1.13)	(-0.68)	(-0.16)	(-0.19)	(-1.62)	(-2.47)
Isolation	1.26*	1.08*	0.97*	0.76	0.94***	0.85**	0.73**	0.60*	0.66**	0.49	0.59*	0.62*	0.15	0.58
	(1.67)	(1.96)	(1.92)	(1.61)	(2.80)	(2.57)	(2.18)	(1.92)	(2.14)	(1.50)	(1.72)	(1.81)	(0.34)	(1.06)
cape	-0.01	-0.02	-0.02	-0.02	-0.01	-0.00	-0.00	-0.02	-0.02	-0.02	-0.01	-0.01	0.02	0.01
	(-0.24)	(-0.73)	(-0.73)	(-0.68)	(-0.40)	(-0.28)	(-0.06)	(-1.35)	(-1.38)	(-1.15)	(-0.75)	(-0.65)	(0.94)	(0.30)
johannesburg	0.06**	0.03	0.01	0.01	0.00	0.00	0.00	-0.01	-0.01	-0.02	-0.01	-0.01	-0.01	-0.00
	(2.26)	(1.52)	(0.51)	(0.54)	(0.01)	(0.08)	(0.02)	(-0.53)	(-0.50)	(-1.45)	(-0.81)	(-0.86)	(-1.20)	(-0.58)
Constant	-6.22	-4.13	-3.50	-2.65	1.92	2.24	2.38	2.24	2.26	0.94	0.06	0.07	3.16	7.47**
	(-1.11)	(-1.03)	(-0.90)	(-0.73)	(0.85)	(1.02)	(1.12)	(1.01)	(1.02)	(0.63)	(0.04)	(0.04)	(1.46)	(2.29)
Total	0.53***	0.66***	0.73***	0.74***	0.74***	0.73***	0.71***	0.69***	0.68***	0.66***	0.51***	0.47***	-0.07	-0.05
	(3.34)	(6.04)	(7.12)	(7.59)	(10.84)	(10.78)	(10.48)	(10.79)	(10.63)	(10.06)	(7.38)	(6.65)	(-0.69)	(-0.37)
Observations	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Even though segregation is diluted when introducing these metropolitan dummies, our main message remains. Segregation is still an important contributor to the income gap. In the composition effect, differences in segregation levels contribute positively to the income gap. They are as important as the differences in education for the bottom of the distribution, while they still account for half of the contribution of education at higher quantiles. In the structure effect, differences in the returns of segregation contribute positively to the income gap. They are also the most important factor of the structure effect.

This message is also confirmed when we look at the RIF-regressions. Even with the six metropolitan dummies, Whites still benefit from segregation at the higher quantiles, whereas Blacks suffer from segregation at the bottom of the distribution. Results are reported in Table 17 for Whites and Table 18 for Blacks.[44]

[44] We do not report results of the RIF-regressions with only two dummies since they convey exactly the same message. Interested readers might still contact the authors to get them.

Table 17: Impact of segregation (isolation) on the White income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
2008	-0.41	-0.11	-0.22	0.48	-0.07	-0.18	-0.03	-0.25	-0.33	-0.17	-0.29	-0.34	0.45	1.28**
	(-0.37)	(-0.14)	(-0.31)	(0.63)	(-0.12)	(-0.32)	(-0.05)	(-0.46)	(-0.62)	(-0.32)	(-0.53)	(-0.62)	(0.86)	(2.34)
2010	-0.51	-0.99	-1.20	-0.81	-0.33	-0.05	-0.07	1.72**	1.56**	1.12	1.23*	1.22*	1.56***	0.28
	(-0.44)	(-0.96)	(-1.23)	(-0.78)	(-0.38)	(-0.06)	(-0.08)	(2.24)	(2.01)	(1.51)	(1.79)	(1.81)	(2.61)	(0.46)
2012	1.43	-0.01	-0.55	-0.56	-0.43	-0.69	-0.73	-0.25	-0.40	-0.49	-0.51	-0.51	0.06	-0.53
	(1.01)	(-0.01)	(-0.58)	(-0.64)	(-0.56)	(-0.89)	(-0.94)	(-0.32)	(-0.52)	(-0.64)	(-0.64)	(-0.64)	(0.08)	(-0.36)
2014	-0.07	0.43	0.14	-0.13	0.29	0.14	0.05	0.12	0.32	0.30	0.54	0.54	1.00	1.75
	(-0.05)	(0.40)	(0.14)	(-0.13)	(0.43)	(0.21)	(0.07)	(0.19)	(0.54)	(0.49)	(0.87)	(0.87)	(1.19)	(1.64)
Pooled	-0.24	-0.05	-0.04	-0.04	-0.19	-0.16	-0.23	0.03	0.15	0.14	0.05	0.11	0.56*	1.11**
	(-0.38)	(-0.10)	(-0.07)	(-0.08)	(-0.54)	(-0.47)	(-0.66)	(0.08)	(0.46)	(0.43)	(0.17)	(0.35)	(1.77)	(2.41)

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Impact of segregation (isolation) on the Black income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
2008	-0.71**	-0.74***	-0.71***	-0.78***	-0.61***	-0.61***	-0.67***	-0.72***	-0.64***	-0.70***	-0.44**	-0.48**	0.07	0.02
	(-2.10)	(-3.46)	(-3.37)	(-3.88)	(-3.18)	(-3.21)	(-3.55)	(-3.83)	(-3.39)	(-3.59)	(-2.24)	(-2.42)	(0.28)	(0.07)
2010	-1.46***	-1.10***	-1.09***	-1.22***	-0.72***	-0.62***	-0.60***	-0.35	-0.39*	-0.54**	-0.64***	-0.58**	-0.06	-0.26
	(-4.07)	(-3.57)	(-3.51)	(-4.31)	(-3.06)	(-2.65)	(-2.61)	(-1.56)	(-1.73)	(-2.40)	(-2.69)	(-2.44)	(-0.27)	(-0.98)
2012	-0.81***	-0.46*	-0.47**	-0.30	-0.67***	-0.60***	-0.52**	-0.17	-0.23	-0.00	0.22	0.21	0.02	-0.22
	(-3.33)	(-1.89)	(-2.23)	(-1.46)	(-3.17)	(-2.81)	(-2.40)	(-0.81)	(-1.09)	(-0.02)	(1.28)	(1.22)	(0.11)	(-0.95)
2014	-1.44***	-1.21***	-1.15***	-0.97***	-0.76***	-0.68***	-0.52***	-0.30*	-0.29*	-0.08	-0.01	-0.05	0.22	0.15
	(-5.63)	(-5.20)	(-5.30)	(-4.80)	(-4.53)	(-4.24)	(-3.19)	(-1.95)	(-1.85)	(-0.49)	(-0.08)	(-0.31)	(1.13)	(0.77)
Pooled	-1.08***	-0.81***	-0.94***	-0.83***	-0.61***	-0.59***	-0.59***	-0.41***	-0.36***	-0.31***	-0.10	-0.09	0.09	-0.08
	(-6.61)	(-6.73)	(-8.05)	(-7.13)	(-5.93)	(-5.82)	(-5.77)	(-4.24)	(-3.73)	(-3.26)	(-1.03)	(-0.99)	(0.91)	(-0.68)

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

9 Conclusion

In this paper, we have proposed an analysis of the impact of segregation on the South African income distribution. Using generalization of the standard Oaxaca-Blinder method to the decomposition of distribution allows us first to overcome some of the previously encountered difficulties of the literature such as specification issues and nonlinear patterns. After controlling for selection issues, we show that segregation can account for up to 40% of the income gap between Blacks and Whites. We also put some lights on the heterogeneity of the impact of segregation. Poor Black ghettos and rich White ghettos experience a negative effect, and a positive effect respectively, which strengthen as we look at increasingly extreme quantiles of the income distribution. This suggests that different economic mechanisms (vulnerability to economic shocks, ease of doing business...) are at play in these locations. Even when we find a negative effect of segregation for Whites in the Census, Blacks still suffer more. The negative effect for Whites may arise from a negative correlation between income and segregation if the most affluent Whites are also the most integrated.

Further studying the different economic mechanisms at play is one potential avenue for future research. Because there is still a debate on whether there are neighborhood effects or not,[45] no government is actually really fighting segregation seriously. So, in a world where we have to live with segregation, it is important to understand the different mechanisms at play in the different type of ghettos. Moreover, knowing the welfare effects implied by residential segregation are important. Along this line of research, some work has already been done on the welfare effects of occupational segregation and might be a good beginning for residential segregation.[46] In this respect, one of our finding for South Africa is that racial segregation contributes to within-group income inequality in this context.

Finally, another caveat of our analysis is the lack of causal evidence. As in most of the decomposition analyses, the causal aspects are not studied and the analysis is only descriptive. Having causal estimates would be a great deal in order to draw attention of policy makers to the problem of segregation. This would also make a great question to explore in future research.

[45]See again the discussion in the introduction about this question.
[46]See Alonso-Villar and Del Rìo[1] for instance.

10 Appendices

10.1 Example of the difference between dissimilarity and isolation measures in decomposition

Imagine the fictional country depicted in Figure 5. It is composed of four cities themselves divided into four neighborhoods. Red and Grey dots represent individuals of two groups (which may be Blacks and Whites). Segregation is complete in two of them, while the last two neighborhoods are respectively almost completely integrated and moderately segregated according to the dissimilarity index.[47]

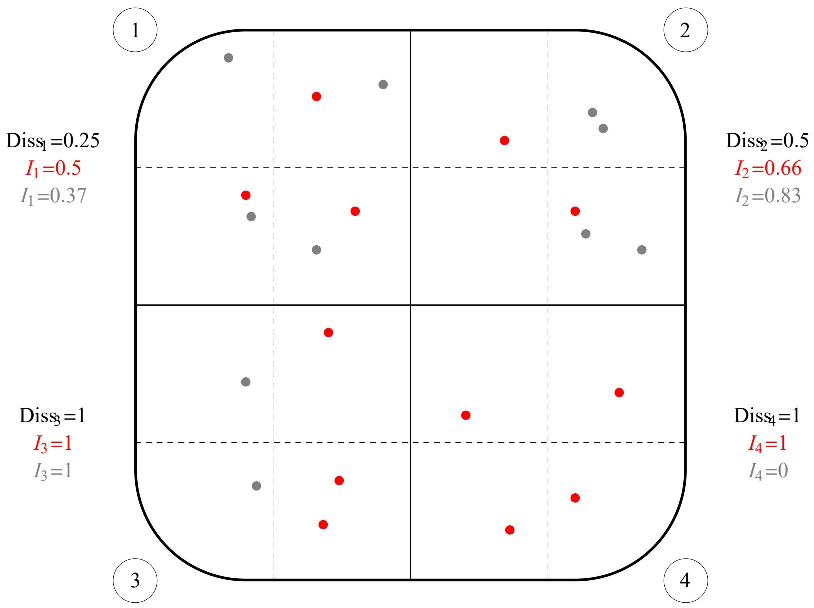


Figure 5: An example of the different impact of segregation indices for decompositions. "Diss" stands for dissimilarity while "I" means isolation.

Now if we compute the average segregation experienced by the Reds with the dissimilarity index, we have the following computation:

$$\bar{S}_R = \frac{3 \times 0.25 + 2 \times 0.5 + 7 \times 1}{12} \tag{12}$$

while for the Grays, we have:

$$\bar{S}_G = \frac{4 \times 0.25 + 4 \times 0.5 + 2 \times 1}{10} \tag{13}$$

[47]Segregation is considered severe for measures of the dissimilarity index above 0.7, while it is moderate between 0.3 and 0.7, and low for values below 0.3 (Kantrowitz[24]).

Thus, the only thing that is varying for the difference in the average segregation (that would appear in the composition term of the decomposition) is the probability for the two groups to live in a location with a given level of segregation. Let rewrite it to make it clearer:

$$
\begin{aligned}
\bar{S}_G - \bar{S}_R &= 0.25 \times (\frac{4}{10} - \frac{3}{12}) + 0.5 \times (\frac{4}{10} - \frac{2}{12}) + 1 \times (\frac{2}{10} - \frac{7}{12}) \\
&= 0.25 \times (\mathbb{P}[A|D = G] - \mathbb{P}[A|D = R]) \\
&+ 0.5 \times (\mathbb{P}[B|D = G] - \mathbb{P}[B|D = R]) \\
&+ 1 \times (\mathbb{P}[C|D = G] - \mathbb{P}[C|D = R])
\end{aligned}
\tag{14}
$$

where the event A,B, and C are respectively "living in a location segregated at $Diss = 0.25$", "... at $Diss = 0.5$", and "... at $Diss = 1$". In the extreme case where these probabilities are the same for the two groups, the whole composition effect cancel out from the decomposition. The further they drift apart, the less diluted the composition effect. Moreover, it is not clear whether this difference in the probability to live in a location at a given segregation level is exactly measuring differences in segregation. Finally, in small samples, the share of individuals of each group living in a location with a given level of segregation are probably a biased estimator of these probabilities. This might even be accentuated when the locations considered are small areas like census blocks.

The average isolation measure for the Reds will take the following value:

$$\bar{S}_R = \frac{3 \times 0.5 + 2 \times 0.66 + 7 \times 1}{12} \tag{15}$$

and for the Greys:

$$\bar{S}_G = \frac{4 \times 0.37 + 4 \times 0.83 + 2 \times 1}{10} \tag{16}$$

We cannot factorize by the level of segregation anymore with such measure. However, the average level of exposure of one type has a nice interpretation as the probability to pick (with replacement) two individuals of the same type in the same location in the country, given the type. Considering equation (16), we know that exposure measures the probability to pick two individuals in a location that are of the same/different groups, so we can rewrite the average as:

$$
\begin{aligned}
\bar{S}_G &= \frac{4}{10} \times 0.37 + \frac{4}{10} \times 0.83 + \frac{2}{10} \times 1 \\
&= \mathbb{P}[B = i_1 | D = G] \times \mathbb{P}[A_i | B = i_1, D = G] \\
&+ \mathbb{P}[B = i_2 | D = G] \times \mathbb{P}[A_i | B = i_2, D = G] \\
&+ \mathbb{P}[B = i_3 | D = G] \times \mathbb{P}[A_i | B = i_3, D = G] \\
&= \sum_{i \in I} \mathbb{P}[A_i \cap B = i | D = G] \\
&= \mathbb{P}[A | D = G]
\end{aligned}
\tag{17}
$$

where "B=i" is the event "Live in location i", "A_i" is the event "Draw two individuals of the same type in location i". Thus, Differences in the average isolation levels will correspond to percentage point differences in the probability to draw two identical individuals in the country.

10.2 Additional results with the index of Dissimilarity

Table 19: Linear regressions (dissimilarity)

	2008			2010			2012		
	White	Black	Pooled	White	Black	Pooled	White	Black	Pooled
Experience	0.08***	0.04***	0.05***	0.07***	0.04***	0.04***	0.04***	0.03***	0.03***
	(6.92)	(10.76)	(12.33)	(5.36)	(8.83)	(10.19)	(2.69)	(9.67)	(10.13)
Experience (square)	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00*	-0.00***	-0.00***
	(-5.94)	(-5.96)	(-6.79)	(-4.18)	(-4.72)	(-5.26)	(-1.95)	(-3.52)	(-3.50)
Years of schooling	-0.17	-0.05***	-0.05***	0.22	-0.03**	-0.03*	-0.02	-0.03**	-0.03**
	(-1.52)	(-3.97)	(-4.47)	(1.25)	(-2.05)	(-1.87)	(-0.10)	(-2.52)	(-2.45)
Years of schooling (square)	0.01***	0.01***	0.01***	-0.00	0.01***	0.01***	0.01	0.01***	0.01***
	(3.23)	(16.83)	(19.93)	(-0.26)	(12.41)	(13.83)	(0.89)	(16.06)	(17.08)
Dissimilarity	1.40***	-0.71***	-0.56***	1.05*	-0.80***	-0.69***	-0.17	-0.72***	-0.77***
	(3.12)	(-3.74)	(-3.03)	(1.69)	(-3.56)	(-3.19)	(-0.26)	(-4.12)	(-4.46)
Constant	6.84***	6.90***	6.68***	4.80***	7.00***	6.77***	7.80***	7.11***	7.09***
	(8.51)	(39.19)	(38.92)	(3.83)	(32.92)	(32.68)	(4.91)	(43.08)	(43.49)
Observations	440	2922	3362	243	3088	3331	261	3996	4257
R^2	0.349	0.347	0.424	0.364	0.263	0.319	0.227	0.317	0.350

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Linear regressions (dissimilarity, continued)

	2014			Pooled			Census		
	White	Black	Pooled	White	Black	Pooled	White	Black	Pooled
Experience	0.03**	0.02***	0.02***	0.05***	0.03***	0.03***	0.17***	0.12***	0.13***
	(2.43)	(7.46)	(7.92)	(8.78)	(17.37)	(19.54)	(129.46)	(311.08)	(348.00)
Experience (square)	-0.00	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***	-0.00***
	(-1.22)	(-2.94)	(-2.66)	(-6.76)	(-7.58)	(-8.34)	(-106.65)	(-77.07)	(-110.75)
Years of schooling	-0.32	-0.06***	-0.06***	-0.08	-0.04***	-0.04***	-0.02	-0.17***	-0.20***
	(-1.18)	(-4.95)	(-4.70)	(-0.91)	(-6.02)	(-6.01)	(-1.61)	(-84.32)	(-108.95)
Years of schooling (square)	0.02*	0.01***	0.01***	0.01***	0.01***	0.01***	0.02***	0.03***	0.03***
	(1.92)	(17.25)	(17.98)	(3.20)	(30.94)	(33.78)	(35.13)	(233.56)	(292.08)
Dissimilarity	0.94	-0.63***	-0.64***	0.94***	-0.67***	-0.65***	-0.54***	-4.70***	-5.00***
	(1.35)	(-3.77)	(-3.93)	(3.26)	(-7.13)	(-7.15)	(-3.00)	(-83.57)	(-92.18)
Constant	8.47***	7.39***	7.32***	6.86***	7.08***	6.97***	2.61***	4.56***	4.86***
	(4.59)	(46.53)	(46.64)	(11.68)	(80.43)	(80.89)	(14.98)	(88.33)	(97.67)
Observations	229	5291	5520	1173	15297	16470	263334	2248132	2511466
R^2	0.279	0.260	0.287	0.300	0.295	0.339	0.146	0.174	0.207

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 20: Oaxaca decompositions

	2008	2010	2012	2014	Pooled	Census
Differential						
Prediction_1	9.09***	9.20***	9.19***	9.20***	9.16***	9.14***
	(214.76)	(167.89)	(175.69)	(161.92)	(361.67)	(3179.77)
Prediction_2	7.70***	7.80***	7.96***	8.01***	7.89***	7.39***
	(455.18)	(429.09)	(560.12)	(626.79)	(1035.33)	(6246.91)
Difference	1.39***	1.40***	1.23***	1.20***	1.26***	1.75***
	(30.56)	(24.20)	(22.66)	(20.55)	(47.78)	(561.74)
Explained						
Experience	-0.01	0.06***	0.04***	0.06***	0.04***	0.11***
	(-0.57)	(3.67)	(2.73)	(4.13)	(5.37)	(120.49)
Education	0.64***	0.57***	0.51***	0.46***	0.52***	0.79***
	(20.83)	(14.86)	(15.77)	(14.83)	(32.33)	(341.67)
Dissimilarity	0.00	0.00	0.01	0.01**	0.01***	0.03***
	(1.44)	(0.97)	(1.51)	(2.07)	(3.55)	(72.83)
Total	0.63***	0.63***	0.56***	0.53***	0.56***	0.92***
	(19.65)	(14.77)	(15.43)	(15.03)	(31.93)	(383.95)
Unexplained						
Experience	0.18	0.24	-0.12	0.10	0.13*	-0.62***
	(1.46)	(1.58)	(-0.85)	(0.87)	(1.95)	(-81.67)
Education	-1.15**	1.15	-0.37	-1.84	-0.58	-0.32***
	(-1.96)	(0.93)	(-0.23)	(-1.05)	(-1.18)	(-9.66)
Dissimilarity	1.78***	1.57***	0.47	1.33**	1.36***	0.45***
	(4.47)	(2.87)	(0.77)	(2.33)	(5.34)	(7.57)
Constant	-0.06	-2.20	0.69	1.08	-0.22	1.32***
	(-0.08)	(-1.56)	(0.39)	(0.58)	(-0.38)	(19.01)
Total	0.76***	0.76***	0.67***	0.67***	0.70***	0.82***
	(17.90)	(14.50)	(13.17)	(12.75)	(29.15)	(250.67)
Observations	3362	3331	4257	5520	16470	1334468

t statistics in parentheses

Prediction_1 is the predicted logarithm of the real monthly income of Whites (in 2014(Nov.) Rands).

Prediction_2 is the predicted logarithm of the real monthly income of Blacks (in 2014(Nov.) Rands).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 21: Impact of segregation (dissimilarity) on the White income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
2008	2.75**	2.63***	2.53***	2.10***	0.92	0.90	0.97	0.75	0.62	0.65	0.59	0.54	1.30**	2.14***
	(2.39)	(3.43)	(3.39)	(2.86)	(1.50)	(1.47)	(1.58)	(1.24)	(1.04)	(1.12)	(1.04)	(0.94)	(2.47)	(3.07)
2010	2.32*	1.47	1.55	1.09	0.24	0.05	0.20	0.96	0.95	0.90	1.18	1.28*	1.76***	0.39
	(1.77)	(1.15)	(1.26)	(0.92)	(0.28)	(0.05)	(0.24)	(1.18)	(1.18)	(1.16)	(1.54)	(1.68)	(2.82)	(0.68)
2012	0.40	-0.34	-0.29	0.00	0.17	0.04	-0.35	0.31	0.38	0.24	0.15	0.15	-0.38	-0.36
	(0.22)	(-0.30)	(-0.27)	(0.00)	(0.19)	(0.05)	(-0.39)	(0.36)	(0.45)	(0.28)	(0.18)	(0.18)	(-0.44)	(-0.23)
2014	1.62	0.56	0.29	0.12	1.22	1.16	1.43*	0.88	0.74	0.85	1.23	1.23	1.79*	2.74***
	(0.73)	(0.40)	(0.23)	(0.10)	(1.48)	(1.44)	(1.74)	(1.17)	(0.99)	(1.08)	(1.50)	(1.50)	(1.90)	(2.65)
Pooled	2.44***	1.51***	1.55***	1.19**	0.56	0.64*	0.64	0.82**	0.65*	0.69*	0.50	0.53	1.00***	1.72***
	(3.26)	(2.70)	(2.79)	(2.28)	(1.44)	(1.65)	(1.64)	(2.28)	(1.83)	(1.94)	(1.44)	(1.51)	(2.83)	(3.66)

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 22: Impact of segregation (dissimilarity) on the Black income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
2008	-0.76*	-0.80***	-0.82***	-0.91***	-0.78***	-0.69***	-0.67***	-0.60***	-0.50**	-0.66***	-0.84***	-0.96***	-1.15***	-0.15
	(-1.90)	(-3.19)	(-3.38)	(-3.87)	(-3.61)	(-3.20)	(-3.08)	(-2.60)	(-2.13)	(-2.65)	(-3.12)	(-3.51)	(-3.42)	(-0.44)
2010	-0.98**	-0.77**	-0.73**	-0.75**	-0.10	-0.10	-0.02	-0.04	-0.06	-0.31	-0.83***	-0.72**	-1.31***	-1.29***
	(-2.14)	(-2.17)	(-2.03)	(-2.24)	(-0.37)	(-0.41)	(-0.08)	(-0.17)	(-0.22)	(-1.15)	(-2.92)	(-2.49)	(-4.21)	(-3.73)
2012	-0.97***	-0.63**	-0.57**	-0.43*	-0.47*	-0.49**	-0.49**	-0.57**	-0.53**	-0.50**	-0.55**	-0.55**	-0.91***	-1.06***
	(-3.30)	(-2.27)	(-2.37)	(-1.91)	(-1.95)	(-1.98)	(-1.98)	(-2.36)	(-2.16)	(-2.10)	(-2.40)	(-2.43)	(-3.24)	(-3.26)
2014	-1.28***	-1.00***	-0.90***	-0.81***	-0.39**	-0.32*	-0.34*	-0.14	-0.16	-0.30	-0.31*	-0.40**	-0.84***	-0.68***
	(-3.65)	(-3.14)	(-3.02)	(-2.99)	(-1.97)	(-1.71)	(-1.82)	(-0.79)	(-0.91)	(-1.64)	(-1.65)	(-2.06)	(-3.13)	(-2.60)
Pooled	-1.12***	-0.78***	-0.78***	-0.70***	-0.38***	-0.34***	-0.35***	-0.36***	-0.37***	-0.45***	-0.60***	-0.63***	-0.90***	-0.86***
	(-5.72)	(-5.49)	(-5.50)	(-5.06)	(-3.24)	(-2.88)	(-2.99)	(-3.19)	(-3.23)	(-3.93)	(-5.06)	(-5.27)	(-6.26)	(-5.40)

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 23: Oaxaca decomposition of the 2014 income distribution

	10th	15th	16th	17th	33rd	34th	35th	45th	46th	50th	56th	57th	75th	90th
Differential														
Prediction_1	8.09***	8.45***	8.50***	8.55***	8.93***	8.96***	8.98***	9.14***	9.17***	9.25***	9.32***	9.34***	9.77***	10.24***
	(54.46)	(83.28)	(89.54)	(95.35)	(140.32)	(142.87)	(144.25)	(150.31)	(150.78)	(149.70)	(148.11)	(148.41)	(126.45)	(105.05)
Prediction_2	6.90***	7.14***	7.19***	7.25***	7.61***	7.67***	7.70***	7.86***	7.88***	7.98***	8.04***	8.08***	8.62***	9.22***
	(295.77)	(341.02)	(366.81)	(405.27)	(558.75)	(586.65)	(598.91)	(624.77)	(624.90)	(608.66)	(580.57)	(561.83)	(407.88)	(453.22)
Difference	1.18***	1.31***	1.31***	1.30***	1.31***	1.29***	1.28***	1.28***	1.29***	1.27***	1.28***	1.26***	1.15***	1.02***
	(7.87)	(12.65)	(13.49)	(14.21)	(20.21)	(20.17)	(20.14)	(20.66)	(20.72)	(20.16)	(19.88)	(19.56)	(14.40)	(10.22)
Explained														
Experience	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***	0.06***
	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)
Education	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***	0.46***
	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)	(14.84)
Dissimilarity	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**	0.01**
	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)	(2.11)
Total	0.53***	0.53***	0.53***	0.53***	0.53***	0.53***	0.53***	0.53***	0.53***	0.53***	0.53***	0.53***	0.53***	0.53***
	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)	(15.06)
Unexplained														
Experience	0.26	0.30	0.25	0.16	0.49***	0.46***	0.46***	0.22	0.19	0.24	0.16	0.13	-0.10	-0.32*
	(0.58)	(1.01)	(0.91)	(0.63)	(2.94)	(2.86)	(2.78)	(1.57)	(1.34)	(1.64)	(1.07)	(0.85)	(-0.62)	(-1.69)
Education	3.27	2.18	2.26	1.78	-3.09	-3.31	-3.34	-2.67	-2.79	-1.03	-0.43	-0.46	-2.90	-7.70**
	(0.57)	(0.51)	(0.56)	(0.47)	(-1.36)	(-1.50)	(-1.57)	(-1.22)	(-1.29)	(-0.80)	(-0.29)	(-0.31)	(-1.58)	(-2.48)
Dissimilarity	2.47	1.33	1.02	0.79	1.36*	1.26*	1.49**	0.86	0.76	0.97	1.30*	1.38*	2.23***	2.91***
	(1.30)	(1.10)	(0.92)	(0.76)	(1.89)	(1.78)	(2.09)	(1.31)	(1.16)	(1.41)	(1.82)	(1.93)	(2.68)	(3.20)
Constant	-5.34	-3.03	-2.75	-1.97	2.02	2.35	2.14	2.34	2.60	0.57	-0.27	-0.31	1.39	5.60*
	(-0.86)	(-0.68)	(-0.65)	(-0.50)	(0.85)	(1.02)	(0.95)	(1.00)	(1.13)	(0.37)	(-0.15)	(-0.17)	(0.66)	(1.68)
Total	0.66***	0.78***	0.78***	0.77***	0.79***	0.76***	0.75***	0.76***	0.76***	0.75***	0.75***	0.73***	0.63***	0.49***
	(4.45)	(7.83)	(8.34)	(8.70)	(12.89)	(12.70)	(12.40)	(13.33)	(13.42)	(12.94)	(12.68)	(12.35)	(8.46)	(5.18)
Observations	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520	5520

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

References

[1] Olga Alonso-Villar and Coral Del Río. Local Segregation and Well-Being. *Review of Income and Wealth*, pages 1–19, 2016.

[2] Patrick Bayer, Hanming Fang, and Robert McMillan. Separate When Equal? Racial Inequality and Residential Segregation. *Journal of Urban Economics*, (May), may 2014.

[3] Patrick Bayer, Stephen L. Ross, and Giorgio Topa. Place of Work and Place of Residence : Informal Hiring Networks and Labor Market Outcomes. *Journal of Political Economy*, 116(6):1150–1196, 2008.

[4] Alan S. Blinder. Wage Discrimination : Reduced Form and Structural Estimates. *Journal of Human Resources*, 8(4):436–455, 1973.

[5] George J. Borjas. Ethnic capital and intergenerational mobility. *Quarterly Journal of Economics*, (February):123–150, 1992.

[6] George J. Borjas. Ethnicity, Neighborhoods, and Human-Capital Externalities. *American Economic Review*, 85(3):365–390, 1995.

[7] George J. Borjas. To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation. *Journal of Urban Economics*, 44(2):228–253, sep 1998.

[8] Yann Bramoullé and Sanjeev Goyal. Favoritism. *Journal of Development Economics*, 122:16–27, 2016.

[9] Yann Bramoullé and Rachel Kranton. Risk Sharing across Communities. *American Economic Review Papers and Proceedings*, 97(2):70–74, 2007.

[10] Jan K. Brueckner, Jacques-François Thisse, and Yves Zenou. Why is central Paris rich and downtown Detroit poor ? An amenity-based theory. *European Economic Review*, 43:91–107, 1999.

[11] Raj Chetty, Nathaniel Hendren, and Lawrence F. Katz. The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, 106(4):855–902, 2016.

[12] William J Collins and Robert A Margo. Residential segregation and socioeconomic outcomes: When did ghettos go bad? *Economics Letters*, 69(2):239–243, 2000.

[13] David M. Cutler and Edward L. Glaeser. Are ghettos good or bad? *Quarterly Journal of Economics*, pages 827–872, 1997.

[14] David M Cutler, Edward L Glaeser, and Jacob L Vigdor. Is the Melting Pot Still Hot? Explaining the Resurgence of Immigrant Segregation. *The Review of Economics and Statistics*, 90(3):478–497, 2008.

[15] David M. Cutler, Edward L. Glaeser, and Jacob L. Vigdor. When are ghettos bad? Lessons from immigrant segregation in the United States. *Journal of Urban Economics*, 63(3):759–774, may 2008.

[16] Federico Echenique and Roland G. Jr. Fryer. A Measure of Segregation Based on Social Interactions. *Quarterly Journal of Economics*, 72(2):441–486, 2007.

[17] Per-Anders Edin, Peter Fredriksson, and Olof Aslund. Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment. *Quarterly Journal of Economics*, 118(1):329–357, 2003.

[18] Sergio Firpo, Nicole M. Fortin, and Thomas Lemieux. Unconditional Quantile Regressions. *Econometrica*, 77(3):953–973, 2009.

[19] Nicole Fortin, Thomas Lemieux, and Sergio Firpo. Decomposition Methods in Economics. In David Card and Orley Ashenfelter, editors, *Handbook of Labor Economics*, volume 4, chapter 1, pages 1–102. Elsevier, 2011.

[20] David M. Frankel and Oscar Volij. Measuring school segregation. *Journal of Economic Theory*, 146(1):1–38, jan 2011.

[21] Robert Hutchens. One Measure of Segregation. *International Economic Review*, 45(2):555–578, may 2004.

[22] Ben Jann. The Blinder-Oaxaca Decomposition for Linear Regression Models. *The Stata Journal*, 8(4):453–479, 2008.

[23] John F. Kain. Housing Segregation, Negro Employment, and Metropolitan Decentralization. *Quarterly Journal of Economics*, 82(2):175–197, 1968.

[24] Nathan Kantrowitz. Racial and Ethnic Residential Segregation in Boston 1830-1970. *Annals of the American Academy of Political and Social Science*, 441(January):41–54, 1979.

[25] Lawrence F. Katz, Jeffrey R. Kling, and Jeffrey B. Liebman. Moving to Opportunity in Boston : Early Results of a Randomized Mobility Experiment. *Quarterly Journal of Economics*, 116(2):607–654, 2001.

[26] Micere Keels, Greg J. Duncan, Stefanie Deluca, Ruby Mendenhall, and James Rosenbaum. Fifteen years later: can residential mobility programs provide a long-term escape from neighborhood segregation, crime, and poverty? *Demography*, 42(1):51–73, feb 2005.

[27] Jeffrey R. Kling, Jeffrey B. Liebman, and Lawrence F. Katz. Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1):83–119, 2007.

[28] Stanley Lieberson and Donna K Carter. Temporal Changes and Urban Differences in Residential Segregation : A Reconsideration. *American Journal of Sociology*, 88(2):296–310, 1982.

[29] Jens Ludwig, Greg J Duncan, Lisa A Gennetian, Lawrence F Katz, Ronald C Kessler, Jeffrey R Kling, and Lisa Sanbonmatsu. Long-term neighborhood effects on low-income families: evidence from moving to opportunity. *American Economic Review Papers and Proceedings*, 103(3):226–231, 2013.

[30] Douglas S. Massey and Nancy A Denton. The Dimensions of Residential Segregation. *Social Forces*, 67(2):281–315, 1988.

[31] Angelo Mele. Poisson indices of segregation. *Regional Science and Urban Economics*, 43(1):65–85, jan 2013.

[32] Ronald L. Oaxaca. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709, 1973.

[33] Philip Oreopoulos. The Long-Run Consequences of Living in a Poor Neighborhood. *Quarterly Journal of Economics*, 118(4):1533–1575, 2003.

[34] Paul B. Rich. Race, Science, and the Legitimization of White Supremacy in South Africa, 1902-1940. *International Journal of African Historical Studies*, 23(4):665–686, 1990.

[35] Maria Sáez-Martí and Yves Zenou. Cultural transmission and discrimination. *Journal of Urban Economics*, 72(2-3):137–146, sep 2012.

[36] Rajiv Sethi and Rohini Somanathan. Inequality and Segregation. *Journal of Political Economy*, 112(6):1296–1321, 2004.

[37] Nicholas Spaull. Poverty & Privilege: Primary School Inequality in South Africa. *International Journal of Educational Development*, 33(5):436–447, 2013.

[38] Leonard Thompson. *A History of South Africa.* Yale University Press, 3rd revise edition, 2001.