https://economix.fr/

The Contribution of Residential Segregation to Racial Income Gaps: Evidence from South Africa

Document de Travail
Working Paper
2020-20

Florent Dubois
Christophe Muller

UMR 7235

EconomiX - UMR7235
Université Paris Nanterre
Bâtiment G - Maurice Allais, 200, Avenue de la République
92001 Nanterre cedex

Email : secretariat@economix.fr

THE CONTRIBUTION OF RESIDENTIAL SEGREGATION TO RACIAL INCOME GAPS: EVIDENCE FROM SOUTH AFRICA*

FLORENT DUBOIS† AND CHRISTOPHE MULLER‡

July 2020

Abstract

Persistent racial income disparities cannot only be explained by differences in socio-economic characteristics. In this paper, we contend that local segregation should be an essential component of the determination of socio-ethnic income gaps using the contemporary White/African gap in South Africa. First, we complete Mincer wage equations with an Isolation index. Second, we decompose the income gap distribution into detailed composition and structure components. Third, we explore the heterogeneity of segregation effects along three theoretical lines: racial preferences, labor market segmentation, and networks effects. Segregation is found to be the main contributor of the structure effect, ahead of education and experience, and to make a sizable contribution to the composition effect. Moreover, segregation is detrimental to incomes at the bottom of the African distribution, notably in association with local informal job-search networks, while it is beneficial at the top of the White distribution. Only minor influences of racial preferences and labor market segmentation are found. Specific subpopulations are identified that suffer and benefit most from segregation, including for the former, little educated workers in agriculture and mining, often female, immersed in their personal networks. Finally, minimum wage policies are found likely to attenuate the segregation's noxious mechanisms.

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

*We are grateful for financial support from the TMENA2 project (ANR-17-CE39-0009-01) granted by the French Agence Nationale de la Recherche. We would like to thank Jean-Paul Azam, Yann Bramoullé, Olga Cantó-Sánchez, Eliane El Badaoui, Fabien Moizeau, and Michel Lubrano for their invaluable comments. This paper also benefited from insights of conference participants at the 2018 LAGV conference (Aix-en-Provence) and the 2020 ADRES conference (Lyon). We would like to express our gratitude to Takwanisa Machemedze, Alex Montgomery, and Lynn Woolfrey at DataFirst for their expert assistance with the data. All remaining errors are ours. Any views expressed in this paper are solely the authors' and do not reflect the position of any institution.

†corresponding author: florent.dubois@parisnanterre.fr

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

I. Introduction

Persistent racial differences in socio-economic outcomes is a major concern as it threatens the stability of increasingly diverse modern societies. Earlier attempts to legally justify such differences were only met with a fierce resistence of civil rights movements. The Apartheid regime clashed with the African resistence during most of the twentieth century in South Africa while, in the United States, the Civil Rights movement battled relentlessly against the Jim Crow legislation. However, explaining the persistence of such differences, most notably income differences (Bayer and Charles 2018; Leibbrandt et al. 2010b), even after these discriminatory legislations were repealed still constitutes a challenge for social scientists.

Most of the literature usually invokes labour market mechanisms, operating either through human capital endowment gaps or discrimination (Altonji and Blank 1999). A notable exception is the recent contribution of Chetty et al. (2020) which shed lights on the role played by differences in intergenerational mobility. Yet, among pre-market factors, residential segregation has received much less attention than education or family backgrounds. Most of the time, it is limited to the analysis of the spatial mismatch hypothesis, the tendency for minority members living far from job centers to select poorer alternatives (Kain 1968). However, because the working decision and the opportunity selected may depends as well on people's interactions within their neighborhood, alternative mechanisms based on racial preferences or neighborhood effects may as well generate income differences.

Nevertheless, there is limited and mixed evidence on the effect of segregation on incomes, almost exclusively based on the United States, and the relative importance of segregation compared to other determinants is rarely assessed. Cutler and Glaeser (1997) find on average a negative effect, while Edin et al. (2003) estimate a positive effect and Oreopoulos (2003), consistently with the results on the Moving To Opportunity experiment (Katz et al. 2001), do not spot any effect at all. These disparate findings may be partly explained by the heterogeneous effect of segregation. In this regards, Cutler et al. (2008) show that segregation has a different impact on immigrants' wages depending on their education group, and Chetty et al. (2016) uncover a positive effect on the children's wages if they move into lower-poverty neighborhoods before they are 13, but a small negative effect for older ones.

In this paper, we provide new evidence on the importance of segregation for racial income differences. We use South Africa as a benchmark case for examining these issues. Arguably, this is a most relevant case as the country combines the longest and most pronounced experience of legally enforced segregation with one of the largest racial income gaps in the world.

Our analysis is organized into two parts. First, we document the effect of segregation on incomes in South Africa using a simple Mincerian framework. This analysis of the mean effect of segregation has a twofold virtue. It establishes a premier estimate for South Africa, as the literature is actually silent (Sherer 2000; Gradín 2012, 2014; Leibbrandt et al. 2010b), and provides a useful benchmark of the aggregated pattern. Then, we scrutinize the heterogeneity of the association of incomes with residential segregation with RIF-regressions and generalized decompositions (Firpo et al. 2009; Fortin et al. 2011), which allows us to identify for whom segregation matters most.

Consistently with Cutler and Glaeser (1997), we find that segregation has, on average, a negative effect on incomes for Africans. The average effect for Whites is positive but unstable over time. In terms of relative contributions, segregation is the main contributor to the racial income gap, even ahead of education. This is particularly true for the structure effect. The differential effect of segregation becomes even clearer in the distributional analysis. We find that segregation is associated negatively with income at the bottom of the African distribution, while positively for the top of the White distribution. Again, segregation pops up as the main contributor in the structure effect with the strongest positive contribution below the median.

In the second part of the paper, we exploit our estimations through the prism of the 2018 minimum wage reform in South Africa. We observe that the distinct positions advocated about this reform by the major political actors imply contrasting views about the resulting influence of segregation on the income gap. Searching for natural constituencies of these parties, we examine three socio-economic mechanisms that might explain how segregation impacts incomes, namely: network effects, racial preferences, and labor market segmentation. For doing so, we employ identifying variables for each mechanism in ancillary regressions of the estimated quantile effect of segregation. We complete our analysis of the heterogeneity of segregation effects by detailing the socio-demo-economic characteristics of the African main winners and losers from segregation with a classification analysis in the spirit of Chernozhukov et al. (2018).

As a result, we exhibit the local within-group network as an important channel mediating the effect of segregation on income gaps. Moreover, these network effects are heterogeneous, and even opposed, in different sections of the distributions and across groups. Second, the 2018 minimum wage reform is found likely to attenuate segregation influence on wages for Africans, in a fashion consistent with the political constituencies of the main political parties in this country. Last, the classification analysis points at gender differences, network effects, and unionship as important sources of heterogeneity between winners and losers. However, education is unlikely to matter much for heterogeneity as both winners and losers, on average, quit schooling without any diploma.

The remainder of the paper proceeds as follows: Section II. discusses the potential economic channels through which segregation might affect income. Section III. describes the measure of segregation and the inference problem in mean and distribution decompositions that some measures elicit. Section IV. reviews segregation during Apartheid and the post-Apartheid trends of income inequality, and presents the data. Section V. expounds on the results obtained by decomposing the mean income gap. Section VI. extends the analysis to the entire income distribution. Section VII. discusses the potential consequences of the 2018 minimum wage reform for the racial gap and segregation. This section also examines the potential socio-economic mechanisms conveying the effect of segregation on incomes. The last section concludes the paper.

II. How Segregation Relates to Incomes

II.A. Individual Preferences for Segregation

In the housing market, segregation, through racial preferences, transforms neighborhoods into clubs and restrict the access to their amenities. Individuals, when deciding where to live, takes into account the racial composition of their targeted neighborhood. Schelling (1971) demonstrates that only a small preference for their own ethnic group is enough to yield highly segregated local contexts. Realtors also play an important role. They can employ discriminatory tactics, such as redlining, because they are themselves racist or because they care for the racial preferences of their current or potential customers (Yinger 1986). Once established somewhere, individuals will vote for their contribution to local public goods. Alesina et al. (1999) show that individuals in more diverse communities are voting for less spending in education when it also benefits to the minority group. Therefore, it generates differences in human capital accumulation through differences in education quality, finally materializing into differences in wages. Then, local segregation and income levels correlates.

This mechanism is amplified if there is already an initial correlation between income and segregation. This would be the case if racial groups are hierarchized by income. Note that this group positioning may be itself at the origin of racial prejudice (Blumer 1958).

II.B. Neighborhood and Peer Effects

Segregation may act on income by the facilitated diffusion of labor-related behaviors inside an homogeneous population. Then, individuals living in highly segregated areas may be more prone to develop specific work habits when they belong to some specific group and therefore to be subjected to group-specific income processes. For example, in the US, Black workers living in ghettos are sometimes believed to be characterized by tardiness, absenteeism, or unreliability, and this may be one reason for their lower incomes. Wilson (1987) claimed that it was inner-city isolation that generated bad work habits. In particular, there is some evidence of a 'ghetto culture' of bad habits that tends to reinforce these habits through social pressure. Even children often feel peer pressure to perform poorly at school. In these conditions, it may be difficult to escape unemployment and poorly paying jobs from within the ghetto. Bénabou (1993) shows that neighborhood and peer effects can explain some individuals' low quality of work. If one's 'peers' are defined in close connection to ethnicity, then the isolation index that we use measures the extent of such social pressure. Besides, social pressure may foster bad practices in one group and good work habits in another, which may further pull apart the incomes of the two groups under segregation.

In addition, ethnic networks may provide differential access to jobs and work promotions (Magruder 2010). In particular, local segregation against one group may limit its access to professional information obtained by other groups (Ioannides and Loury 2004). Ethnically isolated individuals may have lower incomes, ceteris paribus, because their information set is smaller.

II.C. Segmented Labor Markets, Capital Ownership, Trade Unions, and Spatial Mismatch

Segregation may affect income levels by contributing to the segmentation of the labor market (Dickens and Lang 1985; Magnac 1991). Entrepreneurs may pay lower wages to the discriminated group if recruited because they are themselves racist and perceive a cost of employing a minority worker (Becker 1957). In addition, if racist entrepreneurs settle disproportionately in the same segregated areas, due to the proximity of an industrial park, for instance, then a correlation between local segregation and the wage gap across groups emerges.

Alternatively, racial discrimination from the employees, potentially sustained by trade unions, may serve as a device for protecting some insider workers' privileges and higher wages in the primary sector. Historically, this was the case in the mining industry in South Africa (Thompson 2001, chap. 4). White miners were collectively organized and had laws passed that gave them a monopoly on well-paid jobs in mines, whereas African miners could have done the same work for a small fraction of their wage. Segregation eases the formation of such collective action by a larger diffusion of information into homogeneous populations. Then, once again, segregation and incomes correlates. Note that in post-Apartheid South Africa, trade unions may instead try to reduce the wage gap, yielding a negative correlation with segregation instead.

If racial discrimination is statistical instead of taste-based, segregation still contributes to segmenting the labor market by limiting the information about minority workers reaching recruiters, then generating wage differences correlated with segregation. A large difference in capital ownership (and/or human capital levels) across groups, as it is the case in South Africa, would strenghten this mechanism.

Segregation may affect income levels by forcing minority workers to live far away from job opportunities (Kain 1968). This "spatial mismatch" raises the search cost with the distance to job opportunities which incites minority to choose lower paying jobs closer to their place, or even stay unemployed. In South Africa, post-Apartheid housing programs have been reinforcing the estrangement of many African workers from job opportunities for at least a decade (Bebbington et al. 2010).

Lastly, some mechanisms might interact together due to segregation. For instance, in South Africa, traditional redistribution within extended families, neighborhoods and kin groups, in African communities, is often such a burden that it may discourage workers to search for well-paid jobs (Mhlongo 2019).

III. Methodology

III.A. *Measuring Segregation*

1. *Segregation Indices*

One often measures segregation as the propensity of individuals to live with similar peers separately from other groups. The most standard measures assume a partition of the city[1] as given and use information on the subdivision of the city's population to compute an index. Massey and Denton (1988) propose considering five dimensions of segregation: evenness, exposure, concentration, centralization, and clustering. In this paper, we focus on evenness and exposure for several reasons. First, they are, by far, the most popular approaches in the segregation literature. Second, the other dimensions appear less specific to the notion of segregation, less politically salient, and may require fine-gridded data, which are typically not available.

Evenness refers to the degree of overlap between the spatial distributions of the two groups. The most common index in the empirical literature on segregation is the Dissimilarity Index, which quantifies the proportion of the minority group that would have to relocate to achieve an equal spatial distribution. Its formula in the case of two groups, say Africans and Whites, for a partition of the city into a set I of locations, is:

$$Dissimilarity = \frac{1}{2}\sum_{i\in I}\left|\frac{White_i}{White_{Population}} - \frac{African_i}{African_{Population}}\right| \quad (1)$$

where $Group_i$ is the number of Group individuals in location i, $Group_{Population}$ is the total number of Group individuals in the population, and the two groups are Africans and Whites.

In contrast, exposure measures the degree of potential contact between the two groups. One widely used measure of exposure is the Isolation index, which measures the probability of

[1] Segregation measures can also consider the country level. We will use the 'city' vocabulary in the remainder of this paper, as residential segregation is typically discussed at the city level.

interacting with a member of the same group. Its formula, for example, for a White individual is:

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

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

Since we cannot observe the local network structure, our approach is pragmatic and relies on the Isolation index. Beyond its attractive axiomatic properties,[2] this choice is motivated by econometric identification assumptions that are discussed below.

III.B. The Contribution of Segregation to an Income Decomposition

1. General Issues

A central aim of this paper is to quantify the relative contribution of segregation to income gap distribution. Oaxaca-Blinder decompositions help quantify additive contributions of variables to the relationship between factors and outcomes. They often suggest explanations by factors or reciprocal links between outcomes and factors. As is typical in decomposition approaches (DiNardo et al. 1996; Sherer 2000), selection or endogeneity issues are not addressed and there is no causal interpretation of the decomposition, in general. The role of decomposition methods is to provide an initial, preliminary examination of the data, perhaps before specifying a causal or a theoretical model that would include factors found with substantial contributions. This descriptive-predictive approach is endorsed, for instance, in the survey of Fortin et al. (2011, pp. 96-97) on decomposition methods.

In a linear setting, the difference in mean outcome Y between two groups, A and B, is usually decomposed as follows:

$$\mathbb{E}[Y_A] - \mathbb{E}[Y_B] = (\mathbb{E}[X_A] - \mathbb{E}[X_B])\beta_A + \mathbb{E}[X_B](\beta_A - \beta_B)$$

where the composition effect, $(\mathbb{E}[X_A] - \mathbb{E}[X_B])\beta_A$, stems from the average difference in the characteristics X between the two groups, and the structure effect, $\mathbb{E}[X_B](\beta_A - \beta_B)$, comes from the difference in the coefficients β between the two groups (Jann 2008; Fortin et al. 2011). In particular, this simple adding-up property is automatically satisfied in the above standard Oaxaca-Blinder decomposition that relies on linear regressions to describe the means of the compared distributions. This is also the case when examining unconditional quantiles with RIF regressions because the last stage of their estimation is a linear regression. Each of the expectations and parameter vectors that appear in the above decomposition must be estimated from some dataset, which may involve usual sampling, estimation, specification, and measurement errors. In that sense, we examine the potential specification error associated with the usual omission of the segregation variable.

More generally, decompositions allow some quantitative assessment of the relative size of the covariates' contributions to the gap between the distributions of two groups. In this paper, we are interested in the contribution of the local segregation variable, while controlling for essential explanatory factors of earnings: the education and experience of the individuals. A

[2] It is notably asymmetric (Massey and Denton 1988) and respects the Independence and School Division properties (Frankel and Volij 2011).

debatable, albeit rather common, interpretation of the structural component is as a measure of the discrimination in the labor market.[3]

2. The Contribution of Segregation

However, if we directly include as a regressor in a linear equation a segregation measure that is specific to the location (such as the Dissimilarity Index), not to the group, it is no longer clear what the composition effect will capture. Indeed, any such fixed regressor at the location level may be highly correlated with local fixed effects. Then, there is a risk that incorrect shares of the gap between the two groups will be attributed for the composition and structure effects. Moreover, it would also amalgamate symmetric situations that describe different contexts. In South Africa, and for analyzing the link between segregation and income, an all-African township is clearly different from an all-White suburb.

As we need an asymmetric measure of segregation,[4] we primarily use the Isolation Index. Since, in our application, segregation is measured using the initial information taken from the 2001 Census and is fixed for all individuals, it is consistent with the idea that segregation may act on income over relatively long run. This allows the measure of segregation to be the same in the two studied periods. We now turn to the data used in the estimations.

III.C. Endogeneity and Selection

1. Endogeneity

In addition to preventing from a causal interpretation of the decomposition, endogeneity also introduces potential bias in the estimation of the model for each group. In our case, endogeneity may arise through reverse causality between segregation and income, or an omitted factor bias if criminality would be correlated to segregation, for instance. However, as we fix the measure of segregation to its level observed in 2001 for both waves (2008 and 2014), this time lag mitigates any potential feedback loop that would be made possible by wage expectations seven or thirteen years later.

Regarding the problem of omitting criminality, as an outside option for legal jobs, it is potentially correlated with income and segregation, especially if segregation affects employment. In the Online Appendix, we provide 2SLS results with segregation instrumented a la Cutler and Glaeser (1997) with a quadratic polynomial of between and within-municipal district rivers. The number of rivers in a municipal district directly inform on the land fragmentation and natural boundaries that could be used to segregate people while it does not affect criminality by itself. We also provide Unconditional Quantile IV regression following Powell (forthcoming). Our main results appear to be robust to endogeneity.

[3] For instance, Sherer (2000, p.319), referring to the structure effect, states that: "*To determine the extent to which earnings differentials represent labour-market discrimination, Blinder-Oaxaca decompositions are constructed using the output from OLS earnings regressions.*".

[4] In the words of Frankel and Volij (2011) (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."*

2. Selection

Labor market selection may affect the results obtained since workers may have specific characteristics that over or underrepresented in our sample. As described in Section II., several mechanisms influence the job search process which suggests that labor market selection and segregation may interact in a very complex way.

On one hand, if segregation reduces employment prospects for the minority group, the ones getting jobs are the most able and most determined individuals. It suggests that the unemployed would have been paid a lower wage if employed, which would have increased the wage gap. Therefore, segregation would have been contributing more to the gap. In this case, we underestimate the effect of segregation.

On the other hand, if segregation improves employment prospects for the minority group, more unskilled workers are employed than what would have been the case if segregation had no effect on employment. Therefore, the wage gap and the effect of segregation are overestimated.

Note that employment prospects and wages may be correlated differently across other dimensions. For example, if segregation facilitates employment only in low-paying alternatives. In this regards, the results in Table IV may as well be interpreted as a measure to how segregation is affected by different selection process. For instance, using acquaintances to get a job for individuals below the national minimum wage results in a more negative effect of segregation on wages, therefore an underestimation of the effect of segregation which we explained by the quality of the informal network.

Modelling explicitly the complex interactions between labor market selection and segregation is beyond the scope of this paper. Nevertheless, in the Online Appendix, we provide a bound approach in the spirit of Chandra (2003) as a robustness check regarding labor market selection. This allows us to remain agnostic about the true selection process at play. Our main results appear to be robust to labor market selection.

IV. Context and Data

IV.A. Segregation in South Africa

There is a long history of racial segregation in South Africa (Thompson 2001). 'Color bar' discriminatory legislation, against Africans and other non-White inhabitants, was in force from the early days of the Union of South Africa. This culminated during the Apartheid period, which was a nationwide social policy of separate development supported by the Afrikaner minority (Thompson 2001, Chap. 5-6; Giliomee 2003). The 1950 Population Registration Act categorized and recorded racial identities on individual identification documents into 'Blacks', 'Whites', 'Coloureds' and 'Indians'.[5] The 1950 Group Area Act allocated separate settlement regions to distinct races. A permit was needed to cross the internal borders of racial regions, which contributed to stabilizing the population composition of each region. Under the 1953 Reservation of Separate Amenities Act, the different races had access to separate hospitals, universities and other public amenities. The 1953 Bantu Education Act introduced separate schools for different races. Over time, several additional laws restricted a citizen's travel within the country. In practice, Africans were often excluded from cities and towns, unless they could justify their presence there with a work permit. Although spatial racial segregation has

[5] We employ these categories, except replacing 'Blacks' with 'Africans' as, in the post-Apartheid South Africa, 'Blacks' refers to all the non-Whites groups together.

clearly declined since the end of the Apartheid in 1994, it remains very common de facto and would not fail to strike any casual observer traveling across the country. The coincidence of these facts, along with the history of discriminatory remuneration practices along ethnic lines, suggests that different work compensation rules, notably with respect to ethnicity, may have been implemented in low- and high-segregation areas.

IV.B. Racial income gaps in South Africa

At the demise of Apartheid in 1994, people's aspirations and expectations turned toward greater economic equality and improvements in their standards of living. However, the following decade was instead characterized by increasing inequality (Leibbrandt et al. 2012), poverty traps (Adato et al. 2006), and anti-poor growth (Özler 2007). South Africa is one the most unequal countries in the world, with especially large racial gaps in living conditions.

Over the 1993-2008 period, aggregate inequality increased (Agüero et al. 2007; Leibbrandt et al. 2010b, 2012). By contrast, Statistics South Africa (2017) notes that while the Gini index modestly declined from 0.72 to 0.68 over the period 2006-2015, it has remained stable since 2009. Most of this increase in aggregate inequality is associated with an increase in within-group inequality (Leibbrandt et al. 2012), especially for Africans (Özler 2007; Leibbrandt et al. 2010b). Despite an initial reduction in within-group inequality after 2006, by 2015, every group had nearly returned to its original level (Statistics South Africa 2017). On the other hand, evidence regarding between-group inequality is rather scarce. Leibbrandt et al. (2010b) find an increase in between-group inequality, whereas Leibbrandt et al. (2012) report a decreasing contribution of between-group inequality to aggregate inequality. However, this is relative to an extreme maximal counterfactual, which does not imply that between-group inequality, in absolute terms, has decreased. Finally, the emergence of an African middle class is a major novelty in the South African society (Statistics South Africa 2017). However, the size of the phenomenon might have been overestimated (Bhorat and Khan 2018).

IV.C. Data

Our data source is the National Income Dynamics Study (NIDS hereafter), which is an individual panel data survey conducted every two years with a nationally representative sample. There are four waves available that cover the period 2008-2014. However, we will use only the 2008 and 2014 waves to avoid the short-term fluctuations due to the 2008-2009 economic crisis[6] that may obscure the contributions of the main regressors in the decomposition. Data on incomes are usually considered relatively reliable (Leibbrandt et al. 2012). But, the sub-sample sizes by racial groups are sometimes relatively small.

Our second source of data is the community profiles from the 2001 South African Census. They provide the total counts of the South African population aggregated at geographic levels ranging from enumeration areas to provinces. The data are exhaustive but only provide summary statistics about the distributions of some socio-demographic characteristics within each location.

The NIDS are used for individual characteristics and income, while the community profiles serve for calculating the measure of segregation in each municipal district, subdivided by sub-

[6]According to a report on poverty levels by Statistics South Africa (2017), *"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."*

places to obtain a more precise sense of local segregation. The lowest geographic sampling level in the NIDS is the municipal district. As we are interested in studying income differences, we restrict our sample to individuals older than 15 who report a positive total personal monthly real income. Amounts are deflated to November 2014 rands with the CPI. Income is measured as the monthly take-home pay from the main job. Other secondary sources of income are excluded to avoid contaminating the analysis with substantial measurement errors, notably from omissions and non-responses. The design of the NIDS explicitly separates self-employed from employees on which we concentrate. As a result, our base sample consists of 2922 Africans and 440 Whites in 2008 and 5291 Africans and 229 Whites in 2014. In the Online Appendix, we provide evidence of robustness regarding more restrictive sample selection criteria addressing seasonal and part-time workers, workers older than the retirement age, and early retirement.

We focus on the African-White gap only, as these are the two most prominent groups in South Africa. Whites occupy the best economic positions and are the most advantaged group, while Africans are the most disadvantaged group and crystalized the fear of the Afrikaner minority during Apartheid. Though also often discriminated against, Coloureds and Asians stand economically between Africans and Whites. Table I reports the mean and standard deviation of the variables used in the analysis, across ethnic groups and survey rounds. As expected, Whites are generally more educated, older, and richer than Africans. They usually have more interactions with the other group, as shown by the statistics on isolation. In the next section, we report the results of the decomposition.

[Table 1 about here.]

V. Mean Analysis

We assume that expected incomes are determined by the individuals' education and experience, possibly quadratically. Then, we augment this specification with a measure of segregation. This will allow for comparisons with the literature and serve as a benchmark for the distributional analysis, in terms of the aggregate pattern of the income gap. As stated above, our measure of segregation is fixed in the year 2001 because we cannot measure segregation from the NIDS and have to rely on a measure constructed from the 2001 Census.[7] However, since segregation might affect income levels with a delay, it does not seem unreasonable to adopt this approach. For example, bad habits may develop over several years before becoming ingrained. Our model takes the following form for each individual i:

$$\begin{aligned} Income_i = \alpha + \beta_1 \times Education_i + \beta_2 \times Education_i^2 + \beta_3 \times Experience_i \\ + \beta_4 \times Experience_i^2 + \beta_5 \times Segregation(2001)_i + \epsilon_i \end{aligned} \tag{3}$$

where $\alpha, \beta_1, \beta_2, \beta_3, \beta_4$, and β_5 are parameters to estimate, and ϵ_i is a centered error term. We first run this OLS regression separately for Africans and Whites. The results are displayed in Table II.

[Table 2 about here.]

[7] We also considered a measure of segregation coming from the 2011 Census (adjusted to the 2001 administrative boundaries) for 2014. It produced very similar results.

Several expected effects emerge. We find a positive and significant effect of experience for both groups with decreasing returns, as the coefficient for its square is negative. However, as experience is a function of age, these coefficient estimates might also capture a life-cycle phenomenon, older people being generally wealthier than their younger counterparts. The effect of isolation on mean income is positive for Whites and negative for Africans, although it loses its significance in 2014 for Whites. Finally, the effect of education is dominated by the quadratic term, which yields an overall positive effect for Whites and for Africans with more than 6 years of schooling, which concerns at least 72 percent of Africans aged 15 or older. This U-shape pattern may be explained by the skills mismatch characterizing the South African labor market.

Then, we decompose the mean, using the pooled sample as the reference group, in Table III, to elicit the magnitudes of the roles of the different correlates, notably segregation.[8] The income gaps between the two groups correspond to the difference between the mean predicted incomes of the two groups obtained via OLS. First, consistent with the high and rising inequality levels observed since the end of Apartheid (Agüero et al. 2007; Leibbrandt et al. 2012), the average real monthly income gap is considerable. It corresponds to 6658 rands in 2008 and rises to 6886 rands in 2014, almost twice the national minimum wage in 2019.

[Table 3 about here.]

The magnitudes of the composition and structure effects are comparable. Despite the emergence of an African middle class, Africans continue to lag behind on many socio-economic characteristics. A sizable and significant composition effect is thus an expected finding. However, the finding that this composition effect is roughly equal to the structure effect is less expected. It implies that Africans with similar socioeconomic characteristics as Whites benefit much less on average than Whites from these characteristics and that this is as important as the differences in socio-economic characteristics. This might be a consequence of racial discrimination in the job or housing markets (Kain 1968). Alternatively, it might reflect different work habits between Africans and Whites or different professions and activity sectors. For instance, if Africans work mainly in rural areas or the industrial sector, having a master's degree might give them access to a lower wage than Whites working in the financial service sector in an urban area. Thus, the racial gap in returns to education might signal a premium for urban areas and/or the financial sector. Over time, the share of the composition effect increases from 46.4 percent in 2008 to 52.7 percent in 2014, while the structure effect decreases from 53.6 percent in 2008 to 47.5 percent in 2014.

When we more closely examine the detailed decomposition, we first note that all the groups of variables contribute positively to the gap through the composition effect. Education is the main contributor to the composition effect, accounting for 95.5 percent of the effect in 2008 and for 73 percent in 2014. This reinforces our discussion above: Africans lag behind Whites in terms of education and experience.[9] This may be partly due to the dual school system inherited from Apartheid. For instance, in 2009, grade three pupils in formerly White schools outperformed grade five pupils in formerly African schools on a standardized test designed for

[8] To avoid transferring part of the structure effect into the composition effect, we add a group dummy to the pooled model for the decomposition (Jann 2008).

[9] As the contributions of each factor sum up to the total effect, the total contribution of education is the sum of the contribution of the education variable and that of education squared. This is also true for experience and the distributional analysis in the next section.

grade three students (Spaull 2013b). This discussion also suggests that segregation may be responsible for parts of the contribution of education through this duality.

The results for segregation present a different pattern. Its mean composition effect is close to zero and not significant in 2008, whereas it is positive and significant in 2014. However, the contribution of segregation in the composition effect increased by more than fivefold between 2008 and 2014, and while, in 2008, it represents around 3 percent of the total composition effect, it accounts for more than 15 percent in 2014, being the second-greatest contributor to the composition effect, after education. According to Statistics South Africa (2017), the 2008-2009 economic crisis, which struck South Africa during the last quarter of 2008, and the following turmoil drove many people into poverty.[10] As the most deprived are usually the most isolated, this might explain the massive increase of the role of segregation between the two periods.

In the structure effect, segregation emerges as the main relative contribution to the total structure effect. Its contribution is of similar order to education in 2008, although education reduces the gap while segregation increases it, and segregation is the only significant contributor in 2014. Education comes second, accounting for 30 percent of the structure effect in 2008, while its effect is not precisely estimated. When interpreting these figures, we should bear in mind that the constant term still represents a large share of the structure effect (17.6 percent in 2008 and 37.7 percent in 2014), hinting at substantial group-specific hidden factors.

VI. Decomposing Income Distributions

VI.A. Distribution Analysis

The main interest in a distribution analysis of the racial income gap is as a device for investigating the heterogeneity of the effects of segregation on this gap. Again, we pursue an agnostic perspective on endogeneity and selectivity phenomena. As a matter of fact, the distribution analysis may provide hints about where in the distribution these issues may matter most.

Then, instead of comparing the distribution means of the two groups and decomposing the mean gap, one can compare the marginal distribution quantiles of the two groups for the same quantile index (for example, for the median). In that case, the composition effect still solely describes the effect of the differences in the characteristics between the two groups, while permitting the comparison for the same given quantile index in the two distributions.

We depart from common approaches by decomposing the income distribution with RIF regressions (Firpo et al. 2009) instead of the reweighting approach (DiNardo et al. 1996). The reweighting approach suffers from path-dependence in the detailed decomposition, which does not sum to the aggregate decomposition. RIF regressions are much simpler and perform better in practice for detailed decompositions.

[10] "it is clear that the [2008 global financial] crisis was particularly tough on those most deprived in our society" [...] "The last five years, notably between 2011 and 2015, have been a rough economic rollercoaster for South Africa" [...] "This period has seen the financial health of South African households decline under the weight of [this] economic [pressure] and, in turn, pulled more households and individuals down into poverty." (Statistics South Africa 2017, p.14 and 16)

VI.B. Decompositions

The estimation results of the detailed Oaxaca-Blinder decompositions applied to the RIF(y, q_τ) dependent variable are presented in Figure 1. In the top-left panel, we first display the quantiles of the racial gap in log income [11] for the two studied years. For both 2008 and 2014, the gap is always significantly different from zero and keeps the same sign for all quantiles. This first-order stochastic dominance result implies that an aggregate utilitarian social welfare of Whites is unambiguously higher than the corresponding figure for Africans, in both years. Finally, 2014 also first-order stochastic dominates 2008, within each racial group, which confirms the unambiguous improvement of each of these two income distributions over the studied period.

[Figure 1 about here.]

However, we observe two distinct patterns. In 2008, the income gap increases from the bottom quantile to the median and decreases thereafter. In 2014, it is relatively stable from the second decile to a little before the sixth decile and then declines as we approach the top of the distribution. More important, at any quantile, the income gap is always smaller, in log points, in 2014 than in 2008, but at a higher real income level than in 2008, which corresponds to an increase in the income gap. For instance, at the first quartile, the income gap of 1.45 log points in 2008 and 1.25 log points in 2014 coincides to gaps of 4026 rands in 2008 and 4390 rands in 2014.

We test the null hypothesis of no differences between quantiles levels in each year with t-tests on each quantile index. Dashed quantiles represent quantiles for which the null hypothesis is not rejected. The decline in log income differences occurs only significantly for intermediate quantiles ranging from the 32nd to the 68th income quantile. Therefore, this slower increase of the income gap for middle classes might be due to the emergence of an African middle class.

We report the aggregate decomposition of the racial gap, for each year and each quantile, in the top-right panel of Figure 1. This decomposition disentangles the differences in observed characteristics from the influences of market and social mechanisms that are captured by differences in the parameters. The dashed parts of the curves represent quantile composition and structure effects that are significantly different from zero at the 5 percent level. In 2008, the structure effect continuously decreases with quantiles, while the composition effect is increasing and plateaus near the 6th decile. At the upper end of the distribution, the structure effect actually contributes to reducing the income gap. In terms of relative size, the structure effect is slightly larger than the composition effect up to the 35th quantile. In 2014, the pattern is similar, although the two elicited components of the income gap are much closer and their change over quantiles much slower. Thus, the magnitude of the composition effect overtakes that of the structure effect beginning at the median, with the latter not contributing at all after the 65th quantile. This suggests that the hidden mechanisms that separate the incomes of Whites and of Africans operate primarily among the lower classes of these groups. This particularity will be exploited below in the analysis of the minimum wage reform.

To complete this description, t-tests are performed to compare the structure and composition effects in 2008 with their respective counterparts in 2014. Then, we examine whether the structure effect is significantly different from the composition effect in 2008 and in 2014.

[11] In the remainder of the paper, we refer to 'income' for simplicity, but it should be understood that we employ the natural logarithm of income.

Regarding the temporal trend, there are no significant variations for the structure effect, except potentially for a small group of quantiles after the median. This stability over time may indicate relatively permanent socio-economic mechanisms, some of which might be linked to segregation. For the composition effect, a notable decline over time is observed from the 33rd quantile. Thus, the relative importance of each effect has changed over time. In 2008, the two effects are significantly different except around the 35th quantile. However, in 2014, both the structure and composition effects contribute equally below the median. Ultimately, the observed reduction in the income gap observed for the middle quantiles appears to be driven primarily by the reduction in the composition effect.

We delve deeper into the relative contribution of each factor to the composition (Figure I, middle panels) and the structure (Figure I, bottom panels) effects. In 2008, experience does not play any role in the composition effect. Education is the most important contributor to the composition effect, followed by segregation, with the former representing twice the latter's contribution across almost the entire distribution.[12] Both are increasing throughout the distribution, except after the 6th decile, after which the contribution of education slightly decreases and that of segregation continues to increase. This parallel pattern explains the increasing contribution of the composition effect across quantiles, and when education decreases, segregation compensates for its reduction to form the plateau observed. In 2014, each contribution is ranked similarly as in 2008, but experience now contributes positively to the income gap from the first quintile, although it remains the smallest contributor. Education's contribution is stable across quantiles up to the median, at which point it begins to rise to its maximum around the third quartile, and declines slightly thereafter. In 2014, segregation's contribution slowly decreases until the median before recovering from its minimum around the third quartile. Then, it plateaus until it spikes dramatically in the very top quantiles. Both the rise and decline of education's contribution from the median and the tremendous spike exhibited by segregation at the very top materialize directly in the aggregate composition effect. Its relative stability in the first half of the distribution comes from the contributions of education and experience compensating for the weakening of the contribution of isolation. As is typical in quantile analyses, substantial variations at extreme quantiles are likely to be statistical artifacts due to the restricted sample sizes used in the calculations for these quantiles.

For the structure effect, the ranking of the contributors differs drastically from that for the composition effect. Segregation is now the dominant factor at almost all quantiles, before experience, followed by education. The intercept parameter is specific to the structure effect and bears a precise interpretation in this context. Usually, in mean regressions, the intercept is viewed as the average income level individuals obtain once the effects of the other covariates have been removed. In quantile regressions, it is instead the minimum income level at the specified quantile regardless of the effect of other covariates. Thus, in the decomposition, a significant difference between two intercepts may suggest intrinsic discrimination between the two groups. However, one cannot infer anything about the origin of this discrimination, whether it is true racial discrimination inherited from Apartheid or something else related to omitted factors. In both 2008 and 2014, the contribution of the intercept is positive and significant for approximately 20 percent of the population above the 6th decile. However, in both years, this positive contribution is systematically compensated for by a negative contribution of the same magnitude from education. The two terms statistically cancel out throughout the

[12] A t-test suggests that the contribution of education is significantly equal to twice that of segregation up to the 84th quantile.

distribution, which is a consequence of the additive normalization of the decomposition and of the insignificance of the contributions of the other factors at these quantiles. Overall, the residual structure effect, after accounting for the intercept gap, is due primarily to segregation and experience. In both years, segregation contributes positively in the lower halves of the distributions and loses significance for the upper halves. Experience follows a similar pattern, except that it contributes negatively to the income gap above the 6th decile in 2008. A plausible explication for these findings is that in a particularly harsh dualistic labor market for Africans, experience in better jobs represents a signal of reliability and skills for Africans, whereas low-productivity Whites might be protected by discrimination. Alternatively, it is possible that affirmative action legislation adds a premium on experienced African workers. However, this effect disappears in 2014.

VI.C. RIF Regressions

To better understand the structure effect, we now examine the estimation results of the RIF regressions used for the above decompositions. The estimation results are displayed in Figure II for each group, each year, and across quantiles. The partial relationship between education and income is identical in 2008 and 2014 for Africans, indicating that the variation in the structure effect is essentially due to changes in the returns to education of Whites. Moreover, the incomes of Africans display little sensitivity to their education level, whereas the incomes of Whites obey a more complex educational pattern, which is particularly pronounced at the top of their distribution. This might reflect greater heterogeneity in bargaining power for highly educated Whites occupying top positions.

[Figure 2 about here.]

Regarding experience, the pattern for Africans is similar in both years across quantiles. It differs only by its level. In 2008, the linear part is slightly higher, while the quadratic term is slightly lower but only negatively significant from the 4th decile. In 2014, the linear part is not positively significant before the 3rd decile, while the quadratic part is negatively significant after the median. Therefore, Africans enjoyed some little linear experience premium in the bottom of the distribution in 2008, while it was destroyed for at least the first quartile by 2014, maybe due to the 2009-2010 economic crisis. At the top of the distribution, the marginal returns to experience are decreasing, but slightly less in 2014 than in 2008. In both years, Whites always experienced a better marginal return to experience, the only exception is the reversal of the linear component of experience at the top of the distribution in 2008, which explains the negative significant contribution to the structure effect.

The most interesting lesson from these RIF regressions concerns the relationship between segregation and income. Segregation is negatively associated with income only for Africans at the bottom of the distribution and in the lower-middle class (up to the median in 2008 and to the 6th decile in 2014) in both years. On the other hand, it is positively associated with income for Whites in 2014 in the upper half of the distribution. It appears to have a positive effect for all Whites in 2008. Hence, the structural effect of segregation is substantial below the median because the gap between the quantile effects of the two groups is at its maximum. Then, it loses significance as the quantile effect for Africans fades away for the upper quantiles. This suggests that the economic mechanisms at work behind the effect of segregation are most likely different for Africans and Whites, and thus, policies addressing this concern for segregation should also differ. In the next section, we exploit the 2018 minimum wage reform to shed some lights on potential explanations of the effect of segregation.

VII. Segregation and the 2018 Minimum Wage Reform

VII.A. Minimum Wage in South Africa

Before January 2019, the legal minimum wage varied across activity sectors, ranging in 2015 from 1813 rands per month for Domestic Workers to 2844 rands per month for Contract Cleaners (Bhorat et al. 2016). In 2015, it implemented to approximately 39 percent of formal employees.

In 2018, the National Minimum Wage Bill was passed (for its enforcement in 2019) with the support of the ANC (African National Congress) members of parliament and opposition from the other parties. The minimum wage was set at 3500 rands per month, for 40 worked hours per week. While there is some doubt about its universal practical implementation, given the limited capacity of the monitoring agency, it is still a major shock on the economic and political system. For comparison, the median salary of workers covered by sector agreements is approximately 2447 rands per month, and 3400 rands per month for all workers in the formal sector (Bhorat et al. 2016). In any case, almost half of the South African labor force should benefit from the reform (47 percent according to the COSATU (Congress of South African Trade Unions)).[13]

VII.B. The Respective Positions of the Political Parties

The three main political parties have taken sharply contrasting positions on segregation, income sharing across racial groups and minimum wage policies. When the minimum wage law was passed, the Democratic Alliance voted against it, as they favor no minimum wage at all, to preserve jobs, or a much lower minimum. Moreover, they propose introducing different minimum wages in different sectors and allowing workers to accept wages below the minimum. In contrast, the Economic Freedom Fighters (EFF), which is the other important opposition party, voted against the law because they wanted a much higher minimum wage, ranging from 4500 to 12,500 rands per month depending on the sector. They even spurned the government's proposal, likening it to a 'slave wage'.[14] The unions also argue for a higher minimum wage, with 4500 rands per month having been proposed by the COSATU, which is fairly aligned with the government but did not sign an agreement on the law, and 12,500 rands per month being proposed by the SAFTU (South African Federation of Trade Unions).

VII.C. How May the Reform Affect the Relationship Between Segregation and Income

We now propose an innovative approach that uses income decomposition estimates to advance the policy debate on the minimum wage reform. To do so, we examine how well the claims of the main political parties about the minimum wage, on the one hand, accord with the intervals of quantiles, for each group, in the graphs of quantile decomposition, on the other hand.

Clearly, some caution must be taken. In particular, if the reform substantially changes the structural data generating processes of incomes in the country, nothing should be deduced from the graphs. However, if one assumes that this is not the case and that, overall, the current

[13] See AllAfrica.com (2018), last accessed on the 14th of November, 2019.

[14] See citizen.co.za (2018), last accessed on the 14th of November, 2019.

relationship of incomes with education, experience and, specifically in our case, local segregation will not be substantially affected by the reform, then one could use the graphs to identify the populations most likely to be affected by the current reform or by the reforms proposed by the opposition parties. This can easily be achieved by examining the corresponding quantile intervals for each group. In that case, one could argue that, post reform, only the part of the curves that exceed the quantiles corresponding to the considered minimum wage should apply. This provides us with a quick and simple graphical diagnostic device.

Our identification strategy can be compared with the first identification assumption in (Chernozhukov et al. 2013, pp. 2236-2237). These authors assume, for the US, that the conditional density of wages below or at the minimum wage depends only on the value of the minimum wage; that the minimum wage has no effect on unemployment; and that there are no spillover effects onto wages above the minimum. While all of these assumptions are debatable, they seem to correspond to a benchmark for minimum wage effects. Our approach can be seen as another simplifying perspective in that it assumes some rigidity of the studied phenomena across quantiles.[15]

Under these tentative diagnostic rules, the government reform, and to a greater extent the reforms proposed by the EFF and the unions, would lead to the elimination of precisely the areas of the curves in which the segregation variable makes a significant contribution to the racial income gap. Although more causal studies would be necessary to confirm them, these results hint at the possibility that the minimum wage reform might eliminate, or at least substantially reduce, the factors that make local segregation contribute to the wage gap between Whites and Africans.

1. Our approach

We now probe potential mechanisms at play regarding the effects of segregation on incomes and how they relate to the programs of the political parties. To do so, we do no longer look at structure and composition effects of segregation, but instead at the 'local' marginal effects of segregation for diverse income subgroups that correspond to natural constituencies of political movements, by referring to minimum wage policies. Indeed, the distinctive political programs of the parties can be regarded as potential treatments. Because of the reduced sample sizes, we mostly focus on Africans.

Individual incomes are categorized into wage intervals according to whether they would be treated under the hypothesis that a given party's minimum wage proposal would be enforced. As we consider several political movements, a specific target group is composed of individuals with wages below the minimum wage proposal for this movement, while above the minimal wage proposals of the other organizations with lower proposals. In that sense, a specific target group can be regarded as an intersection of potential treated and non-treated group definitions. The first specific target group, labelled NMW, gathers all individuals below the National Minimum Wage. Therefore, it can be seen as the ANC constituency, at least as far as the minimum wage policy is concerned, and will be the most important group in our discussions. It is relatively consistent with the ANC's supporters profiles found in a recent survey (citizensurveys.net 2018) that shows that low-income individuals are overrepresented when compared to supporters of

[15] In contrast, simulations based on Computable General Equilibrium Models have been used for assessing the effect of minimum wages in South Africa (Pauw and Leibbrandt 2012). They seem to display negligible impact of these policies on poverty.

the other two main parties. The second group, labelled COSATU, comprises people below the COSATU's proposition of 4500 rands and above the NMW. The third group, labelled EFF, regroups individuals below the EFF's proposition of 12500 rands and above the COSATU's one. The last target group, labelled N.C., are Not Concerned since they earn more than 12500 rands. We found in Figure III that the implicitly targeted individuals by these political movements are also the most affected by segregation among the Africans. This supports our strategy of analysis.

[Figure 3 about here.]

Then, we use the estimated RIF-regression coefficients of segregation as the dependent variable in descriptive regression models to capture the heterogeneity of the effect of segregation in each target group.[16] This is possible because each observed individual corresponds to a specific quantile in the income distribution, and for its specific quantile index, we have a specific estimated RIF-regression coefficient of Isolation. All these individual-specific estimated coefficients can be partitioned in subsamples of interest. The dependent variable is therefore the marginal quantile effect of segregation experienced by an individual. We use robust standard error estimators to allow for the expected correlations of errors for different quantiles.

The set of independent variables in these regressions includes: household size and household income, gender, some labor-related characteristics (affiliation to a trade union, weekly worked hours, reservation wage), some dummies for industrial sectors (agriculture, mining, finance, private households, and community, social, and personal services), the province of residence, and the identifying variables related to theories that we discuss further below.

2. *Socio-economic mechanisms for segregation and constituencies*

First, individuals were asked how they obtained their current job. In particular, they may have accessed it through other household members, which denotes strong network links, or through friends and relatives in another household, which signals weak network links. These informal channels for job information and access have been found to be dominant in developing countries (Ioannides and Loury 2004). This provides us with two proxy variables hinting at the salience of local network effects. The other answers to this question seem to rule out network effects.[17]

Regarding racial preferences, individuals were asked about their desire to leave their current neighborhood. They could state a strong or moderate preference to stay (or leave), or that they were unsure whether to stay or leave. We posit that if this predilection is somewhat correlated with racial residential segregation, this is likely to be related to racial preferences for a particular racial mix in their neighborhood.

As additional evidence that the desire to leave their current neighborhood conveys, at least to some extent, racial preferences, we study, in the Online Appendix, its association with trust in the own racial group and trust in the other racial groups. We also conduct a one-way ANOVA between the level of segregation and the desire to leave their neighborhood. Although the results

[16] We do not account for the dependent variable already being an estimate when estimating the standard errors. Indeed, because of the large degrees of freedom in each of these regressions, the neglected uncertainty coming from this preliminary estimation step is unlikely to change the inference.

[17] However, whenever there was a doubt that an answer might also convey information on networks, like when individuals got their job through a previous employer, we controlled for these variables. However, we only discuss the most parsimonious model as these variants do not change the results.

are not clear cut, they provide some weak evidence that a desire to leave a neighborhood conveys some information about racial preferences.

Then, labor markets are segmented when, in a given sector, the same skill level is associated with a substantial remuneration gap, dividing two segments of workers. In South Africa, segmentation occurred during the Apartheid because jobs were mostly attributed along racial lines in favor of the Whites. Although the remaining racist labor regulations were abolished in 1994, some discriminating practices may have subsisted up to now. However, starting from 1998, segmentation might be attenuated, or diversified, by the Affirmative Action policy that constrains some firms to hire disadvantaged African workers according to some inter-racial equity criteria.

The South African government passed the Employment Equity Act in 1998, which is the core of the South African Affirmative Action policy. This bill aims at achieving "Equity in the workplace by (a) promoting equal opportunity and fair treatment in employment through the elimination of unfair discrimination; and (b) implementing affirmative action measures to redress the disadvantages in employment experienced by designated groups, in order to ensure their equitable representation in all occupational categories and levels in the workforce." (Government Gazette 1998).

During Mbeki's presidency, the Affirmative Action policy was further completed with the Broad-Based Black Economic Empowerment (BBBEE) Act in 2003, later precised by a set of Codes of Good Practice in 2007, which sets the basis for the transfer of physical capital from the Whites to the Blacks (Africans, Coloreds, and Asians). Therefore, if a change should have occured in the segmentation of the labor market, it should have been after the Employment Equity Act.

Therefore, one of our handle on labor market segmentation is whether the current job started before the Affirmative Action Act, or after. We implement this via a dummy variable indicating whether this condition is satisfied, interacted with the years of schooling to account for different skill levels. All in all, workers who started to work just after the Employment Equity Act may face lower market segmentation than those who started before. In the Online Appendix, we explore the relevance of the Employment Equity Act of 1998 as a turning point of the labor market segmentation by using a regression discontinuity design.

Finally, trade unions may leverage better salaries for their adherents irrespective of their skills. If unionism follows racial lines, then it could lead again to a segmentation of the labor market. Therefore, we use the declared unionship status of the individuals to control for this form of segmentation.

Although this is not our object here to disentangle the complex causalities participating in the definition of constituencies and political programs for distinct parties, we believe that this approach provides suggestive information on the potential political processes behind the persistence or the attenuation of the impact of segregation on incomes in South Africa.

3. Results

The estimation results are displayed in Table IV. Each estimated coefficient of an independent variable in the regressions can be interpreted as how this variable affects the mean effect of segregation on income, for a given target group. In the regressions of Table IV, the provincial and sectorial dummies mostly control for some fixed effects, whereas in the classification analysis, they are markers of the specific constituencies of the ANC and the EFF. Since they tell the same story, it is more relevant to discuss them with Table V. The estimated intercept is indicative of the base effect of segregation for each target group when the other included regressors have no effect. The estimated intercepts reflect the general pattern described earlier that Africans at the bottom of the distribution generally suffer from segregation, but contrary to the earlier analysis, they also reveal that rich Africans benefit from it, given that the additional regressors interactions with the segregation effect are controlled for. The other coefficients should be interpreted as depicting how the corresponding independent variables attenuate or reinforce the base effect of segregation on income.

[Table 4 about here.]

Male individuals tend to enjoy an attenuated effect of segregation, while only in the poorest African target groups, which suggests that segregation-based discriminations on race, social class and gender often cumulate. The effect is large, roughly six times greater among those below the national minimum wage than within the COSATU's specific target group. Since choices of places of residence and work may be made at household level, it may be relevant to examine household-level characteristics. A large household size accentuates the effect of segregation for Africans in the poorest group. This is consistent with larger households living farther away from the main employment centers or in deprived areas. In any case, more dependents puts more pressure on financially constrained households: one additional household member implies a 1 percent worsening of the effect of segregation. Household income dampens the effect of segregation in every group, even when segregation is positive. However, in most groups, the effect is very small, which indicates that nonlinearities in the analysis of the level of segregation and incomes are unlikely to matter for this analysis. A higher reservation wage, which should often imply rejection of low-pay jobs, in particular for Africans below the national minimum wage, makes workers less sensitive to the effect of segregation that is partly confined to this type of jobs. However, the effect is limited, with a 1170 rands increase in reservation wage only reducing the effect of segregation by 1 percent for the poorest. For the wealthiest group, the corresponding amount is of 654 rands, and of almost 3000 rands, for the specific target group of the EFF. Finally, omitting instead these household variables does not overly change the qualitative results for the variables describing socio-economic interactions, while some other coefficients lose in significance.

Network effects In the poorest African target group, mobilizing a loosely-connected network (with weak links) reinforces the negative effect of segregation. This may be indicative of a bad-quality network that may only provide information on low-pay jobs. Moreover, the information flowing into this network may be somewhat redundant as similar people, with homogeneous human capital characteristics, are more likely to bring the same information. In that case, more segregation might imply worse quality and fewer job opportunities received through individual networks.

However, the coefficient of the weak links used for job search becomes significantly positive for African middle class workers earning between 4500 and 12 500 rands, which corresponds to EFF's specific constituency. This time, this is a good-quality network that may help workers to access more highly remunerated jobs.

At the top of the distribution, individuals no longer seem to get their jobs through acquaintances but instead relies heavily on their close family. Doing so substantially reinforces the positive effect of segregation, which is consistent with the capture of the best positions by a small elite: having obtained a job through family links increases the positive effect of segregation by 35 percent, even if significant only at the 10 percent level.

Note that this is the separation into the four specific constituencies that allowed us to find significant and somewhat plausible correlations of networks with the segregation-income associations.

Besides, the last column in Table IV, which shows OLS estimates for the full sample, is eloquent about which confusion could arise by pooling all constituencies together. Although these pooled results may sometimes appear to be more significant because of a larger sample size, this is not an important advantage as no new effect arise. In contrast some effects completely vanish with the pooling, such as the few points proxied with the preference to stay.

The heterogeneity of the segregation-income associations, notably for correlations with network variables, emerges much better with the chosen partition of the workers population. The positive significant coefficient of the strong links for the wealthiest and the opposite-sign coefficients of the weak links for the poorest and middle class could not have been found with a pooled specification. Opposite-sign effects would also have been missed for education and reservation wage.

Overall, these results are consistent with the literature. Referral effects depend on the social network built mostly through social interactions within the neighborhood (Bayer et al. 2008). They primarily affect the employment probability while the quality of that network will have an impact on wages. Loury (2006) shows that individuals with few alternatives rely more on their informal network and end choosing alternatives paying lower wages compared to other search methods. In South Africa, Magruder (2010) finds that, in the Cape Metropolitan area, having a present and employed father increases the employment rate of their sons by one-third, on average, but 55 percent of Africans and Coloureds have absent, unemployed, or deceased fathers. Therefore, segregation, by concentrating single parenthood in specific neighborhoods (Crane 1991) might be responsible for their limited job choices. The positive effect of strong links for the richest group is also consistent with this mechanism. Finally, Adato et al. (2006) also provide evidence, in the KwaZulu-Natal, of positive network effects for non-poor households' labor market outcomes, whereas poor households, at best, experience no effect.

Racial preferences The main significant effect of racial preferences in Table IV concerns the wealthiest African target group: a moderate inclination for not moving away reinforces their beneficial effect of segregation quite strongly. It is ten times greater than the effect of an additional rand of household income. These well-off individuals may want to stay next to their relatives as they have often obtained their job thanks to their family links. In the poorest group, individuals with a strong preference for leaving suffer more from segregation, while significantly only at the 10 percent level.

Labor market segmentation As mentioned before, education, interacted with the Affirmative Action dummy, alleviates the effect of segregation for the poorest and the richest African target groups. Education tends to attenuate the negative effect of segregation for the ANC constituency, and its positive effect for the best-off Africans, relatively more inclined to vote for the Democratic Alliance. However, no significant differences in segregative effects can be found between a job starting before or after the Affirmative Action legislation was passed. Therefore, it seems that Affirmative Action did not significantly reshape the labor market segmentation in terms of mean effects. Nonetheless, in the Online Appendix, we provide evidence that Affirmative Action has increased the polarization of the returns to education instead. This change in the labor market may be related to the reduction of stereotypes on African workers' productivity. By forcing firms to engage more African workers, employers learn more about their productivity, which reduces the noise of their belief, therefore remunerating them with wages closer to their true productivity. The negative effect of education for the richest group suggests that the better-off Africans exploited this lack of information to their advantage. This is consistent with individuals getting their jobs through their family links and explanations related to cronyism.

Being unionized attenuates the effect of segregation on incomes, both for the poorest and for the richest. Only the EFF group is not affected by unionism, although the effect is only significant at the 10 percent level for the richest group. In the two poorest groups, trade unions seems to play their role and leverage better wages for their members. The negative effect observed for the richest group might be a side effect of the defense of the poorest ones, as the gains obtained by the unions may be at the expense of the better-off workers.

4. Sorted effects

In Table V, we describe the main winners and losers from segregation: the "winners" (respectively "losers") are defined as being the 10 percent Africans most positively (respectively negatively) affected by segregation. Following Chernozhukov et al. (2018), we explore in this way the heterogeneity of the partial association of segregation and incomes by scanning the characteristics of the winners and losers from segregation.

[Table 5 about here.]

The main African losers from segregation are mostly little educated workers, often female, living in large households with few pecuniary resources. They mostly obtained their job, often in the agriculture and mining sectors, through their extended network, and are ready to work for a wage much below the National Minimum Wage proposed by the ANC. This might be because with a little more than 8 years of education on average, they are merely completing mandatory schooling.[18] They are concentrated in KwaZulu-Natal, Gauteng, Eastern Cape, and Limpopo.

On the opposite, the main African winners from segregation are workers, mostly male, with a household income more than twice the National Minimum Wage. They were generally recruited through their extended network and are unionized. On average, they work more hours than the main losers, and a larger fraction have obtained their job before the Affirmative Action bill. They are also more educated, which may explain why they ask for higher salaries as reflected by their reservation wages. However, they still have fewer than 12 years of education on average which would have earned them a matriculation diploma. Therefore, they are probably as likely as the main losers to quit schooling without any diploma. Contrary to Cutler et al. (2008), this rules out education as an important dimension of heterogeneity. Finally, African winners are mostly living in Gauteng, KwaZulu-Natal, and Mpumalanga, but are only overrepresented in Gauteng and Mpumalanga. They work mainly in Social services, Finance, Manufacturing, and Wholesale and Retail trade.

These results provide another perspective on the above-discussed mechanisms of the association of segregation and incomes. Whether this can be related to racial preferences or not, both

[18] Schooling is mandatory in South Africa from age six turning seven to the age of 15 or completion of grade 9, whichever comes first.

the main losers and winners are reluctant to leave their residential area. One may think that they often adhere to the segregation context that they face. Having found their job through their extended network is a preponderant trait of the losers. This reinforce our suspicions about the negative influence of within-group networks backed up by local segregation, at least for some disadvantaged workers. Finally, there is some weak evidence of the role of labor market segmentation, noticeable through the prominent effect of the Affirmative Action, in providing higher earnings for the main winners of local segregation.

The regional and industrial differences between losers and winners fit well the electoral constituency of the ANC and the political strategy of the EFF. The traditional ANC strongholds are the northern provinces of Limpopo, Mpumalanga, and North West, and the Eastern Cape in the south. Gauteng, the Western Cape, the Northern Cape, and, to a lesser extent, the Free State have always been disputed with the Democratic Alliance. KwaZulu-Natal is currently administered by the ANC, while challenged by the IFP. The mostly rural Eastern Cape, Limpopo, and KwaZulu-Natal, simultaneously with the agricultural sector partially correspond to the ANC specific target group since the ANC largely relied on the rural vote from its inception.[19] Moreover, workers in the agricultural sector are often not unionized which coincides with the low unionization rate of the main losers.[20] Workers in the mining sector, on the other hand, are close to the COSATU, of which the National Union of Mineworkers is a member, founded by the currently sitting president Cyril Ramaphosa in the 1980s. The COSATU has always been supportive of the ANC.

The EFF's strategy is often to outbid the ANC and the trade unions. However, their minimum wage proposal much exceeds realistic earnings for the poorest workers, especially in the agricultural and mining sectors. If voters choose their champion according to their expected gain, then the EFF's proposal of 12500 rands should speak more to individuals working in the sectors of manufacturing, wholesale and retail trade, and construction, hence, mostly urban workers within the EFF's constituency.[21] Indeed, most other political movements made minimal wage proposals too low to affect many of these workers directly. On the whole, although there are certainly many other determinants of political programs, it is intriguing to note that the ANC-led reform is liable to benefit the ANC's constituency in several ways. On the other hand, it should raise the wages of the workers employed in the modern and public sectors, from income categories that predominantly vote for the ANC. On the other hand, by removing wage intervals corresponding to especially noxious associations of segregation and wages, it should contribute to protect some of the main African losers of segregation. Interpretations of the estimates of the minimum wage proposal for the other parties would still be more tentative, as their proposals were not implemented and never had the chance to go through parliaments.

VIII. Conclusion

In this paper, we proposed a new approach to analyzing the contribution of segregation to socio-ethnic income gaps. These new methods highlighting the contribution of segregation to the income distributions of racial groups provide not only indications of the importance

[19] See Afrobarometer.org (2015) , last accessed on the 14th of November 2019.

[20] Thompson (2001) describes the early formation of trade unions in South Africa consecutive to the rise in the cost of urban living. In 1945, 40% of the unionized workers were employed in commerce and manufacturing, and "the crucial terrain for labor relation was, as ever, the mining industries." (p.179). Unionization in agriculture, far from the urban centers, and being heavily mechanized or of the subsistence type, cannot easily develop.

[21] See citizensurveys.net (2018), last accessed on the 14th of November, 2019.

of integrating segregation in earnings models, but also generate hints at the socio-economic mechanisms that may explain income differences between socio-ethnic groups.

Using generalizations of the Oaxaca-Blinder method to distribution decomposition and scrutinizing the heterogeneity of segregation effects allow us to uncover patterns that remain hidden in mean analyses and traditional Mincer equations. One main finding is the essential contribution of segregation to both composition and structure effects of decompositions of income distributions in South Africa. Segregation is found to be the main contributor to the structure effect, ahead of education and experience. This is notably interesting because major socio-economic mechanisms determining incomes may be hidden behind the heterogeneity of these effects. In this regard, local informal job-search networks, often operating within racial groups and stimulated in segregated contexts, are found to affect the association segregation-income in a way that harms the poorest and benefit African middle classes. In contrast, labor market segmentation affects more the polarization of incomes than their mean levels. Finally, only minor influence of racial preferences has been detected, although this may be due to lack of precise measure for these preferences.

Segregation is also a sizeable contributor to the composition effect. This feature, grounded on observational characteristics could be exploited more for policy design. For example, areas with high segregation levels can be improved with mixed residential real estate investment. Moreover, segregation is found to negatively affect incomes at the bottom of the African distribution but to positively affect Whites at the top of their distribution. Finally, subpopulations are identified that suffer and benefit most from segregation through their incomes, including for the former, little educated workers in agriculture and mining, often female, and immersed in local informal job-search networks.

These findings are useful because knowing which part of the distribution is affected by diverse factors can help the government to design more effective public policies. We illustrated this in the case of the minimum wage reform, which appears to be a potential means of attenuating some noxious effects of segregation on poor Africans, beyond its direct effect on their incomes. Another avenue of policy design could be based on exploiting or mitigating the within-group local network effects that seem to be at work. However, an important lesson acquired is that local networks fostered by local segregation can be noxious or beneficial. Better identifying and understanding these diverse cases seem to be preliminary to such policy design.

Moreover, the findings are useful for guiding economic investigations. We emphasized explanations based on job-search networks, while other explanations could also be considered. Such developments could use explicit models of the non-market mechanisms anchored on some found covariates of segregation effects, such as individual preferences for segregation, labor market segmentation, and diverse types of neighborhood and peer effects.

However, the proposed analyses are not without limitations. These issues would extend to the possible endogeneity of segregation and migration to specific neighborhoods. Addressing these issues in a more direct and sophisticated explanatory setting and, notably, providing causal evidence of the precise economic mechanisms is, therefore, an important challenge for future research.

Perhaps, it would be possible to exploit a shock on residential segregation that may change the balance of the education and human capital accumulation in the country. For example, the current segregation in South Africa is likely to sustain the duality of the school system. In reverse, a shock on education may directly affect segregation, especially if racial preferences can change with education levels.

EconomiX, Université Paris Nanterre
Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, IRD, AMSE

References

Adato, Michelle, Michael R. Carter, and Julian May, "Exploring Poverty Traps and Social Exclusion in South Africa Using Qualitative and Quantitative Data," *Journal of Development Studies*, 42 (2006), 226–247.

Afrobarometer, (https://afrobarometer.org/fr/publications/ad65-south-africans-increasingly-dissatisfied-their-elected-leaders-performance), 25 (2015), last accessed on the 14th of November, 2019.

Agüero, Jorge, Michael R. Carter, and Julian May, "Poverty and Inequality in the First Decade of South Africa's Democracy: What can be Learnt from Panel Data from KwaZulu-Natal?" *Journal of African Economies*, 16 (2007), 782–812.

Alesina, Alberto, Reza Baqir, and William Easterly, "Public Goods and Ethnic Divisions," *Quarterly Journal of Economics*, 114 (1999), 1243–1284.

AllAfrica, (https://allafrica.com/stories/201804250851.html), (2018), last accessed on the 14th of November, 2019.

Altonji, Joseph, and Rebecca Blank, "Race and Gender in the Labor Market," in *Handbook of Labor Economics*, Orley Ashenfelter and David Card, eds. (Elsevier Science, 1999).

Bayer, Patrick, and Kerwin Kofi Charles, "Divergent Paths: A New Perspective on Earnings Differences between Black and White Men Since 1940," *Quarterly Journal of Economics*, 133 (2018), 1459–1501.

Bayer, Patrick, 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 (2008), 1150–1196.

Bebbington, Anthony J., D. Mitlin, J. Mogaladi, M. Scurrah, et al., "Decentring Poverty, Reworking Government: Social Movements and States in the Government of Poverty," *Journal of Development Studies*, 46 (2010), 1304–1326.

Becker, Gary, *The Economics of Discrimination*, 2nd ed. (Chicago: University of Chicago Press, 1957).

Bénabou, Roland, "Workings of a City: Location, Education, and Production," *Quarterly Journal of Economics*, 108 (1993), 619–652.

Bhorat, Haroon, Tara Caetano, Benjamin Jourdan, Ravi Kanbur, et al., "Investigating the Feasibility of a National Minimum Wage for South Africa," DPRU Working Paper Series, WP201601, 2016.

Bhorat, Haroon, and Safia Khan, "Structural Change and Patterns of Inequality in the South African Labour Market," DPRU Working Paper Series, WP201801, 2018.

Blumer, Herbert, "Race Prejudice as a Sense of Group Position," *The Pacific Sociological Review*, 1 (1958), 3–7.

Chandra, Amitabh, "Is the Convergence in the Racial Wage Gap Illusory?," NBER Working Paper Series, WP9476, 2003.

Chernozhukov, Victor, Iván Fernández-Val, and Blaise Melly, "Inference on Counterfactual Distributions," *Econometrica*, 81 (2013), 2205–2268.

Chernozhukov, Victor, Iván Fernández-Val, and Ye Luo, "The Sorted Effects Method: Discovering Heterogeneous Effects Beyond their Averages," *Econometrica*, 86 (2018), 1911–1938.

Chetty, Raj, 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 (2016), 855–902.

Chetty, Raj, Nathaniel Hendren, Maggie R. Jones, and Sonya R. Porter, "Race and Economic Opportunity in the United States: An Intergenerational Perspective," *Quarterly Journal of Economics*, 135 (2020), 711–783.

Citizen Surveys, (https://citizensurveys.net/citizen-surveys-how-many-white-eff-supporters- (2018), last accessed on the 14th of November, 2019.

Crane, Jonathan, "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing," *American Journal of Sociology*, 96 (1991), 1226–1259.

Cutler, David M., and Edward L. Glaeser, "Are Ghettos Good or Bad?," *Quarterly Journal of Economics*, 112 (1997), 827–872.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor, "When Are Ghettos Good or Bad? Lessons from Immigrant Segregation in the United States," *Journal of Urban Economics*, 63 (2008), 759–774.

Dickens, William T., and Kevin Lang, "A Test of Dual Labor Market Theory," *American Economic Review*, 75 (1985), 792–805.

DiNardo, John, Nicole M. Fortin, and Thomas Lemieux, "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64 (1996), 1001–1044.

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

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

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux, "Unconditional Quantile Regressions," *Econometrica*, 77 (2009), 953–973.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo, "Decomposition Methods in Economics," in *Handbook of Labor Economics*, David Card and Orley Ashenfelter, eds., vol. 4, chapter 1 (Amsterdam: Elsevier, 2011), 1–102.

Frankel, David M., and Oscar Volij, "Measuring School Segregation," *Journal of Economic Theory*, 146 (2011), 1–38.

Giliomee, Hermann, "The Making of the Apartheid Plan, 1929-1948," *Journal of Southern African Studies*, 29 (2003), 373–392.

Gradín, Carlos, "Race, Poverty and Deprivation in South Africa," *Journal of African Economies*, 22 (2012), 187–238.

——, "Race and Income Distribution: Evidence from the USA, Brazil and South Africa," *Review of Development Economics*, 18 (2014), 73–92.

Ioannides, Yannis M., and Linda Datcher Loury, "Job Information Networks, Neighborhood Effects, and Inequality," *Journal of Economic Literature*, 42 (4), 1056–1093.

Jann, Ben, "The Blinder-Oaxaca Decomposition for Linear Regression Models," *The Stata Journal*, 8 (2008), 453–479.

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

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

Leibbrandt, Murray, Ingrid Woolard, Arden Finn, and Jonathan Argent, "Trends in South African Income Distribution and Poverty Since the Fall of Apartheid," OECD Social, Employment and Migration Working Papers, 101, 2010b.

Leibbrandt, Murray, Arden Finn, and Ingrid Woolard, "Describing and Decomposing Post-Apartheid Income Inequality in South Africa," *Development Southern Africa*, 29 (2012), 19–34.

Loury, Linda Datcher, "Some Contacts Are More Equal than Others: Informal Networks, Job Tenure, and Wages," *Journal of Labor Economics*, 24 (2006), 299–318.

Magnac, Thierry, "Segmented or Competitive Labor Markets?" *Econometrica*, 59 (1991), 165–187.

Magruder, Jeremy R., "Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa," *American Economic Journal: Applied Economics*, 2 (2010), 62–85.

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

Mele, Angelo, "Poisson Indices of Segregation," *Regional Science and Urban Economics*, 43 (2013), 65–85.

Mhlongo, Niq, *Black Tax: Burden or Ubuntu* (Jonathan Ball Publishers SA, 2019).

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

Özler, Berk, "Not Separate, Not Equal: Poverty and Inequality in Post-Apartheid South Africa," *Economic Development and Cultural Change*, 55 (2007), 487–529.

Pauw, Karl, and Murray Leibbrandt, "Minimum Wages and Household Poverty: General Equilibrium Macro-Micro Simulations for South Africa," *World Development*, 40 (2012), 771–783.

Powell, David, "Quantile Treatment Effects in the Presence of Covariates," *Review of Economics and Statistics*, forthcoming.

Republic of South Africa Government Gazette, (https://www.gov.za/sites/default/files/gcis_document/201409/a55-980.pdf), No. 19370 (1998), last accessed on the 28th of October, 2019.

Schelling, Thomas C., "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1 (1971), 143–186.

Sherer, George, "Intergroup Economic Inequality in South Africa: The Post-Apartheid Era," *American Economic Review Papers and Proceedings*, 90 (2000), 317–321.

Spaull, Nicholas, "South Africa's Education Crisis: The Quality of Education in South Africa 1994–2011," Technical Report, 2013.

Statistics South Africa, "Poverty Trends in South Africa: An Examination of Absolute Poverty Between 2006 and 2015," Technical Report, Johannesburg, 2017.

The Citizen, (https://citizen.co.za/news/1941755/eff-calls-new-minimum-wage-bill-legitimis (2018), last accessed on the 14th of November, 2019.

Thompson, Leonard, *A History of South Africa*, 3rd ed. (New Haven, CT: Yale University Press, 2001).

Wilson, William Julius, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policies* (Chicago: University of Chicago Press, 1987).

Yinger, John, "Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act," *American Economic Review*, 76 (1986), 881–896.

TABLE I

Descriptive Statistics

	2008				2014			
	Whites		Africans		Whites		Africans	
	Mean	Std D.	Mean	Std D.	Mean	Std D.	Mean	Std D.
Income	12909.28	13268.02	3534.29	5777.53	14012.4	13514.05	4542.55	5065.90
Isolation	0.65	0.12	0.93	0.09	0.64	0.12	0.93	0.08
Dissimilarity	0.84	0.08	0.85	0.07	0.85	0.07	0.86	0.07
Years of schooling	12.82	2.03	8.92	4.36	13.28	1.83	10.55	3.66
Experience	21.48	11.69	22.68	13.23	24.03	13.07	20.19	12.73
Male	0.51	0.50	0.57	0.50	0.57	0.50	0.50	0.50
Age	40.31	11.68	37.60	11.12	43.31	13.23	36.74	11.01
Weak link	0.28	0.45	0.35	0.48	0.38	0.49	0.45	0.50
Strong link	0.06	0.24	0.07	0.25	0.04	0.19	0.04	0.20
Union membership	0.20	0.40	0.26	0.44	0.23	0.42	0.27	0.44
Household size	3.08	1.25	4.16	3.11	3.25	1.60	4.48	3.22
Household income	23931.78	20486.24	5868.31	8556.82	27279.05	26478.98	7411.73	9150.55
Hours worked weekly	42.82	13.64	36.36	20.41	41.56	14.70	41.59	15.89
Reservation wage	.	.	.	.	12580.69	14510.48	4774.53	5633.57
Firm size (50+)	.	.	.	.	0.30	0.46	0.25	0.43
Western Cape	0.42	0.49	0.06	0.23	0.37	0.48	0.05	0.22
Eastern Cape	0.03	0.18	0.09	0.29	0.07	0.25	0.08	0.28
Northern Cape	0.06	0.24	0.05	0.22	0.07	0.26	0.04	0.20
Free State	0.00	0.07	0.09	0.28	0.01	0.09	0.08	0.27
KwaZulu-Natal	0.05	0.22	0.26	0.44	0.09	0.29	0.27	0.45
North West	0.02	0.15	0.09	0.28	0.01	0.11	0.08	0.27
Gauteng	0.27	0.44	0.20	0.40	0.30	0.46	0.22	0.41
Mpumalanga	0.12	0.32	0.10	0.30	0.06	0.23	0.10	0.30
Limpopo	0.03	0.16	0.07	0.25	0.03	0.16	0.08	0.27
Agriculture	0.05	0.22	0.16	0.37	0.05	0.22	0.09	0.28
Mining	0.03	0.17	0.07	0.25	0.03	0.16	0.04	0.21
Manufacturing	0.15	0.35	0.17	0.38	0.14	0.34	0.11	0.32
Energy	0.01	0.10	0.01	0.09	0.02	0.12	0.01	0.11
Construction	0.03	0.17	0.07	0.25	0.05	0.21	0.08	0.27
Trade; Hotels	0.18	0.39	0.14	0.35	0.18	0.39	0.19	0.39
Transport	0.04	0.20	0.04	0.20	0.05	0.21	0.05	0.22
Finance; Real estate	0.17	0.38	0.08	0.26	0.21	0.41	0.10	0.31
Social services	0.34	0.47	0.26	0.44	0.29	0.45	0.31	0.46
Observations	440		2922		229		5291	

The variable Income, Household income, and Reservation wage are deflated to November 2014 rands. The dummy variables Male, Weak link, Strong link, Union membership, Firm size, and sectoral and provincial dummy variables are expressed as a share of the population. Firm size and reservation wage were only available in 2014.

TABLE II
OLS Regressions

	2008		2014	
	Whites	Africans	Whites	Africans
Experience	0.08***	0.04***	0.03**	0.02***
	(6.84)	(10.82)	(2.47)	(7.46)
Experience squared	-0.0014***	-0.0004***	-0.0003	-0.0002***
	(-5.87)	(-5.99)	(-1.26)	(-2.90)
Years of schooling	-0.17	-0.05***	-0.31	-0.06***
	(-1.53)	(-4.03)	(-1.13)	(-5.04)
Years of schooling squared	0.01***	0.01***	0.02*	0.01***
	(3.28)	(16.85)	(1.87)	(17.32)
Isolation	1.08***	-0.40**	0.55	-0.46***
	(3.70)	(-2.57)	(1.37)	(-3.22)
Constant	7.32***	6.66***	8.83***	7.29***
	(10.07)	(42.09)	(5.00)	(47.91)
Observations	440	2922	229	5291
R^2	0.355	0.346	0.279	0.259

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

TABLE III
OAXACA DECOMPOSITIONS

	2008	2014
Differential		
Prediction_1	9.09***	9.20***
	(214.76)	(161.92)
Prediction_2	7.70***	8.01***
	(455.18)	(626.79)
Difference	1.39***	1.20***
	(30.56)	(20.55)
Composition		
Experience	-0.01	0.06***
	(-0.57)	(4.13)
Education	0.64***	0.46***
	(20.78)	(14.81)
Isolation	0.02	0.11***
	(0.47)	(3.22)
Total	0.65***	0.63***
	(12.75)	(12.82)
Structure		
Experience	0.16	0.11
	(1.34)	(0.91)
Education	-1.13*	-1.76
	(-1.93)	(-0.99)
Isolation	1.05***	0.68**
	(4.80)	(2.53)
Constant	0.66	1.54
	(1.02)	(0.85)
Total	0.75***	0.57***
	(12.69)	(9.02)
Observations	3362	5520

t-statistics in parentheses.
Prediction_1 is the mean of the predicted logarithm of the real monthly income of Whites (in 2014 (Nov.) rands).
Prediction_2 is the mean of the predicted logarithm of the real monthly income of Africans (in 2014 (Nov.) rands).
These predictions also correspond to each subgroup unweighted sample mean when the coefficients are estimated using OLS for each subsample.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

OLS regressions of the RIF-regression coefficients of Isolation

	(1) NMW	(2) COSATU	(3) EFF	(4) N.C.	(5) Pooled sample
Male	0.064***	0.010**	-0.0026	-0.040	0.18***
	(3.48)	(2.20)	(-0.42)	(-1.06)	(10.55)
Household size	-0.011***	-0.00085	0.00026	0.011	-0.0077***
	(-3.99)	(-1.09)	(0.27)	(1.45)	(-2.75)
Household income (thousands ZAR)	0.018***	0.0011**	-0.00051*	-0.016***	0.0049***
	(6.06)	(2.20)	(-1.74)	(-8.39)	(3.28)
Hours worked weekly	0.0014**	0.000019	0.0000056	-0.0018	0.0039***
	(2.16)	(0.10)	(0.02)	(-0.93)	(5.78)
Reservation wage (thousands ZAR)	0.0095***	-0.00094	-0.00085*	-0.0093***	0.0034*
	(3.63)	(-1.37)	(-1.78)	(-4.59)	(1.81)
Network effects:					
Strong link	-0.052	0.0093	0.0027	0.19*	-0.099**
	(-1.29)	(0.72)	(0.17)	(1.78)	(-2.33)
Weak link	-0.050***	0.0054	0.012**	-0.0089	-0.10***
	(-2.86)	(1.25)	(2.06)	(-0.17)	(-5.89)
Labor market segmentation:					
Union	0.072***	0.011**	-0.0019	-0.076*	0.26***
	(2.59)	(2.14)	(-0.33)	(-1.70)	(13.25)
Before Affirmative Action	0.016***	0.000063	-0.00042	-0.037***	0.030***
	(2.92)	(0.04)	(-0.34)	(-2.97)	(8.52)
After Affirmative Action	0.011***	0.00076	-0.00032	-0.033***	0.025***
	(4.52)	(0.99)	(-0.31)	(-2.77)	(9.83)
Racial preferences:					
Strong preference to stay	-0.016	-0.00098	0.00019	0.090	0.018
	(-0.62)	(-0.15)	(0.02)	(1.47)	(0.71)
Moderate preference to stay	-0.0088	-0.0077	-0.0053	0.16**	-0.0036
	(-0.29)	(-1.00)	(-0.51)	(2.27)	(-0.12)
Moderate preference to leave	0.018	-0.00045	0.0011	0.12	0.051
	(0.45)	(-0.05)	(0.08)	(1.47)	(1.34)
Strong preference to leave	-0.074*	0.0031	0.021	0.070	-0.010
	(-1.75)	(0.28)	(1.61)	(0.84)	(-0.25)
Intercept	-1.10***	-0.14***	0.28***	0.60***	-0.96***
	(-18.68)	(-8.39)	(11.36)	(2.59)	(-16.16)
Province dummies	Yes	Yes	Yes	Yes	Yes
Sector dummies	Yes	Yes	Yes	Yes	Yes
N	2228	502	1293	420	4443
R-square	0.14	0.076	0.021	0.39	0.30
Adj. R-square	0.13	0.024	-0.00035	0.34	0.29
F	13.2	1.68	1.02	286.15†	79.0
P-value	<0.00005	0.018	0.44	<0.00005†	<0.00005

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

NMW: National Minimum Wage; COSATU: trade unions; EFF: Economic Freedom Fighters; N.C.: Not Concerned. *Before Affirmative Action* refers to the interaction of a dummy variable, taking the value 1 if the individual started his job before the passing of the Employment Equity Act of 1998, with the number of years of schooling. *After Affirmative Action* refers to the same interaction with the dummy variable taking the value 0 instead. *Province dummies* comprise a list of dummy variables for all the South African provinces (Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, North West, Northern Cape, Western Cape) except Mpumalanga which serves as the reference basis. *Sector dummies* comprise a list of dummy variables for specific South African industrial sector (Private households, Agriculture, Mining and quarrying, Finance, and Community, social and personal services), the excluded sectors (Manufacturing, Energy, Construction, Retail trade, and Logistics) defining the reference basis.

† Due to a clustering of observations when using subsamples, the sandwich estimator of the robust covariance matrix could not always be computed. We use instead the boostrap estimator obtained with the 554 replications over 1000 that could be used for such computations. Therefore, instead of the F-statistics, we report a Wald statistics.

TABLE V

Classification Analysis - Difference in the Average Characteristics of the African main winners and losers from Segregation

	Main losers		Main winners		Difference		
	Mean	S.E.	Mean	S.E.	Mean	S.E.	P-value
Isolation	.942	.003	.929	.003	.013	.004	.001
Male	.3	.02	.64	.02	-.34	.03	< 0.001
Household size	4.95	.14	4.2	.14	.75	.19	< 0.001
Household income	2635	136.77	8693.49	320.81	-6058.49	347.09	< 0.001
Strong link	.07	.01	.02	.01	.04	.01	.0011
Weak link	.65	.02	.39	.02	.26	.03	< 0.001
Union	.05	.01	.37	.02	-.32	.02	< 0.001
Hours worked weekly	38.09	.67	43.91	.49	-5.82	.84	< 0.001
Reservation wage	2247.69	107.77	5583.54	185.32	-3335.85	209.68	< 0.001
Before Affirmative Action	.03	.01	.12	.01	-.09	.02	< 0.001
Years of Schooling	8.32	.17	11.35	.13	-3.03	.21	< 0.001
Moderate preference to stay	.62	.02	.58	.02	.05	.03	.1511
Unsure	.17	.02	.14	.02	.03	.02	.2094
Moderate preference to leave	.1	.01	.13	.02	-.03	.02	.1275
Strong preference to leave	.04	.01	.07	.01	-.03	.01	.0767
Western Cape	.04	.01	.06	.01	-.03	.01	.0465
Eastern Cape	.13	.01	.06	.01	.08	.02	< 0.001
Northen Cape	.02	.01	.05	.01	-.03	.01	.0053
Free State	.09	.01	.08	.01	.01	.02	.6597
KwaZulu-Natal	.32	.02	.2	.02	.11	.03	< 0.001
North West	.08	.01	.07	.01	.01	.02	.646
Gauteng	.15	.02	.28	.02	-.13	.02	< 0.001
Mpumalanga	.06	.01	.13	.01	-.08	.02	< 0.001
Limpopo	.12	.01	.06	.01	.06	.02	< 0.001
Agriculture	.29	.02	.02	.01	.27	.02	< 0.001
Mining and Quarrying	.12	.01	.02	.01	.1	.02	< 0.001
Manufacturing	.01	.004	.09	.01	-.08	.01	< 0.001
Energy supply	.084	.012	.11	.01	-.02	.02	.2479
Construction	.002	.002	.02	.01	-.01	.01	.023
Wholesale and Retail trade	.05	.01	.09	.01	-.04	.02	.0213
Transport	.14	.02	.16	.02	-.02	.02	.4625
Finance	.02	.01	.07	.01	-.06	.01	< 0.001
Social services	.03	.01	.12	.02	-.1	.02	< 0.001

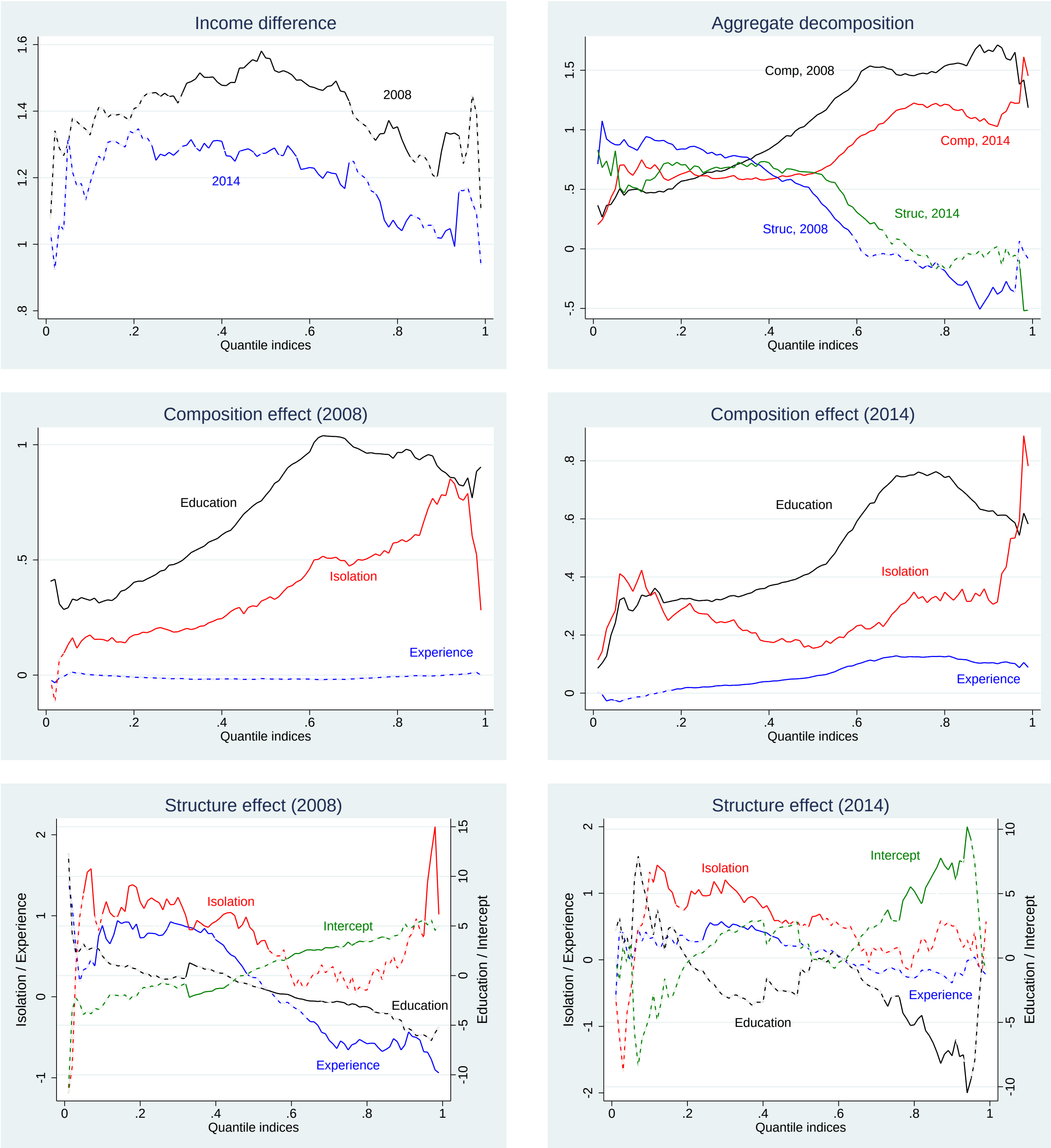

FIGURE I

Aggregate and Detailed Decompositions in 2008 and 2014

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level, except for the income differences, where they represent the quantiles for which the income difference in 2008 is not significantly different from the income difference in 2014. "Comp" and "Struc" refer to the composition effect and the structure effect, respectively.

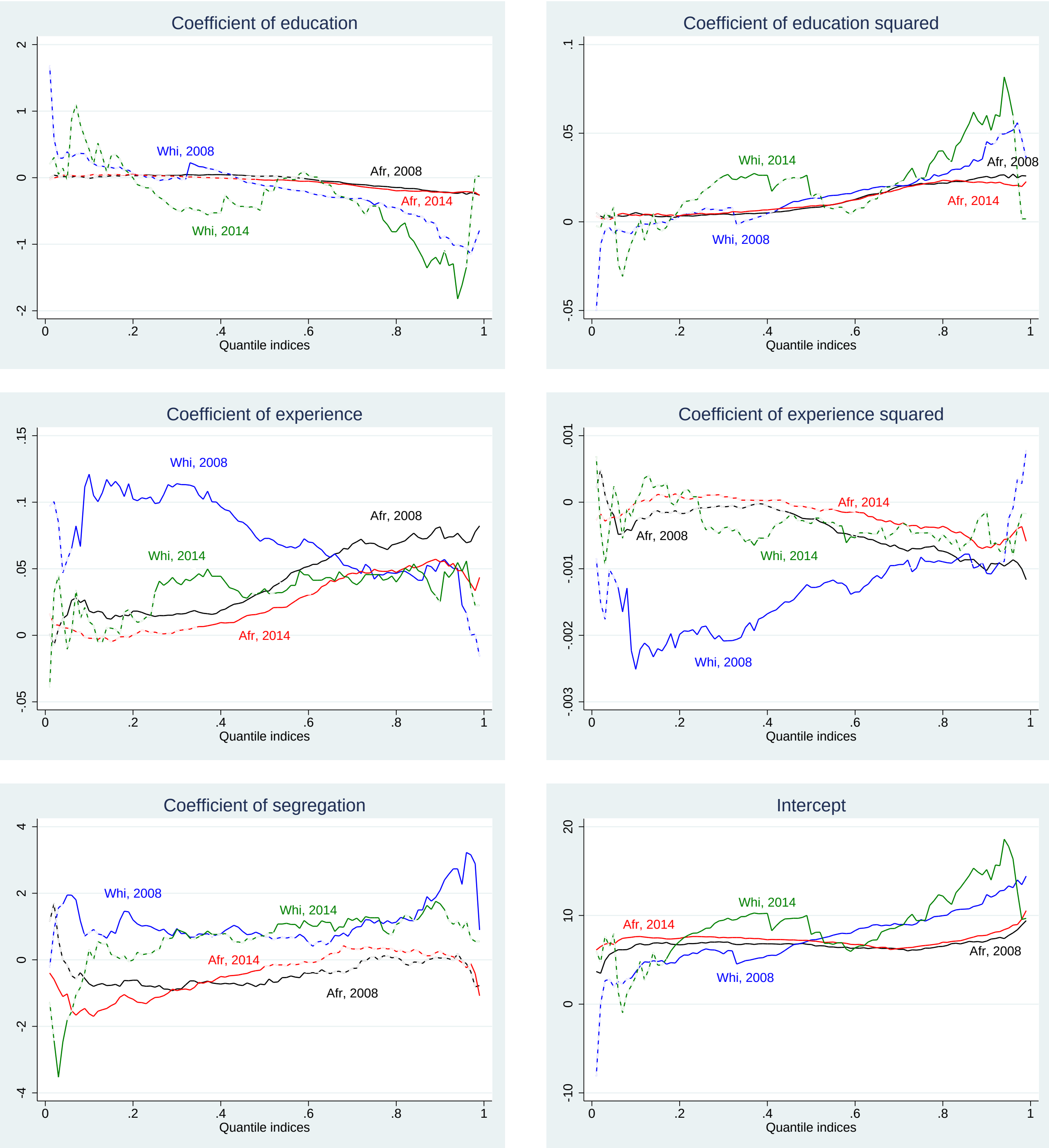

FIGURE II

RIF Regressions of Income by Racial Group and Year

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level. "Afr" and "Whi" refer to Africans and Whites, respectively.

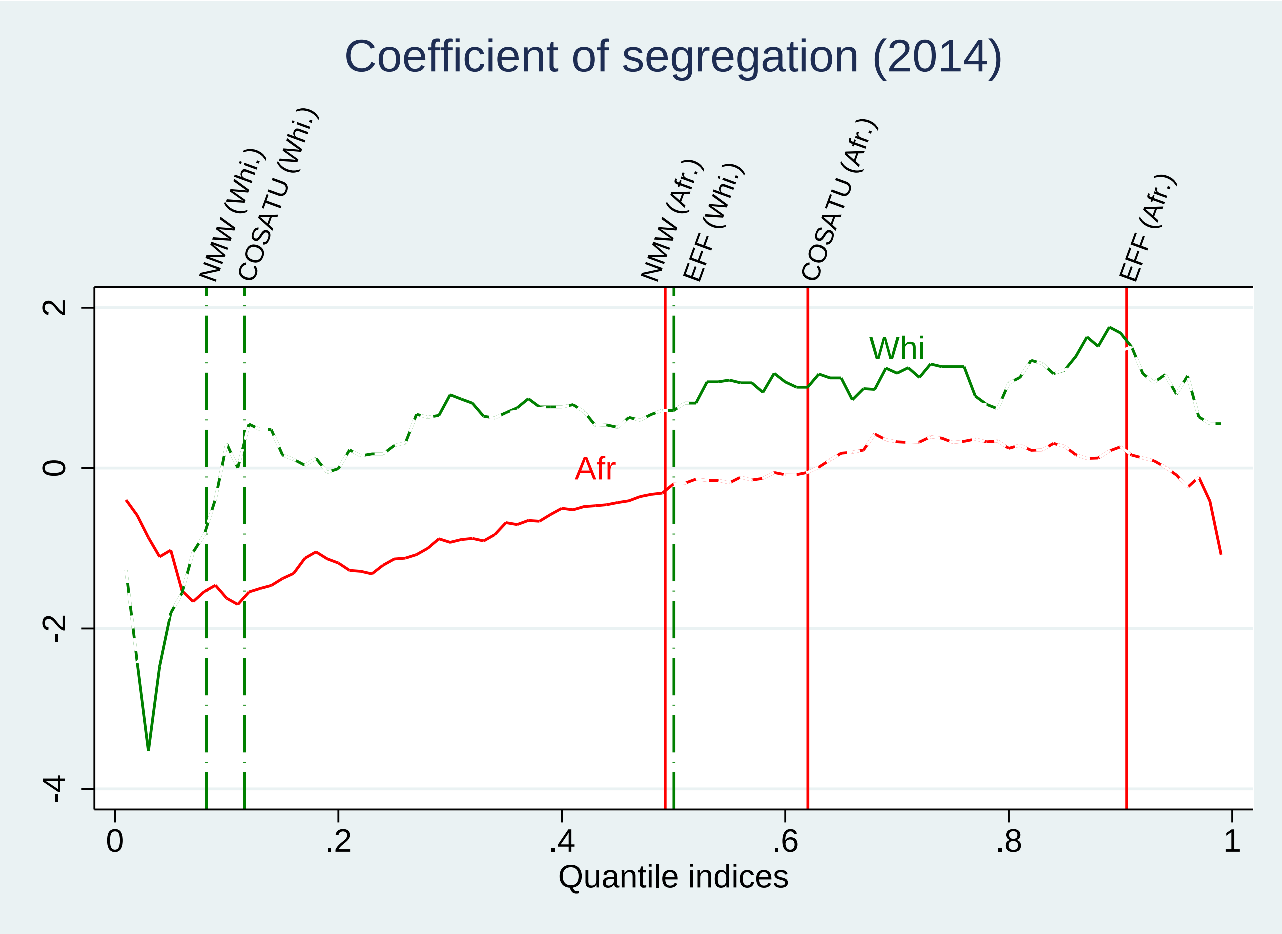

Figure III

RIF-regression coefficients of Segregation with target groups

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level. "Afr" and "Whi" refer to Africans and Whites, respectively.

THE CONTRIBUTION OF RESIDENTIAL SEGREGATION TO RACIAL INCOME GAPS: EVIDENCE FROM SOUTH AFRICA

Appendices for online publication

Florent Dubois* and Christophe Muller†

July 2020

I. Appendix to Section III.C (Endogeneity and Selection)

I.A. *Endogeneity*

As suggested by Cutler and Glaeser (1997), topographical features fragmenting the land provide natural boundaries that can be exploited to segregate individuals. Rivers provide such barriers while they should not have any impact on criminality.

We use data on the South African rivers in 2016 from the International Steering Committee for Global Mapping hosted at the online Stanford Libraries. On average, there are 60 rivers in a district council, 47 between-district and 13 within-district rivers. The rivers are mapped in Figure I.

[Figure 1 about here.]

We follow Cutler and Glaeser (1997) by using a quadratic polynomial of between and within-district council rivers as instruments for segregation. We also transform each instrument into its natural logarithm. We apply the same instrumentation strategy for Unconditional Quantile IV regressions (Powell forthcoming). The estimation procedure is based on Markov-Chain Monte-Carlo methods. For each quantile, we use 7000 draws (10000 draws in total, and the first 3000 burned-in), and an acceptance rate of 0.9. We then use a local polynomial regression on the quantile effects to avoid the estimate to jump erratically from one quantile to another. We use the standard errors of the local polynomial regressions for inference.[1] Results are displayed in Tables I and II for the mean analysis and in Figures II and III for the distributional analysis.

[Table 1 about here.]

[Table 2 about here.]

*corresponding author: florent.dubois@parisnanterre.fr

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

[1] Inference is identical if we use the smoothed standard errors of the quantile IV regressions instead.

The results obtained with IV are qualitatively similar to those obtained with OLS. The coefficients of experience and education are close. The main difference is the large change in magnitude for segregation and the intercept among Africans for both years, though there are still going in the same direction as the OLS estimates. In the decomposition of the mean effect, the same observations apply.

This instrumentation strategy has the same limitations in our case as in Cutler and Glaeser (1997). The instruments generally fail to pass the overidentification test, except for Whites in 2014 (Sargan Chi^2(3)= 1.90, P-value=0.59). But these tests relies on the assumption of i.i.d. error terms which is unlikely to be satisfied in our data set as well as any individual data set. Robust tests for weak instruments a la Montiel Olea and Pflueger (2013) suggest that the rivers instruments may suffer from this problem as they fail to pass the test at the 10 percent level when estimated with 2SLS (Effective F-statistic = $16.935 < 18.195$), although passing at the 10 percent level when estimated via LIML and giving similar results (Effective F-statistic = $16.935 > 14.558$). Finally, Hausman tests suggest that OLS performs as equally as 2SLS for Whites in both years (F-statistic (2008) = 0.13, P-value=0.72; F-statistic (2014) = 0.063, P-value=0.80), while endogeneity seems to affect mostly regressions for Africans in both years. Overall, this instrumentation strategy provides some evidence of a robust effect at least for Whites in 2014. But, it also sheds light on its limits for future research.

[Figure 2 about here.]

In the distributional analysis, the aggregate decomposition slightly changes. The largest difference between the composition and structure effect occurs between the second and fourth deciles instead of the top without instrumentation, while the structure effect exceeds the composition effect at the two extremes of the distribution. The effect is more pronounced in 2014 than in 2008. Quantitatively, although the magnitude is smaller after instrumenting, each component are always positive even at the top of the distribution. The dynamics over the distribution of the composition effect is now mostly driven by segregation as the effects of education and experience are monotone and almost flat over the entire distribution. In the structure effect, segregation contributes to increasing the income gap in a wider range than in the base estimates. In 2008, it increases the gap over all the distribution, while in 2014, the very top is not affected by segregation.

[Figure 3 about here.]

The Unconditional Quantile IV estimates depict a similar picture as RIF-regression estimates. Whites still massively benefit from segregation whereas Africans are negatively affected. However, these new estimates suggest that, in 2014, some very poor Whites may be hurt and some very rich Africans may benefit from segregation as well. Overall, instrumentation does not change the main conclusion of the main analysis.

I.B. Selection

Labor market selection may interact in a very complex manner with segregation. Modelling precisely these interactions is beyond the scope of our paper. However, we can get a sense of the bias that can affect our results due to labor market selection. Labor market selection raises the question of estimating the counterfactual wage that unemployed would get if they would have worked with the wage structure of the currently employed.

Chandra (2003) proposes a bound approach that assumes that all the unemployed would have earned a salary below the median wage earned by similar workers. This scenario is based on the hypotheses that there is no rationing, except for frictional unemployment, in the labor market and the unemployed have lower unobservable characteristics. Therefore, if individuals are not working this is only because their offer wage is below their reservation wage. It has the advantage of avoiding modelling explicitly the selection process.

In South Africa, labor market is plagued by high unemployment and skills mismatch. Hence, the reality of the support of the counterfactual wage distribution of the unemployed is most likely to be something in-between the extreme scenario described by Chandra (2003) and another extreme scenario that would assume that all the unemployed would have earned more than the median wage of similar workers. These two scenari define a lower bound and an upper bound with respect to labor market selection.

Following Chandra (2003), we define 36 cells based on year (2008 and 2014), race (Whites and Africans), age groups[2] (less than 30, between 30 and 55, more than 55), and education groups (less than 12 years of education, 12 years and matriculation, and college educated). Then, we uniformly draw a multiplier to apply to the median wage of the cell to construct a counterfactual wage for those with a missing wage in the cell. In the lower bound scenario, this multiplier is in the range $[0, 1)$, and in the range $(1, 2]$ for the upper bound scenario. We impute wages in this manner to the unemployed, those explicitly discouraged, and the employed having a missing wage. We exclude those declared inactive as they have already decided not working based on the current wage structure of the workers on which we base our imputation. Results are provided in Tables III, IV, V, and VI for the mean analysis, and in Figures IV, V, VI, VII, VIII, and IX for the distributional analysis.

[Table 3 about here.]

In the lower bound scenario, the effect of experience and education is similar. Only the magnitude is reduced and the effect for Whites dominates the effect for Africans, although there are minor differences. Regarding segregation, estimates still indicate a positive effect for Whites and negative for Africans on average, but the magnitude increased.

[Table 4 about here.]

In the upper bound scenario, the effect of experience and education still display a quadratic polynomial but the gap between the two racial groups is reduced. This time, however, education favors Africans in 2008 because this scenario gives a better wage to more individuals with fewer education. In 2014, the situation reverts back to the initial estimates but the gap is still reduced. Regarding segregation, the positive effect for Whites still persists. It is even significant at the 10 percent level in 2014 in this scenario. On the other hand, for Africans, the effect is close to zero with very low t-statistics which suggests a lot of heterogeneity in both years for the distributional analysis.

[Table 5 about here.]

[Table 6 about here.]

[2] We also tried with only 2 age categories (less than 30, and 30 and above) to increase the number of individuals by cells. This gives substantially similar results.

In the lower bound scenario, the main features uncovered previously are still present. Education is still the main contributor to the composition effect. Segregation follows next. Experience still appears as the least important contributor. In the structure effect, segregation is still by far the most important contributor. The novelty comes from the new negative contribution of experience in 2008.

In the upper bound scenario, the composition effect is still similar to the main estimate. Only segregation slightly contributes to reduce the gap in 2008. In the structure effect, segregation keep contributing to increase the gap in 2008, but it is no longer the largest contribution, since education reduces more the gap, although only at the 10 percent level. Finally, in 2014, the contribution of segregation to the structure effect is no longer significant, while education now increases the gap. Overall, the main effect of segregation keeps going in the same direction in both scenari although it is fuzzier in the upper bound scenario.

[Figure 4 about here.]

Moving to the distributional analysis, the two scenari provide a similar picture for the aggregate decomposition. The global pattern is the same as the base sample. The composition and the structure effects are roughly equal and start to diverge around the median or above depending on the year. The main difference is mostly related to the levels of each component. The upper bound scenario widens and stabilizes the gap between composition and structure effects, especially in 2008, before the composition effect takes the lead.

[Figure 5 about here.]

When we look at the detailed decomposition of the composition effect, the situation depicted in 2008 is similar to the base sample. Education is the main contributor before segregation and is especially important in the top part of the distribution. Similarly, segregation becomes more and more important after the median. It even peaks above education for the last decile in the upper bound scenario. Experience is the smallest contribution when it is significant. In 2014, the picture changes in the lower bound scenario. Isolation becomes the main contributor before education, and experience trails behind. In the upper bound scenario, the general pattern resembles the base estimates but at a reduced level for all contributions. Education is still the first contribution but only by a slight margin before segregation. Experience comes last.

[Figure 6 about here.]

In the structure effect, the contribution of segregation is unambiguously positive over almost all the distribution in the lower bound scenario in both years. In the upper bound scenario, segregation contributes to reducing the income gap from the 8th decile in 2008. In 2014, this negative contribution moves slightly down to the distribution from the third quartile to the 85th decile. There is also a small positive contribution around the first decile of the distribution. Overall, the contribution of segregation is never positive at the bottom of the distribution. At best, segregation does not contribute to increasing the income gap, but often, it increases the discrepancy between the poor. At the top, segregation often does not contribute to the income gap, but it has some potential to help reducing the gap in the upper bound scenario. This suggests that the elimination of concentrated poverty in poor ghettos is a major public policy objective. At the top of the distribution, segregation may be a way for individuals to circumvent discriminatory practices and asks for further research.

[Figure 7 about here.]

The main patterns remains valid in both scenari for education. The coefficients describe a quadratic polynomial with increasing returns, but which group is favored depends on the scenario and year. For instance, in the lower bound scenario, the quadratic term is higher or equal for Africans than for Whites and only slightly more negative for the linear term. This suggests that Africans benefit more from education in the lower bound case, especially in 2008 at the top of the distribution, which is confirmed in the structure effect. On the other hand, in the upper bound scenario, in 2014, although the quadratic term is lower for Whites, the linear term always dominate the effect for African. Thus, Whites still benefit more from education at least up to the 7th decile where the quadratic term for Africans exceeds markedly the one for Whites.

[Figure 8 about here.]

For experience, the main results emerge again. The general pattern is quadratic with decreasing returns, except for Whites at the bottom of the distribution in 2014, in both scenari.

[Figure 9 about here.]

The coefficients for segregation of the RIF-regressions are consistent with the base estimates. In the lower bound scenario, the discrepancy between Africans and Whites is more pronounced. Whites benefit from segregation over almost all the distribution in 2008 and only for the last quartile in 2014. For Africans, segregation is negative up to the 8th decile and not significant afterwards, in both years. In the upper bound scenario, this pattern repeats with some new insights. The positive effect of segregation for Whites appears now only in the last quartile in 2008 and between the third and seventh deciles in 2014. It is still not significant elsewhere in the distribution. For Africans, the effect is never significant in 2008. In 2014, it is negative at the bottom of the distribution but for a smaller part stretching up to slightly before the second decile. The effect even becomes positive for most of the middle classes (from the first quartile to the 8th decile), although it is almost always smaller than the effect of Whites. Overall, the effect of segregation is positive for mostly affluent Whites and negative mostly for deprived Africans.

II. Appendix to Section IV.C (Data): sample selection issues

In the paper, we use the largest sample possible, analyzing employed workers older than 15 and declaring a positive monthly wage from their main occupation. In this section, we demonstrate that our main results are robust to alternative sample selection criteria.

Potential concerns regarding part-time and seasonal workers, workers older than the retirement age, or early retirement may affect our results. We address seasonal workers by droping workers whose on-going contract started more than a year ago at the time of the interview. Focusing only on workers declaring between 40 and 45 hours a week to address part-time would have reduced too much the sample to still be informative. We use hourly wages instead of monthly wages as the dependent variables. We compute hourly wages by dividing monthly wages by four times the average working hours per week. Moreover, note that the number of hours worked weekly is, on average, identical for Whites and Africans in 2014. Finally, we address participation around retirement and incorporation of young graduates in the labor market by focusing on individuals between 25 and 55 years old. We only display the results for the more restrictive criteria in Tables VII and VIII for the mean analysis and in Figures X and XI for the distributional analysis.

[Table 7 about here.]

Under this more restrictive sample selection criterion, most of the effects described in the main analysis are reproduced, although some are less precisely estimated due to the reduced sample size. The magnitude of some coefficients changes as well. This is the case for segregation, for the Whites for instance, which is lower in 2008 but larger and now significant at the 5 percent level in 2014.

[Table 8 about here.]

For the decomposition of the mean, in Table VIII, the main results are also confirmed. Differences in education explains almost exclusively the composition effect in both year, while differences in segregation returns explain exclusively the structure effect.

When we turn to the distributional analysis, the results are less black and white than with the mean. The aggregate decomposition still shows a more important contribution of the structure effect at the bottom of the distribution while the composition effect dominates afterward. Magnitudes have changed however. The composition effect is mostly smaller than in the base estimates while the structure effect is either larger than the base estimates in some parts of the distribution or at a similar level.

The detailed decomposition of the composition effect depicts a similar picture. As in the base estimates, education is the most important contributor, followed by segregation, and experience is negligible. Only the levels are slightly reduced compared to the base estimates.

In the detailed analysis of the structure effect, the main conclusions still hold. Segregation still contributes to increasing the income gap at the bottom of the distribution but at the 10 percent level only in 2008. In 2014, the contribution of segregation is even larger than in the base estimates. Experience now only contributes to reducing the gap above the median. In 2014, it contributes only slightly at the bottom of the distribution. Education and the intercept still cancel out.

[Figure 10 about here.]

The RIF-regressions also portray a similar picture. Only the levels change slightly for education and the intercept. Experience loses its statistical significance for Whites in both years. For segregation, the general pattern is identical to the base estimates. Whites benefit from a positive effect over large parts of the distribution whereas Africans experience a negative effect below the median. However, the effect of segregation for Whites is less clearly estimated at the top of the distribution but would appear significant at the 10 percent level.

[Figure 11 about here.]

Firm size may also distort our results if one group works disproportionately in large firms, for instance, as larger firms give higher wages. However, firm size is only available in the 2014 wave. To get a sense of the potential bias generated by firm size, we provide, in Table IX, the results of the regression of income on firm size categories in each racial group. As we use the smallest firm category as reference, this gives us an estimate of how much more a worker is earning due to the size of the firm he is working for. If our results are confounded by firm size, then Whites should get a better premium than Africans as firm size grows. Results from this analysis clearly contradict this hypothesis. Whites never get a premium depending on the size of their employer. Africans, on the other hand, receive an increasing premium from 1050 rands to almost 3500 rands. Therefore, firm size has the opposite effect on the wage gap, and our results underestimate the effect of segregation.

[Table 9 about here.]

III. Appendix to Section VII.C (2018 Minimum Wage Reform)

III.A. RDD and Affirmative Action

Individuals starting their jobs just before and after the bill was passed should not differ too much in terms of education choices as their choices were determined long before the bill was discussed. If the Employment Equity Act has an effect on labor market segmentation, we should observe a change in the returns to education the year of the bill. We estimate below the returns to education with the Kernel-Regularized Least-Square estimator advocated by Hainmueller and Hazlett (2014). This method has two advantages for our purpose. First, it is akin to local regression methods, thereby providing an estimated coefficient per individual. Second, it avoids misspecifications issues of possible interactions between regressors as it is data-driven. The estimated equation is a standard Mincerian equation with the deflated individual wage for the primary occupation as the dependent variable. It includes, as independent variables, years of education, total working experience (age minus years of education minus six), and on-the-job experience (years between the start of the current job and the interview (2014 or 2015)).

We report in Figure XII the results of a regression discontinuity analysis of education returns as a function of the job's starting year. Among the poorest African group (top left panel), the reform seems to have slightly raised the return to education, but this difference is not significant, as reported in Table IV. However, the outlier return to education observed just before 1980 for this group might be responsible for this insignificance. There is no effect of the reform for the COSATU's specific target group. For the EFF's specific target group, although there is no break in return levels at the time of the reform, the positive trend in education returns suddenly vanishes, which creates a kink. In the richest group (bottom right panel), there seems to be both a break in level and a kink, although the gap in return levels before and after the reform is not significant.

[Figure 12 about here.]

However, further insight can be gained by looking at the distribution of effects. Kernel-Regularized Least-Squares density estimates of the education return in a Mincer model are displayed for all African groups in Figure XIII, and by specific target group in Figure XIV. In Figure XIII, if the Employment Equity Act had any effect on education return, it was not through the central distribution location, which is almost the same before and after Affirmative Action. Yet, there is a drastic increase of the polarization of returns after the reform.

[Figure 13 about here.]

This increased polarization is not limited to a particular group, as it appears in Figure XIV, with density estimates by specific target group. The results of the Kolmogorov-Smirnov tests show that, except for the richest group, all distributions are significantly different from their counterpart before the Affirmative Action.

One lesson from all these results is that the segregation effects that are related to market segmentation do not discriminate against the poorest Africans exclusively but rather exert their influence along most of the income distribution. Within each specific constituency, the increased polarization observed after the Affirmative Action suggests that the market segmentation may have worsened, perhaps with the emergence of a new class of African insiders within the formal sector. However, other time-dependent changes in the South African economy, such as structural changes in activity sectors, may also explain this worsened polarization.

[Figure 14 about here.]

III.B. Preferences to stay and racial preferences

We report the results of a one-way ANOVA between the level of segregation and the preference to leave in Table X. ANOVA results are sensitive to the normality assumption and to the equality of variances between groups. As these two assumptions are not fulfilled, we also report the results of the Kruskal-Wallis test, a non-parametric alternative. Both the one-way ANOVA and the Kruskal-Wallis test elicit an association between segregation and preference to leave. It suggests that a racial preference might express itself through the preference to leave.

[Table 10 about here.]

References

Chandra, Amitabh, "Is the Convergence in the Racial Wage Gap Illusory?," NBER Working Paper Series, WP9476, 2003.

Cutler, David M., and Edward L. Glaeser, "Are Ghettos Good or Bad?," *Quarterly Journal of Economics*, 112 (1997), 827–872.

Hainmueller, Jens, and Chad Hazlett, "Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach," *Political Analysis*, 22 (2014), 143–168.

Montiel Olea, Jose Luis, and Carolin Pflueger, "A Robust Test for Weak Instruments," *Journal of Business and Economic Statistics*, 31 (2013), 358–369.

Powell, David, "Quantile Treatment Effects in the Presence of Covariates," *Review of Economics and Statistics*, forthcoming.

TABLE I
IV ESTIMATES

	2008		2014	
	White	African	White	African
Isolation	1.27*	-5.72***	0.74	-7.80***
	(1.89)	(-4.34)	(0.85)	(-3.94)
Experience	0.08***	0.04***	0.03***	0.02***
	(6.12)	(10.30)	(2.90)	(5.57)
Experience (square)	-0.00***	-0.00***	-0.00	-0.00
	(-5.12)	(-5.49)	(-1.39)	(-1.12)
Years of schooling	-0.16*	-0.07***	-0.30	-0.09***
	(-1.86)	(-4.55)	(-1.09)	(-5.38)
Years of schooling (square)	0.01***	0.01***	0.02*	0.01***
	(4.24)	(14.56)	(1.88)	(14.58)
Constant	7.18***	11.50***	8.63***	14.14***
	(9.24)	(9.48)	(4.67)	(7.62)
Observations	440	2922	229	5291
R^2	0.354	0.083	0.278	.

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

TABLE II
Oaxaca Decompositions - IV Estimates

	2008	2014
Differential		
Prediction_1	9.09***	9.20***
	(213.86)	(161.45)
Prediction_2	7.70***	8.01***
	(363.86)	(471.10)
Difference	1.39***	1.20***
	(29.34)	(20.13)
Explained		
Isolation	0.00	0.00
	(.)	(.)
Experience	-0.01	0.06***
	(-0.57)	(4.13)
Education	0.64***	0.46***
	(20.72)	(14.83)
Total	0.63***	0.52***
	(19.53)	(14.90)
Unexplained		
Isolation	6.11***	7.74***
	(3.83)	(3.59)
Experience	0.09	0.11
	(0.75)	(0.76)
Education	-1.12	-1.66
	(-1.63)	(-0.96)
Constant	-4.32**	-5.51*
	(-2.50)	(-1.95)
Total	0.76***	0.68***
	(17.72)	(12.79)
Observations	3362	5520

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 Africans (in 2014(Nov.) Rands).

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

TABLE III

OLS Regressions - Lower Bound

	2008		2014	
	White	African	White	African
Experience	0.04*	0.08***	0.05***	0.03***
	(1.71)	(13.06)	(2.90)	(6.71)
Experience (square)	-0.00	-0.00***	-0.00*	-0.00***
	(-1.49)	(-9.47)	(-1.96)	(-2.98)
Years of schooling	-0.08	-0.10***	0.04	-0.11***
	(-1.04)	(-8.74)	(1.11)	(-12.14)
Years of schooling (square)	0.01***	0.02***	0.01**	0.02***
	(3.58)	(20.96)	(2.06)	(25.80)
Isolation	1.62***	-1.08***	0.59	-0.85***
	(4.82)	(-6.17)	(1.51)	(-5.50)
Constant	6.40***	6.43***	6.03***	7.18***
	(11.40)	(34.29)	(16.26)	(44.82)
Observations	565	5762	448	8536
R^2	0.205	0.199	0.354	0.185

t statistics in parentheses

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

TABLE IV
OLS REGRESSIONS - UPPER BOUND

	2008		2014	
	White	African	White	African
Experience	0.05***	0.05***	0.02	0.02***
	(3.15)	(13.69)	(1.57)	(6.43)
Experience (square)	-0.00*	-0.00***	-0.00	-0.00***
	(-1.94)	(-10.17)	(-0.07)	(-3.89)
Years of schooling	-0.12***	-0.07***	-0.09***	-0.16***
	(-2.68)	(-10.07)	(-4.26)	(-27.01)
Years of schooling (square)	0.01***	0.01***	0.01***	0.01***
	(6.01)	(25.62)	(7.57)	(38.90)
Isolation	0.56***	0.10	0.38*	0.04
	(2.63)	(0.98)	(1.78)	(0.42)
Constant	7.88***	6.74***	8.08***	7.81***
	(22.23)	(64.26)	(39.85)	(76.48)
Observations	565	5762	448	8536
R^2	0.300	0.288	0.372	0.212

t statistics in parentheses

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

TABLE V

Oaxaca decompositions - Lower Bound

	2008	2014
Differential		
Prediction_1	8.83***	8.30***
	(200.37)	(145.18)
Prediction_2	7.06***	7.48***
	(468.39)	(604.23)
Difference	1.76***	0.82***
	(37.90)	(13.96)
Explained		
Experience	0.04***	0.11***
	(3.41)	(9.36)
Education	0.77***	0.22***
	(22.69)	(5.86)
Isolation	0.16***	0.20***
	(3.64)	(4.83)
Total	0.97***	0.53***
	(16.93)	(9.64)
Unexplained		
Experience	-0.66**	0.27
	(-2.41)	(0.97)
Education	-0.40	0.20**
	(-0.79)	(2.36)
Isolation	1.88***	0.97***
	(7.53)	(3.22)
Constant	-0.02	-1.15**
	(-0.04)	(-2.38)
Total	0.79***	0.29***
	(11.76)	(4.38)
Observations	6327	8984

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 Africans (in 2014(Nov.) Rands).

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

TABLE VI

OAXACA DECOMPOSITIONS - UPPER BOUND

	2008	2014
Differential		
Prediction_1	9.37***	9.23***
	(316.21)	(291.57)
Prediction_2	7.87***	8.18***
	(879.44)	(1019.64)
Difference	1.50***	1.05***
	(48.53)	(32.22)
Explained		
Experience	0.03***	0.05***
	(3.64)	(8.02)
Education	0.56***	0.26***
	(24.71)	(12.30)
Isolation	-0.06**	-0.03
	(-2.02)	(-1.07)
Total	0.53***	0.29***
	(14.37)	(9.16)
Unexplained		
Experience	0.06	0.10
	(0.33)	(0.81)
Education	-0.55*	0.18***
	(-1.73)	(4.33)
Isolation	0.32**	0.23
	(2.11)	(1.59)
Constant	1.14***	0.26
	(2.72)	(1.29)
Total	0.97***	0.76***
	(22.51)	(20.45)
Observations	6327	8984

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 Africans (in 2014(Nov.) Rands).

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

TABLE VII
OLS REGRESSIONS - SAMPLE SELECTION

	2008		2014	
	White	African	White	African
Experience	0.05*	0.06***	-0.01	0.02***
	(1.76)	(6.46)	(-0.37)	(2.75)
Experience (square)	-0.00	-0.00***	0.00	-0.00
	(-1.24)	(-4.24)	(0.71)	(-0.09)
Years of schooling	-0.08	-0.05***	-0.44	-0.07***
	(-0.72)	(-2.93)	(-1.01)	(-3.95)
Years of schooling (square)	0.01**	0.01***	0.02	0.01***
	(2.56)	(11.80)	(1.37)	(13.18)
Isolation	0.63*	-0.45*	1.25**	-0.31*
	(1.77)	(-1.79)	(2.06)	(-1.76)
Constant	2.13***	1.58***	4.98*	2.28***
	(2.70)	(5.90)	(1.72)	(11.80)
Observations	217	1480	117	2992
R^2	0.315	0.349	0.181	0.271

t statistics in parentheses

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

TABLE VIII

OAXACA DECOMPOSITIONS - SAMPLE SELECTION

	2008	2014
Differential		
Prediction_1	4.14***	4.16***
	(80.58)	(55.66)
Prediction_2	2.95***	3.13***
	(116.77)	(198.04)
Difference	1.19***	1.03***
	(20.83)	(13.46)
Explained		
Experience	-0.03	0.04***
	(-1.58)	(2.67)
Education	0.68***	0.45***
	(14.94)	(11.31)
Isolation	0.04	0.05
	(0.69)	(1.33)
Total	0.70***	0.54***
	(8.83)	(9.36)
Unexplained		
Experience	-0.27	-0.35
	(-0.85)	(-1.14)
Education	-0.57	-2.90
	(-1.09)	(-1.17)
Isolation	0.78**	1.03***
	(2.36)	(2.86)
Constant	0.55	2.71
	(0.69)	(1.01)
Total	0.49***	0.49***
	(5.73)	(5.75)
Observations	1697	3109

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 Africans (in 2014(Nov.) Rands).

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

TABLE IX

OLS ESTIMATES OF FIRM SIZE ON THE INCOME GAP

	White	African
1-4	0.00	0.00
	(.)	(.)
5-9	-5796.80	1058.01***
	(-1.65)	(4.36)
10-19	-3928.26	1736.10***
	(-1.14)	(7.85)
20-49	-536.82	1878.63***
	(-0.16)	(8.75)
50 or more	3570.46	3424.28***
	(1.20)	(17.33)
Constant	14168.52***	2794.09***
	(5.69)	(19.26)
Observations	185	4649

t statistics in parentheses

The dependent variable is the real monthly income of Africans (in 2014(Nov.) Rands).

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

TABLE X
One-way ANOVA

	Sum of squares	Degrees of freedom	Mean square	F-statistics	P-value
Between groups	.20447	4	.0511	8.90	$< .00005$
Within groups	28.045	4881	.0057		
Total	28.249	4886	.0057		

Kruskal-Wallis test

	Statistics	P-value
Chi-Squared	116.9446	.0001
Chi-Squared with ties	117.0688	.0001

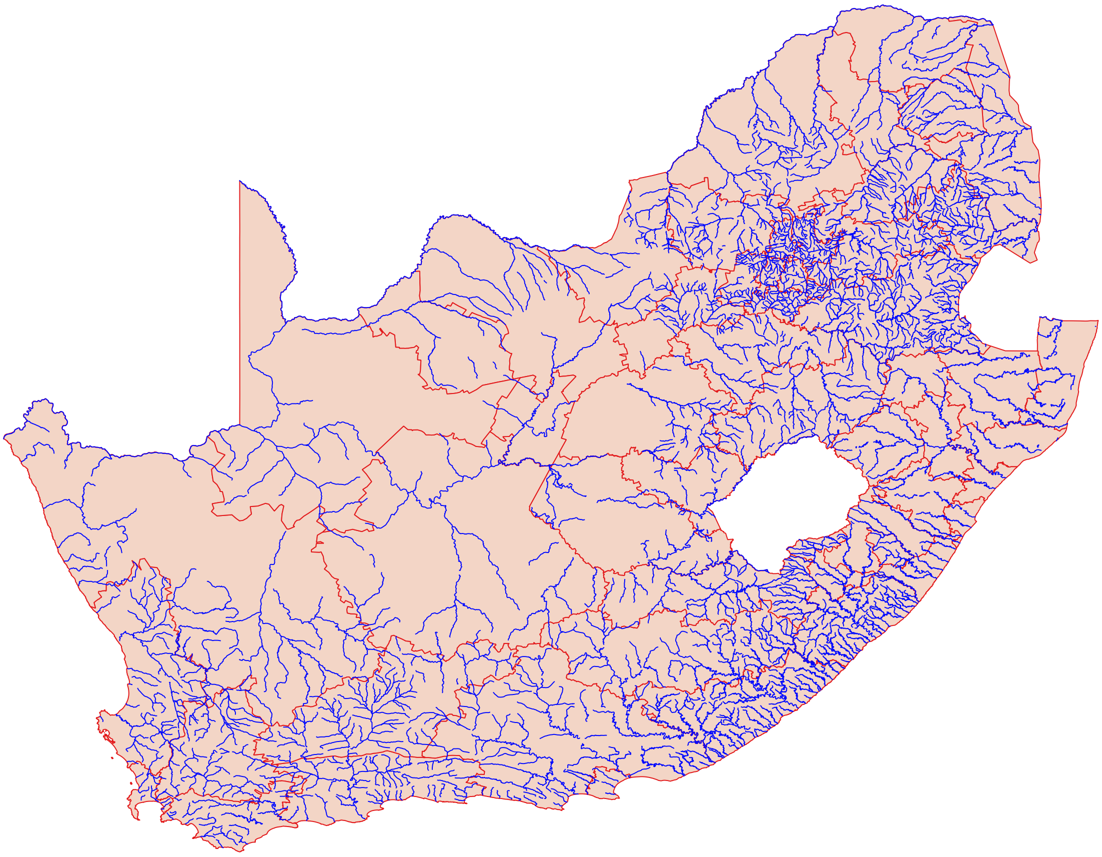

Figure I
South African Rivers
Blue lines are rivers. Red lines are district councils borders as of 2001.

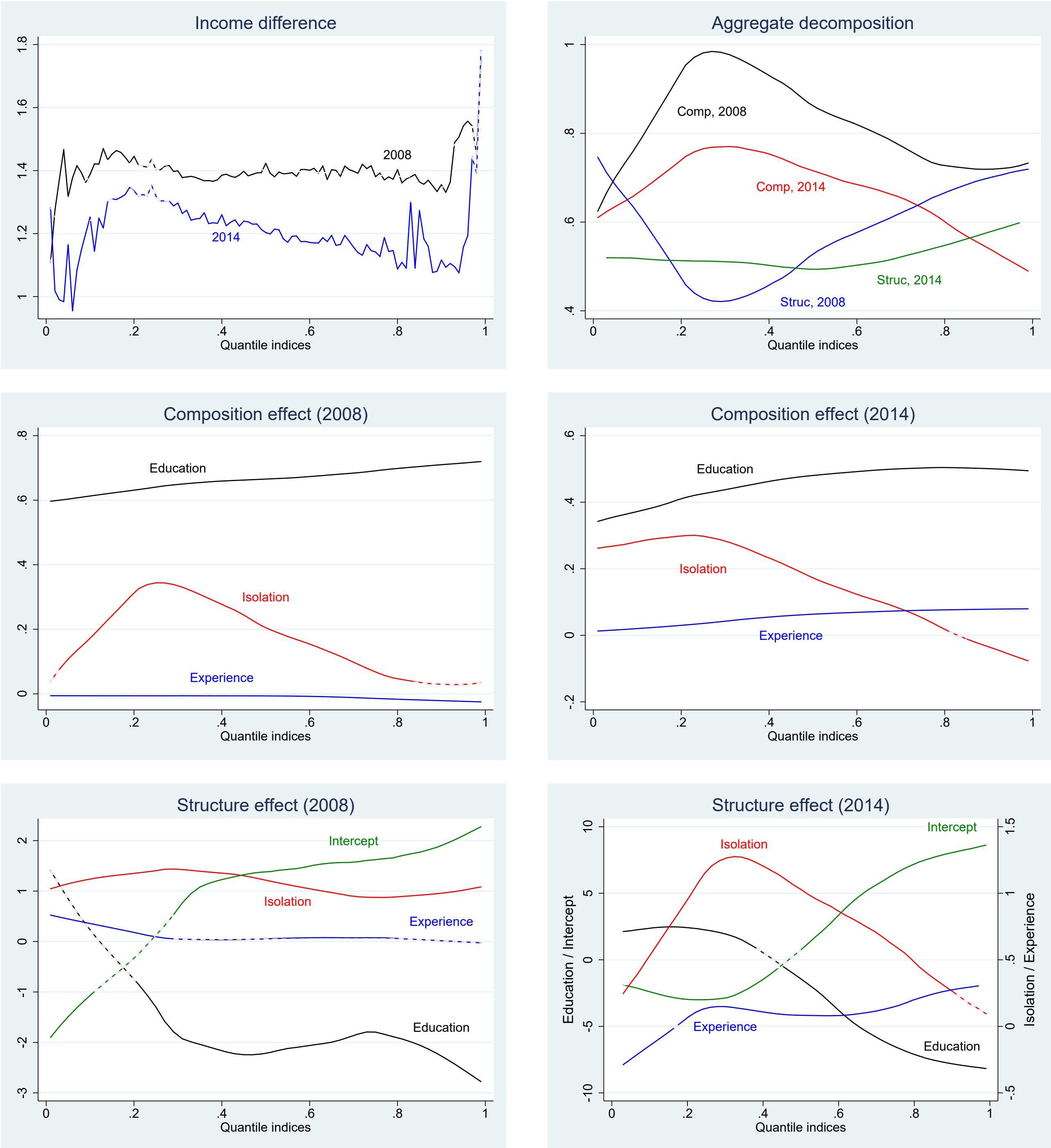

FIGURE II

Aggregate and Detailed Decompositions in 2008 and 2014 - Instrumental Variables Estimates

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level, except for the income differences, where they represent the quantiles for which the income difference in 2008 is not significantly different from the income difference in 2014. “Comp” and “Struc” refer to the composition effect and the structure effect, respectively. Income differences estimates are not smoothed.

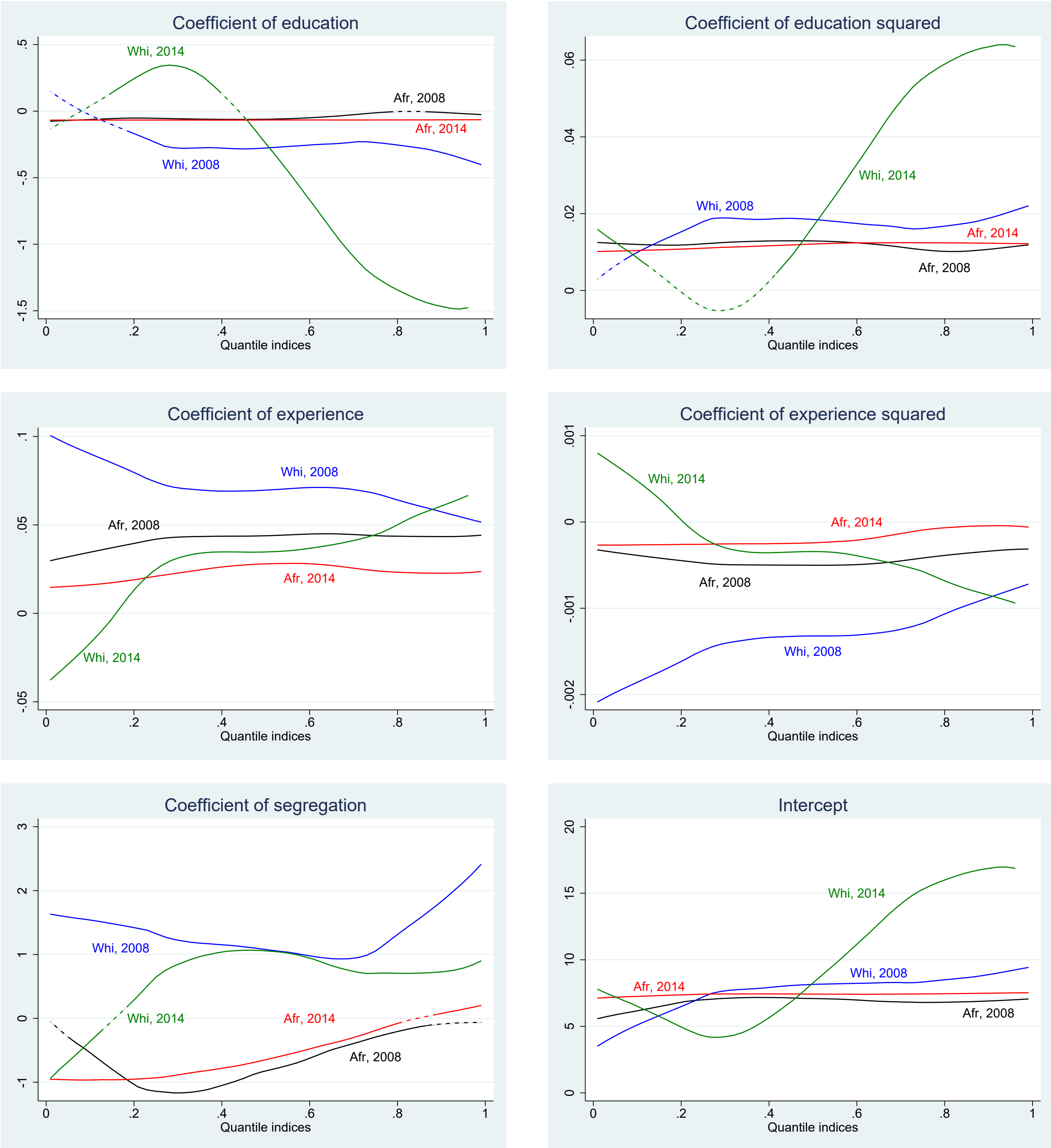

FIGURE III

Unconditional Quantile IV Regressions of Income by Racial Group and Year

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level. "Afr" and "Whi" refer to Africans and Whites, respectively.

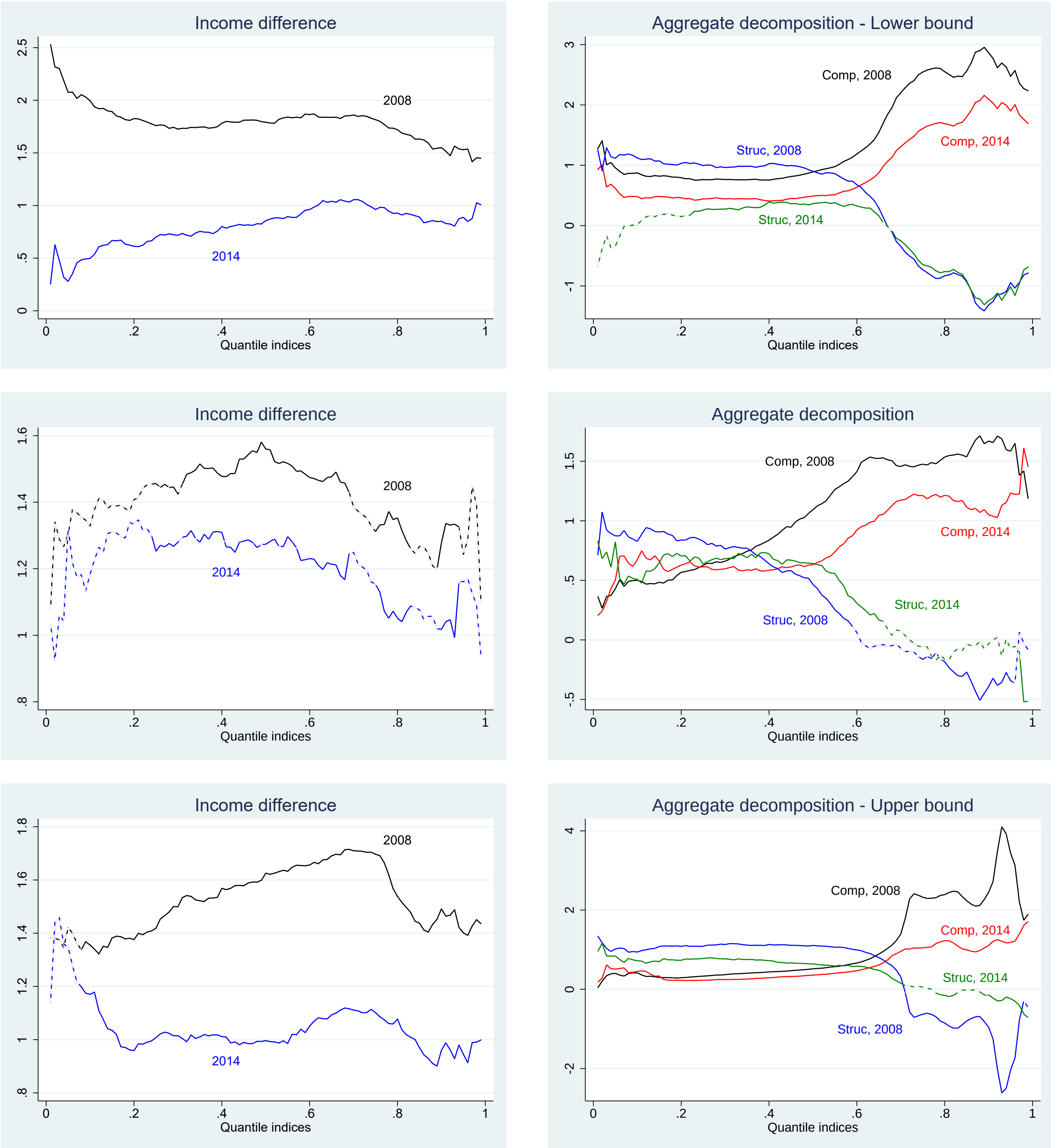

FIGURE IV

Income Gap and Aggregate Decompositions in 2008 and 2014 - Labor Market Selection Scenari

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level, except for the income differences, where they represent the quantiles for which the income difference in 2008 is not significantly different from the income difference in 2014. “Comp” and “Struc” refer to the composition effect and the structure effect, respectively.

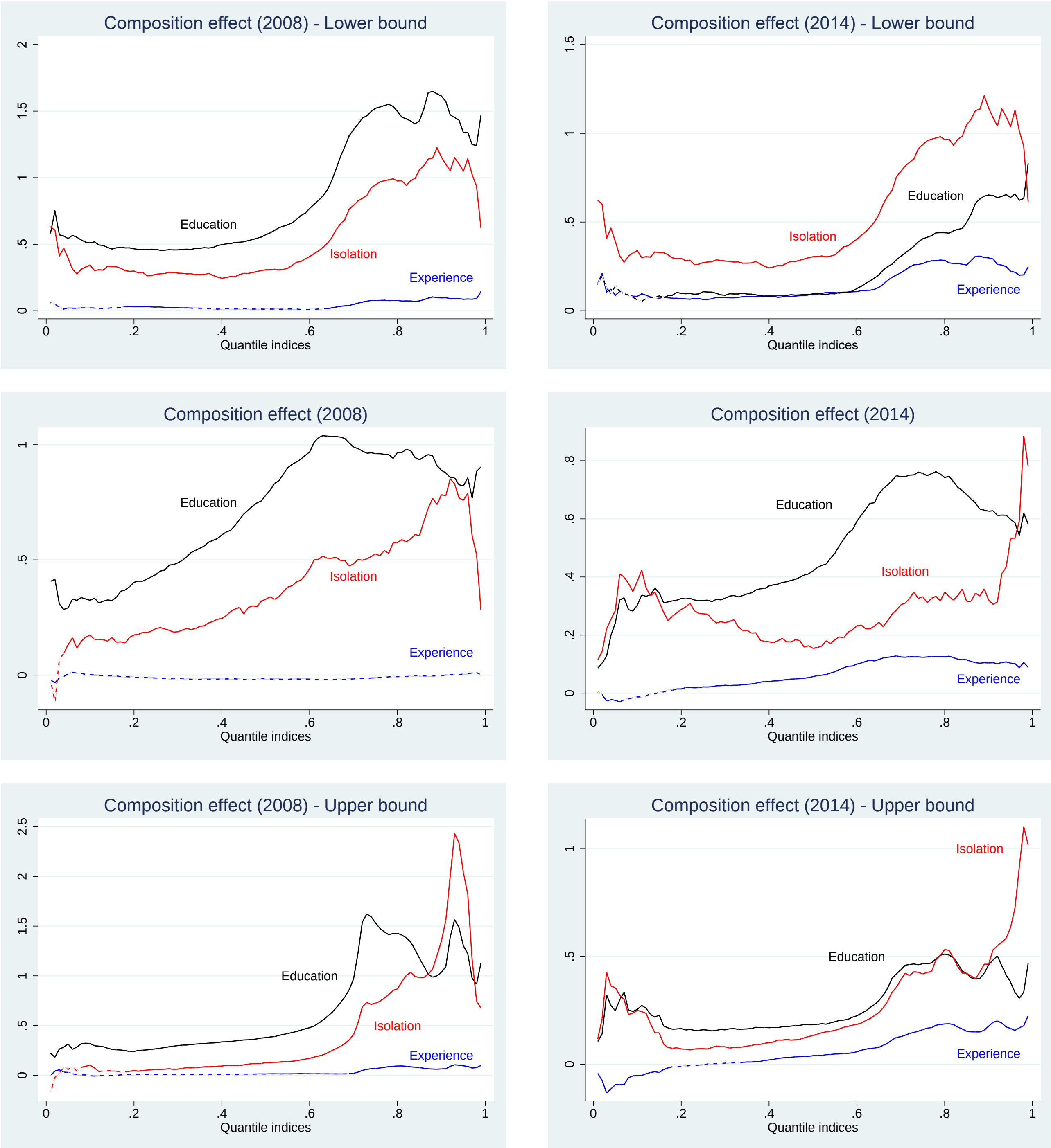

FIGURE V

Composition Effect in 2008 and 2014 - Labor Market Selection Scenari

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level. "10%" refers to an effect significantly different from zero at the 10 percent level.

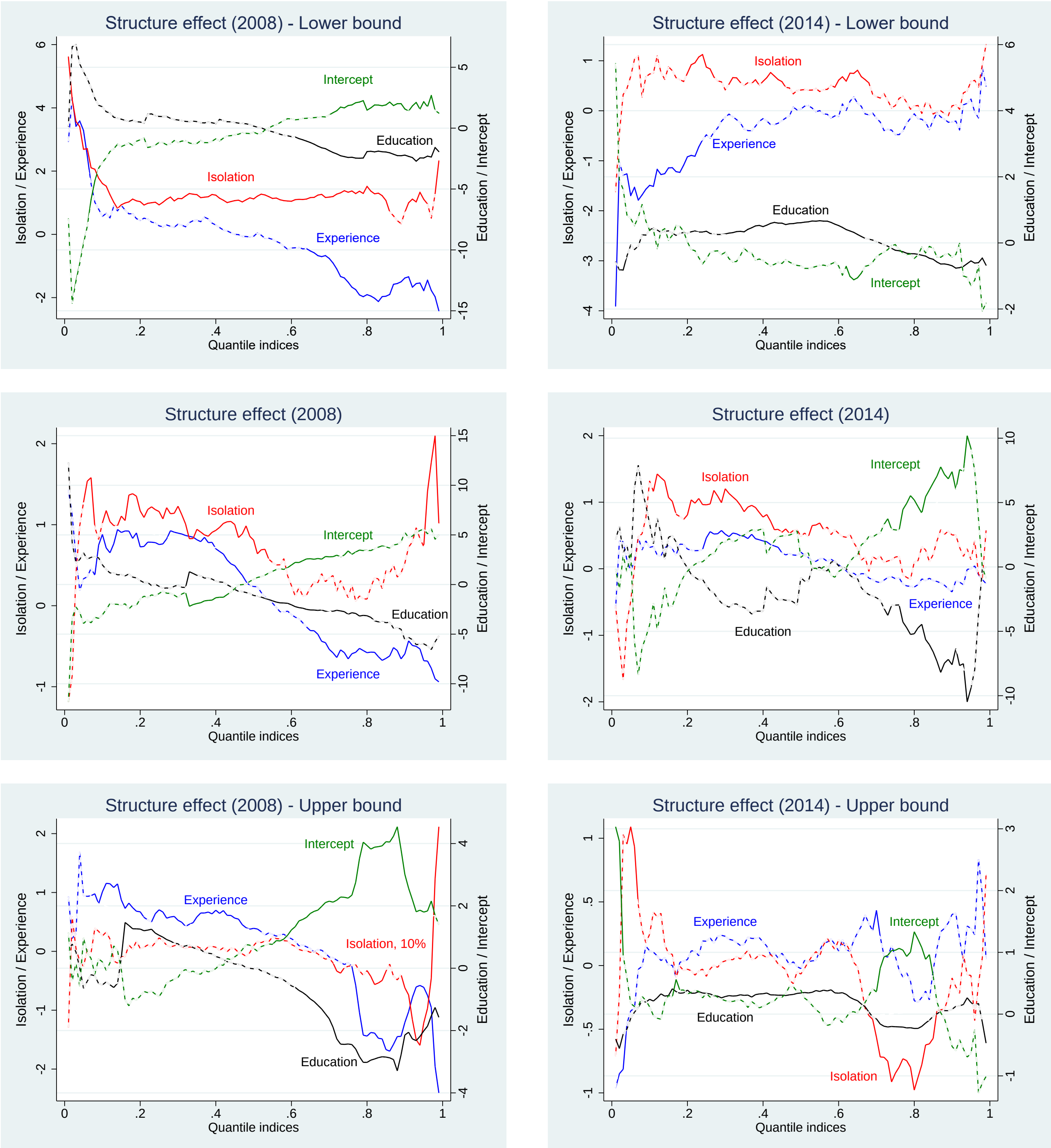

FIGURE VI

Structure Effect in 2008 and 2014 - Labor Market Selection Scenari

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level. "10%" refers to an effect significantly different from zero at the 10 percent level.

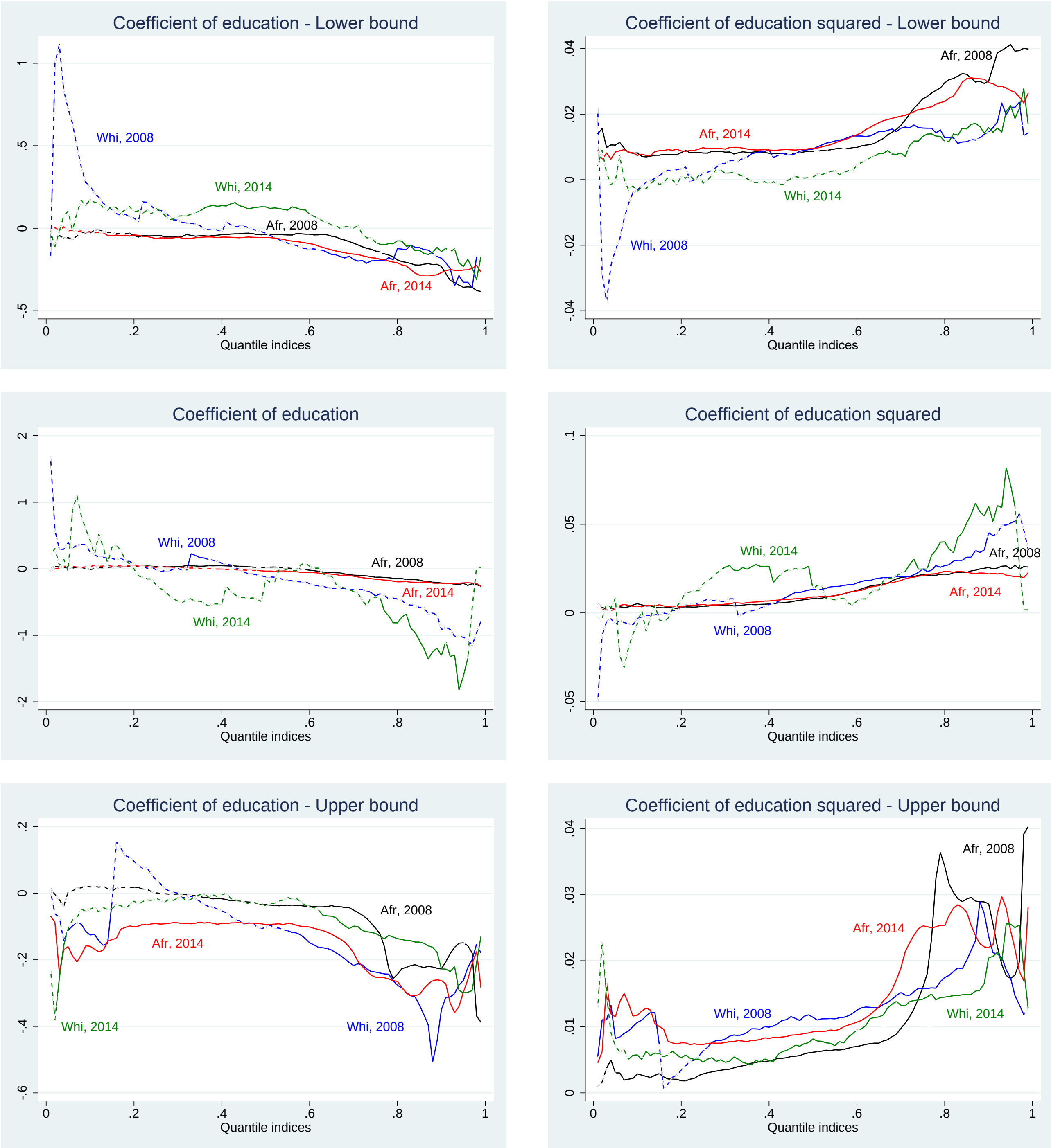

Figure VII
RIF Regressions of Income by Racial Group and Year - Education and Education Squared - Labor Market Selection Scenari

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level. "Afr" and "Whi" refer to Africans and Whites, respectively.

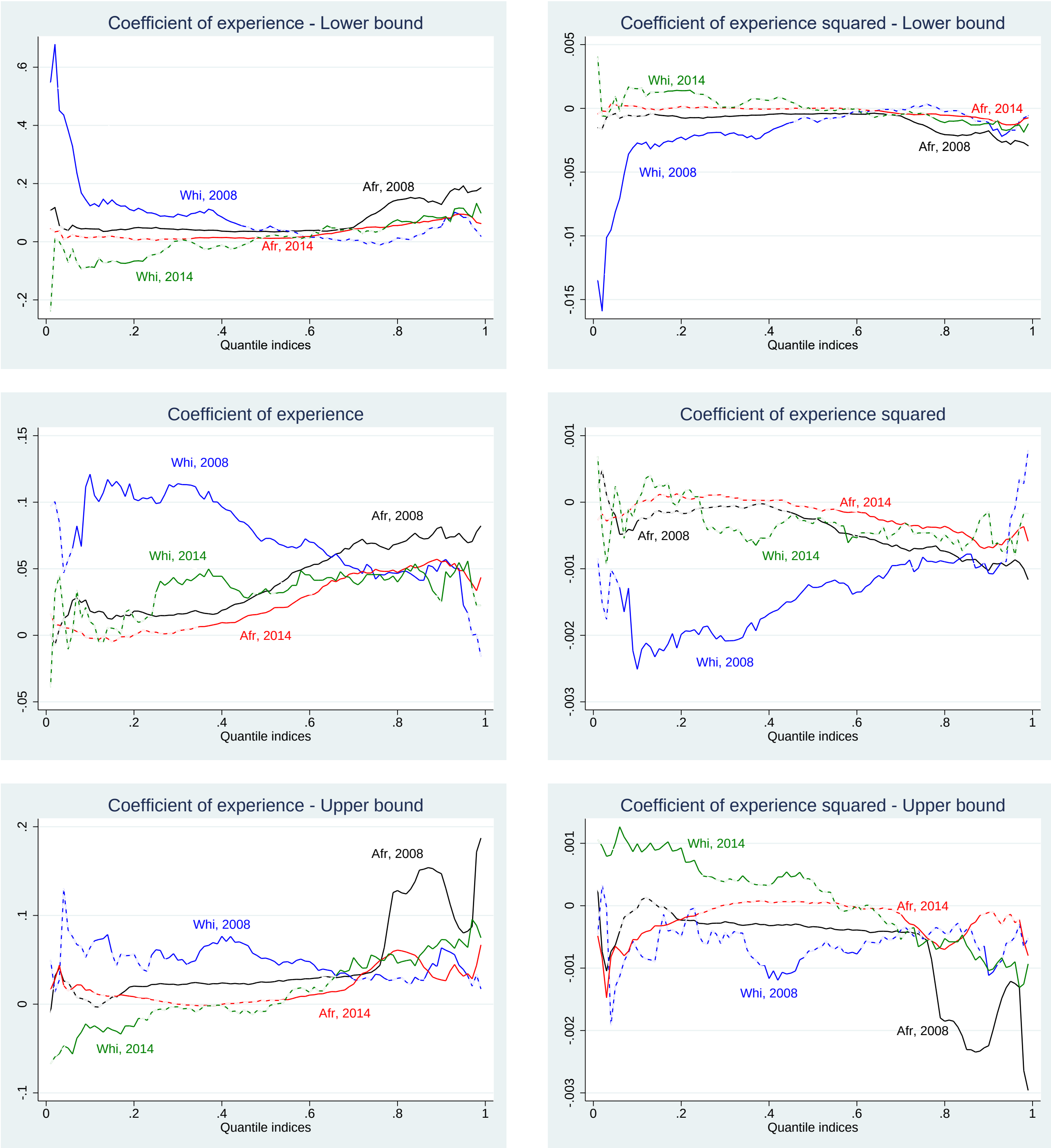

Figure VIII

RIF Regressions of Income by Racial Group and Year - Experience and Experience Squared - Labor Market Selection Scenari

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level. "Afr" and "Whi" refer to Africans and Whites, respectively.

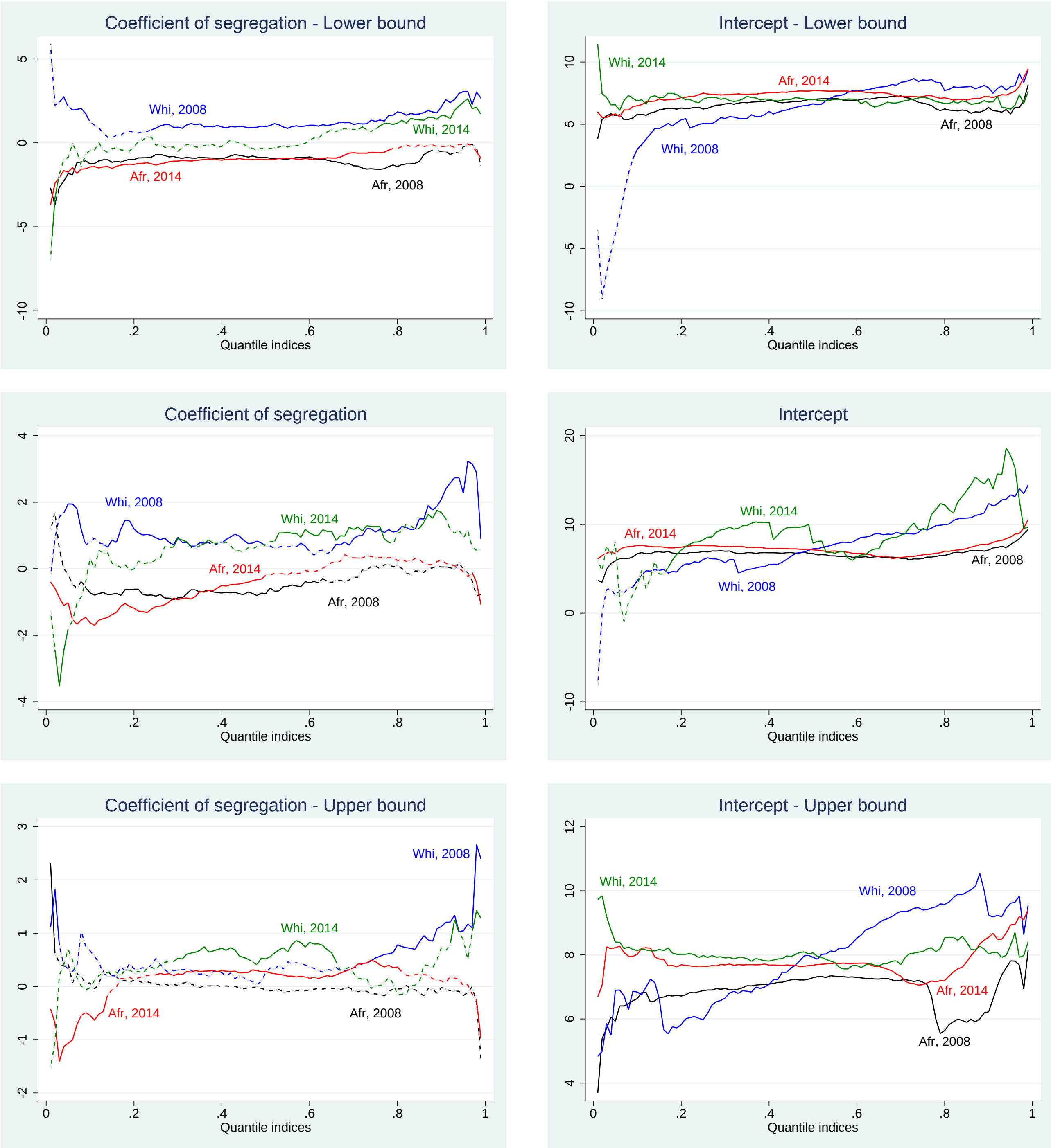

Figure IX

RIF Regressions of Income by Racial Group and Year - Isolation and Intercept - Labor Market Selection Scenari

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level. "Afr" and "Whi" refer to Africans and Whites, respectively.

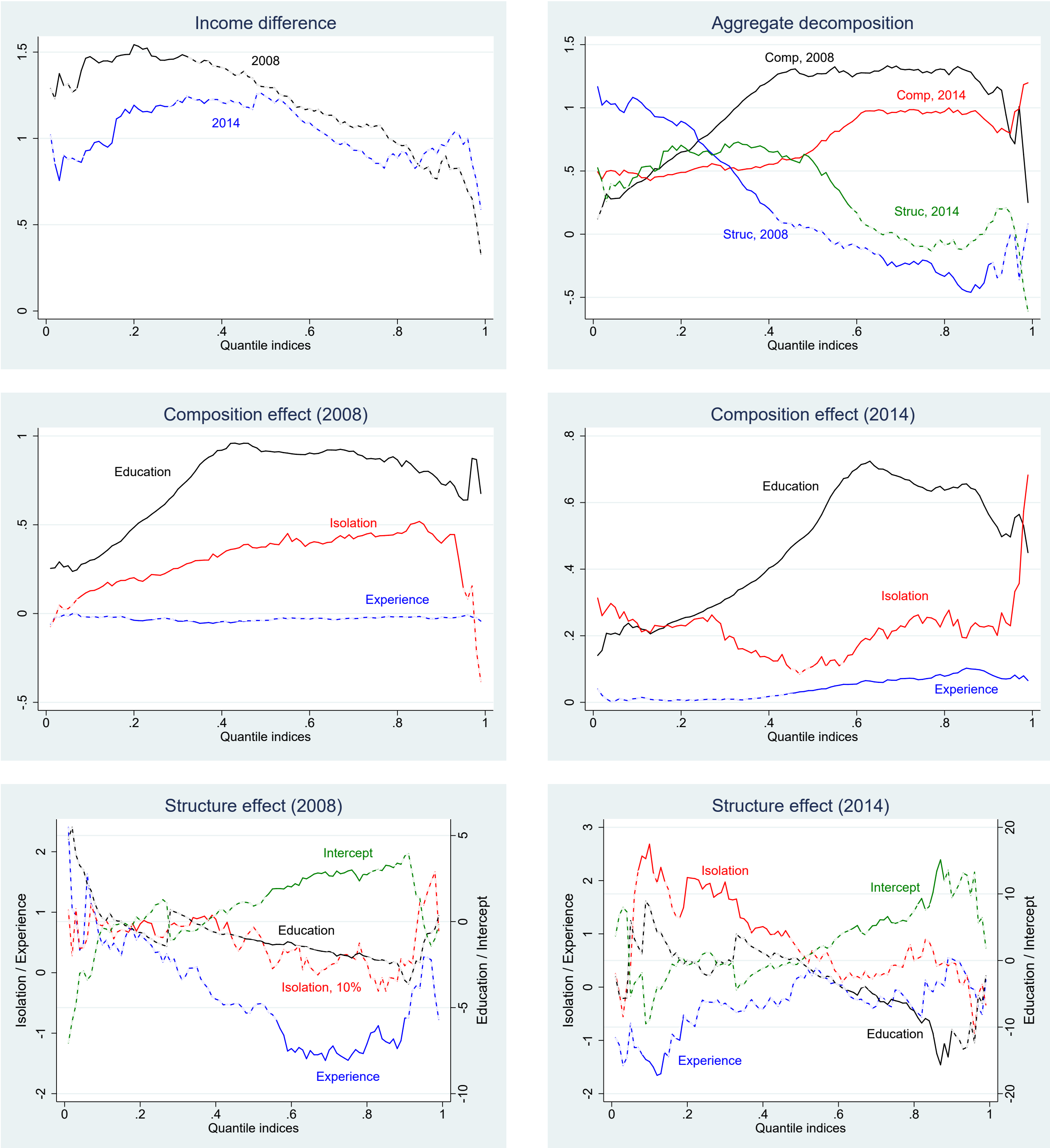

Figure X

Aggregate and Detailed Decompositions in 2008 and 2014 - Sample Selection

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level, except for the income differences, where they represent the quantiles for which the income difference in 2008 is not significantly different from the income difference in 2014. "10%" refers to an effect significantly different from zero at the 10 percent level. "Comp" and "Struc" refer to the composition effect and the structure effect, respectively.

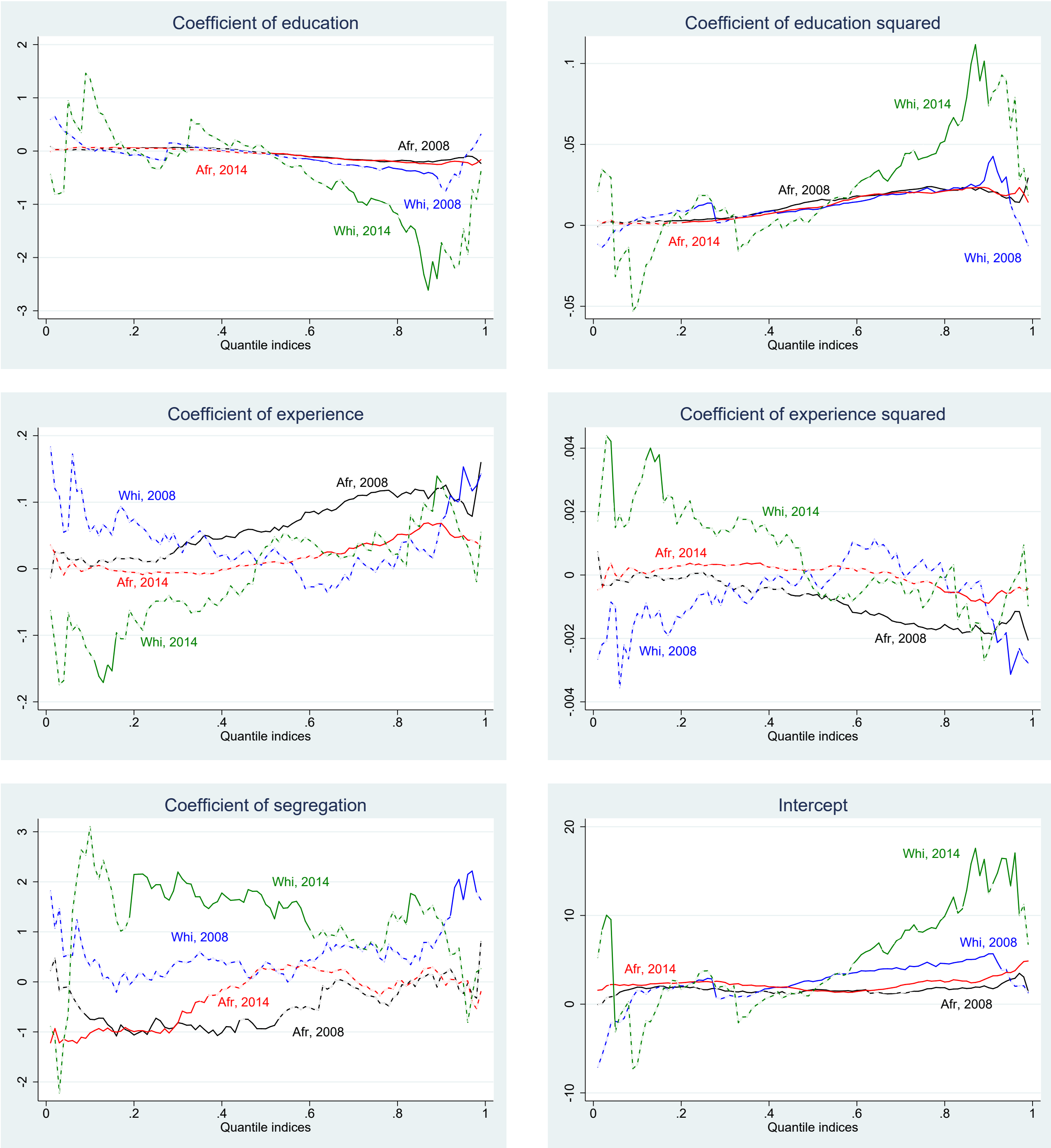

FIGURE XI

RIF Regressions of Income by Racial Group and Year - Sample Selection

The dashed parts of the curves represent the effects not significantly different from zero at the 5 percent level. "Afr" and "Whi" refer to Africans and Whites, respectively.

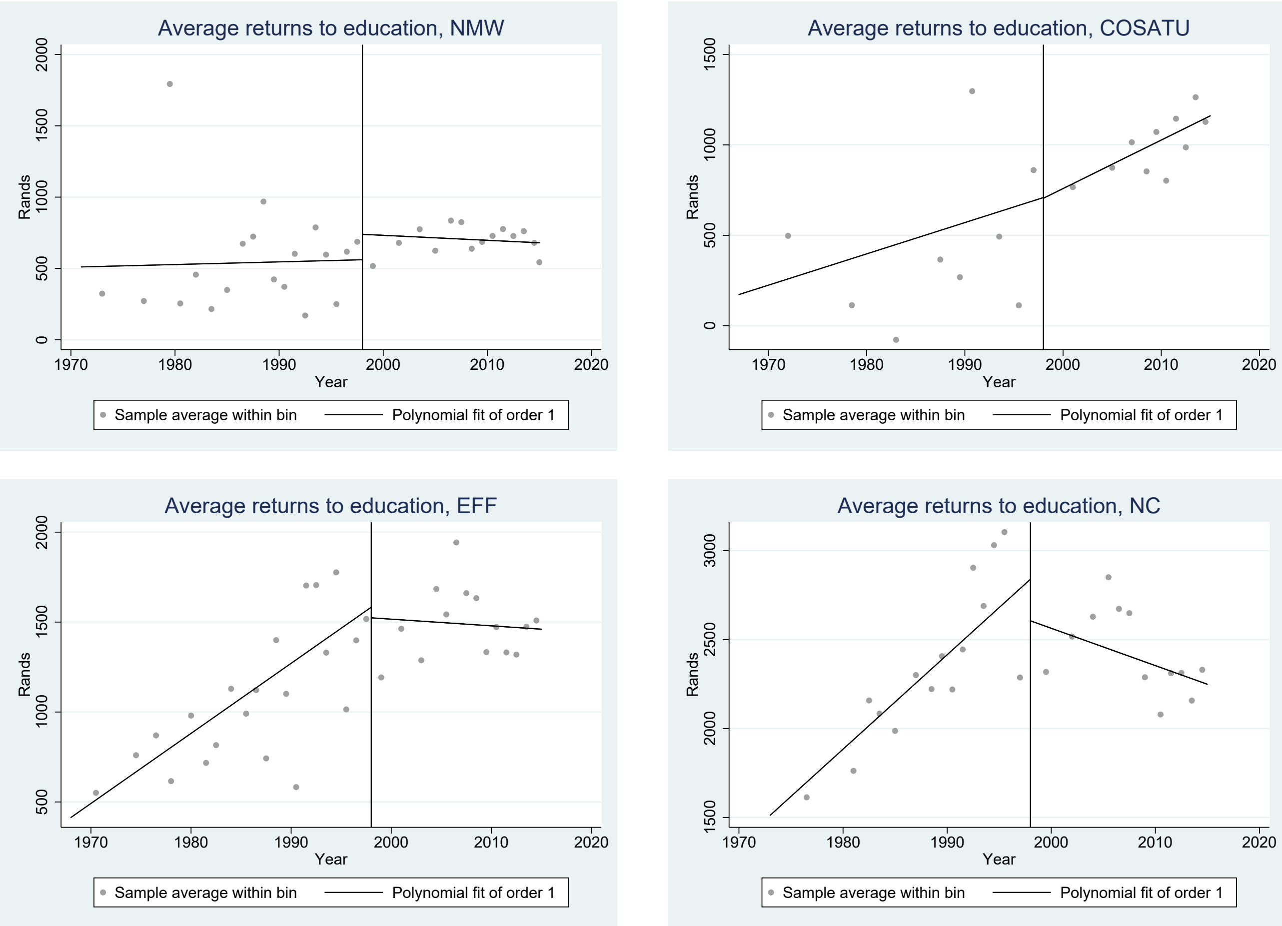

FIGURE XII

RDD plots by target group

NMW: National Minimum Wage; COSATU: trade unions; EFF: Economic Freedom Fighters; N.C.: Not Concerned.

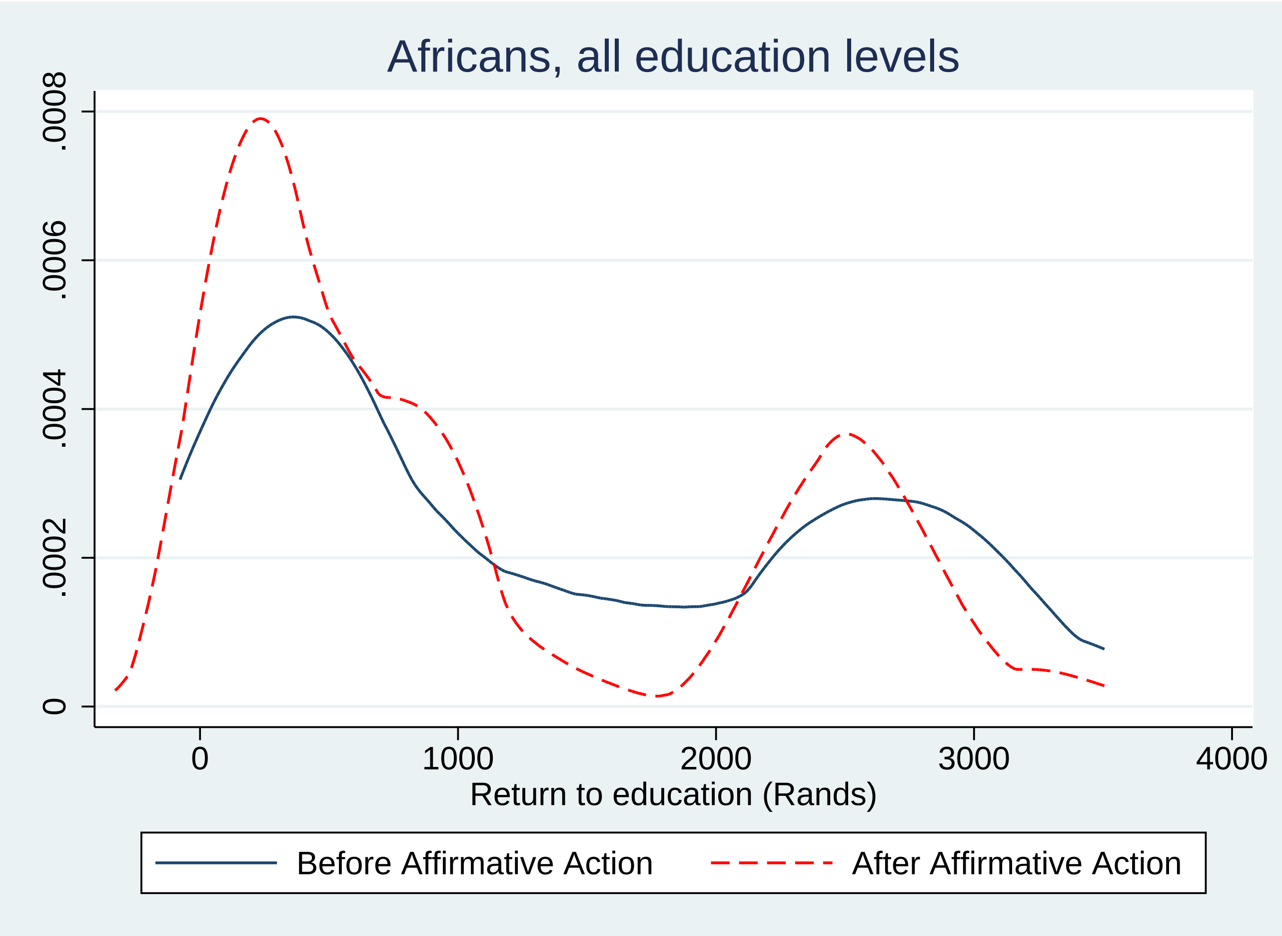

FIGURE XIII
Kernel density plot, all Africans

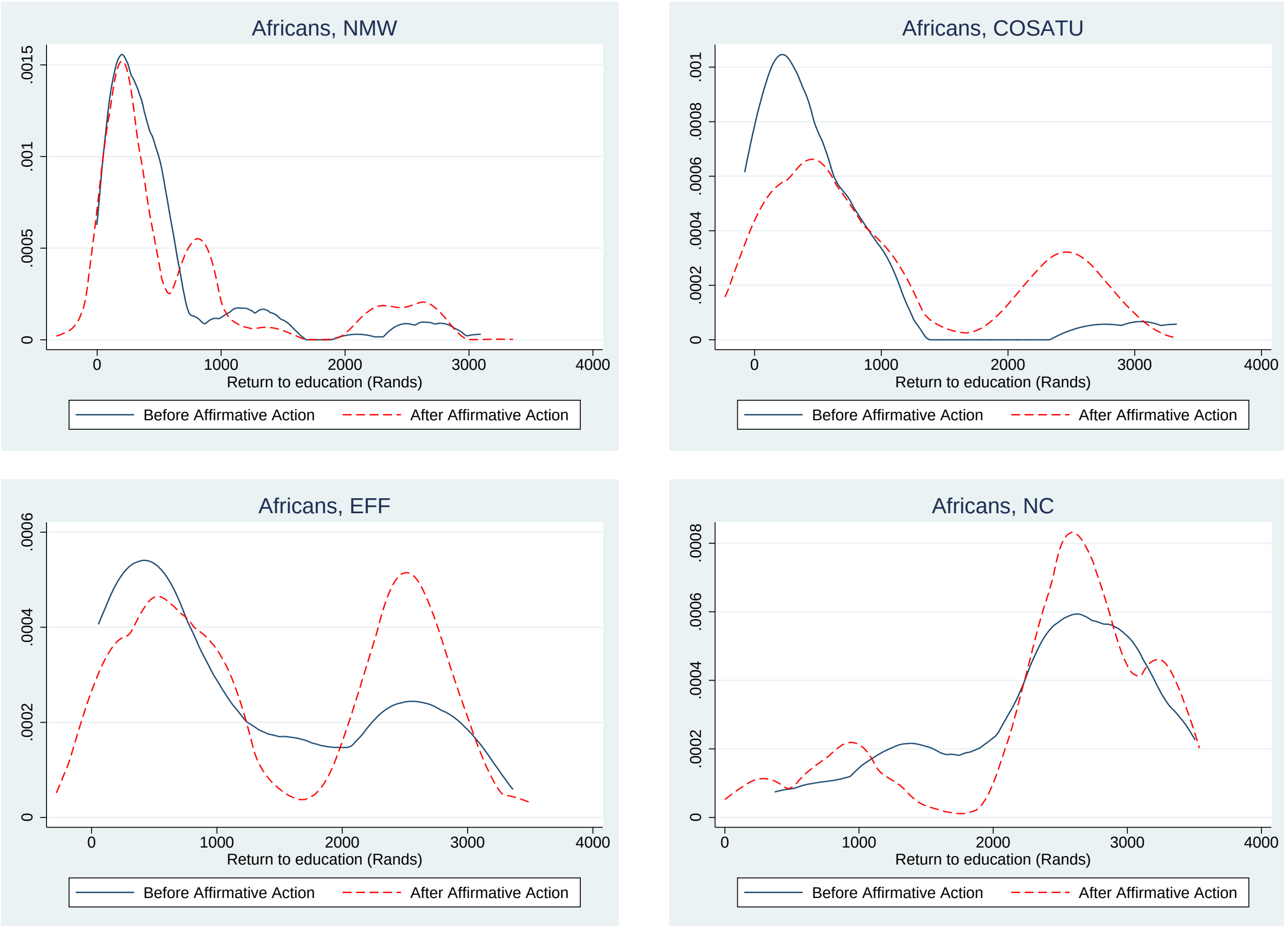

FIGURE XIV

Kernel density plots by target group

NMW: National Minimum Wage; COSATU: trade unions; EFF: Economic Freedom Fighters; N.C.: Not Concerned.

KS-statistics and their associated P-values for each target group: NMW) KS-stat=0.1706, P-value=0.002; COSATU) KS-stat=0.3041, P-value=0.01; EFF) KS-stat=0.2192, P-value<0.001; NC) KS-stat=0.1150, P-value=0.154;