

EQUALITAS Working Paper No. 44

Occupational Segregation by Race in South Africa after the Apartheid

Carlos Gradín

January 2017

# Occupational Segregation by Race in South Africa after the Apartheid

Carlos Gradín\*

Universidade de Vigo and EQUALITAS

## Abstract

This paper investigates if any substantial progress was achieved after the end of the apartheid in the high level of racial stratification of occupations in South Africa, the legacy of a political and social regime constructed on the basis of racial inequality. The empirical analysis is based on census microdata and labor force surveys. Our results do not provide strong evidence of a sustained or significant de-segregation of occupations. These remain highly segmented by race, with black Africans disproportionately holding low-paying jobs (compared with whites). Less than a third of segregation and about half of the racial stratification in the distribution of occupations are related with blacks' characteristics, especially with their lower achieved education, a gap that has been reduced over time. There remains a large unexplained part, however, indicating that segregation and stratification remain even when blacks and whites are compared with similar characteristics.

**Keywords:** occupational segregation, stratification, low-paying, apartheid, South Africa, race.

**JEL Classification:** J15, J42, J71, J82, O15, O55.

**Address:** Facultade de CC. Económicas e Empresariais, Universidade de Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Galicia, Spain. E-mail: cgradin@uvigo.es; Phone: +34 986 813527.

(\*) The paper benefited from its presentation in the *Global Conference on Prosperity, Equality and Sustainability* (New Delhi: World Bank and Institute for Human Development). I also acknowledge the financial support from the Spanish *Ministry of the Economy and Competitiveness* (ECO2016-76506-C4-2-R) and *Xunta de Galicia* (GRC2015/014).

## 1. Introduction

The life of South Africans has been dominated by racial segregation since the first Europeans arrived to the Cape in 1652 beginning the largest European settlement in the continent. The segregation of black Africans, along that of Coloured and Asians, was intensified during the apartheid, the political and social regime enforced by the National Party since it took office in 1948 until the first general democratic elections in 1994.

The ultimate aim of white rulers was to force non-whites to provide seasonal, cheap and abundant labor for farms, mines, and other sectors, while keeping economic and political power in their hands. Segregation in South Africa outstood for the range and extent of its discriminatory legislation that affected every possible sphere in life (e.g. work, education, health, transport, recreation, politics, sexual relationships).<sup>1</sup> Among them, the “colour bar” implied a job reservation for whites that excluded blacks from skilled and semi-skilled jobs, while they were deprived from an adequate education (e.g. 1953 Bantu Education Act). Segregation was also an ideology and set of practices seeking to legitimize social difference and economic inequality (Beinart and Dubow, 1995). Core elements of this segregation, such as the exclusion of blacks from skilled work (especially if they implied supervisory functions over whites), or the system of large-scale oscillating labor migration, were determined by custom as well as legislative bars.

The construction of a new de-racialized South Africa started after the end of the apartheid regime under the rule of the African National Congress. This implied the formal dismantlement of all the remaining segregative legislation, along the introduction of anti-discriminatory and affirmative policies to reverse its effects (i.e. Labour Relations Act, Employment Equity Act, Promotion of Equality and Prevention of Unfair Discrimination Act). Well rooted inequalities along racial lines, however, revealed more difficult to be removed. Especially, in the context of a sluggish economy, the result of the shrinkage of the non-mineral tradable sector since the early-1990s (Rodrik, 2008), with a chronically high level of unemployment.

The aim of this paper is to investigate the extent to which the end of the apartheid produced a sustained process of racial desegregation of the distribution of occupations, thus dismantling one of the core elements of racial inequality in South Africa. This has strong implications in the inclusiveness of black Africans as citizens, as well as in improving their material living conditions.

---

<sup>1</sup> This included the disenfranchisement of blacks (e.g. 1936 Representation of Natives Act), restrictions on their geographical mobility (e.g. pass laws), the seizure of most productive lands (e.g. 1913 Native Land Act), the imposition of hut or poll taxes, urban segregation (1923 Natives Act), etc. Feinstein (2005) provides a detailed historical account of racial segregation in South Africa.

We document the extent and nature of the segregation of black and white workers across occupations based on post-apartheid census and labor force data. For that, we first use the conventional framework based on segregation curves and indices such as Gini and Dissimilarity (Jahn, Schmid and Schrag, 1947; Duncan and Duncan, 1955).<sup>2</sup> We also analyze the vertical or ordinal dimension of segregation, measuring the extent to which the labor market is stratified by race, with black Africans being systematically segregated into low-paying occupations, using concentration curves and indices when occupations are sorted by average earnings (Gradín, 2013b, 2016). Additionally, we attempt at identifying the driving factors of this segregation at each moment of time by measuring the level conditional on workers characteristics using a counterfactual distribution in which black Africans are given the characteristics of whites (Gradín, 2013b). More precisely, we analyze if segregation is driven by workers' endowments such as the lower level of education of black Africans or their over-representation in rural and poorest provinces of the country. Alternatively, segregation might result from the labor market being intrinsically segregative among workers with similar characteristics on the basis of their race. Both sources of segregation might be the result of discrimination (actual or anticipated) but their distinction helps to better understand its nature.

In what follows, the next section briefly reviews the relevant literature. The third and fourth sections describe the methodology and data. The fifth section discusses the empirical results, which are summarized in the last section.

## **2. Race and labor market outcomes in South Africa**

As a legacy of colonialism and the apartheid, the labor market in South Africa is largely stratified by race. Some of these racial inequalities have been extensively addressed trying to measure how much progress, if any, was accomplished after the first democratic elections.<sup>3</sup> Most of the research focused so far on the magnitude of racial gaps in labor market outcomes and the extent to which they could be explained by differences in workers' productivity, especially the large differential in attained education (a gap that is analyzed in detail in van der Berg, 2007).

A primary source of racial inequality in the labor market occurs in the access to employment. Unemployment rapidly increased in South Africa, especially among blacks, during the 1990s and 2000s, when the economy was unable to absorb the growing supply of semi-skilled labor (e.g.

---

<sup>2</sup> We are aware, however, of the complexity of South African demographic groups due to the presence of other racial categories (Coloured and Asians), and the high level of heterogeneity within racial groups by other dimensions, like ethnicity, gender, or area of residence, that need to be analyzed in more detail.

<sup>3</sup> Leibbrandt et al. (2010) provide a throughout description of the situation (trends, institutions, policies) of the labor market in post-apartheid South Africa.

Kingdon and Knight, 2007; Banerjee et al. 2008). This occurred in a context characterized by labor market inflexibility and a small informal sector compared with other developing countries (e.g. Kingdon and Knight, 2007) and was exacerbated by skill-biased technical change (e.g. Banerjee et al. 2008). This large employment gap by race was largely (but not entirely) explained by the workers' characteristics of each group. For example, Kingdon and Knight (2004) found that one fifth (8 out of 34 percentage points) of the unemployment gap between blacks and whites could not be explained by their attributes in 1994. A higher unexplained term was found by Rospabé (2002) for 1993-99 or by Brokes and Hinks (2004) for 1995-2002. Paradoxically, changes in the characteristics of black South African men after the apartheid have made them more employable over time, but their propensity to be employed declined (Wittenberg, 2007).

Once workers entered the labor market, they face another source of racial inequality, the occupational distribution. This was an essential element of the segregation through job reservation policies and discriminatory practices that excluded blacks from accessing any skilled or semi-skilled job. The desegregation started before the end of the apartheid (Mariotti, 2012) with the increasing access of blacks to semi-skilled occupations previously reserved for whites between 1970 and 1980, year of the formal abolition of job reservation, although this was due to the scarcity of white workers (who were increasingly better educated). However, racial segregation continued to be strong, since whites were mostly employed in skilled jobs.

Treiman et al. (1996) estimated that the gap in occupational status (measured by the scale from the International Socioeconomic Index of Occupations) could be largely explained in 1980 and 1991 by the different characteristics of each group. Rospabé (2002) reported a gap between blacks and whites in occupational attainment (the probability of getting a high-ranking job such as manager, professional, semi-professional, or technician) of about 40 percentage points in 1993, of which 32% remained unexplained after controlling for productivity characteristics. In 1999, that gap was similar, but the unexplained part increased until 37%. Similarly, using a multinomial logit model, Treiman (2007) found a large racial penalty in occupational attainment in 1996 that vanished for the very few blacks with tertiary education. As for the reasons of continuing segregation, Keswell et al. (2013) highlighted the importance of education, showing that African female children who inherited the same level of educational opportunity as their parents were 6 or 10 percent more likely to be in the bottom of the occupational distribution than if they were exposed to better educational opportunities. Regarding the consequences, Gradín (2013a) has recently showed that the higher presence of blacks in more skilled occupations (along their improved education) has contributed to reduce the racial poverty gap after the apartheid (which was to a large extent explained by the cumulative disadvantaged

characteristics of black Africans, i.e. education, labor, demographic structure, area of residence, and family background).

We have found, however, very little research attempting to directly measure the extent and nature of occupational segregation, the main purpose of this paper. Among the exceptions, Campbell (1987) reported a steady Dissimilarity index of about 0.540 during apartheid years (1970, 1980 and 1985), a time of rapid industrialization that increased the access of blacks into professional and technical occupations.<sup>4</sup> More recently, in the context of an analysis of sex segregation, Parashar (2008) reported a Dissimilarity index of white-black segregation (2-digit classification of occupations) of 0.572 in 2001 using census data. She also reported that this segregation was larger among women (0.580 vs. 0.512 for men), and in Free State and Western Cape (compared with Gauteng). She highlighted the fact that South Africa, unlike the US, outstands for showing higher segregation by race than by sex.

More attention has received the huge earnings gap found between black and white South Africans. For example, Allanson et al. (2000) reported that one third of that gap in 1994 remained unexplained after controlling for differences in productivity, and Allanson et al. (2002) found no immediate improvement after that date. Keswell (2010), however, reported a modest decline in the white-black wage differential between 1993 and 2002 with an increasing importance of differences in the returns to education, from virtually nothing to account for about 40% of the gap. This finding points to the increasing importance of differences in quality of education received by each population group, while the gap in the quantity of years of education was reduced. Consistently, Rospabé (2002) also reported an increase in the unexplained part of a shrinking gap between 1993 and 1999.

### **3. Methodology**

The conventional framework to measure segregation of two groups across occupations uses the segregation curve and indices such as Dissimilarity or Gini (Jahn, Schmid and Schrag, 1947; Duncan and Duncan, 1955). Gradín (2016) expanded this framework to take into account the extent to which the segregation of one group (blacks) involves their workers holding the lowest-paying jobs using the concentration curve and indices derived from it. We also follow Gradín's (2013b) approach to identify the level of segregation that can be explained by differences in the attributes of workers of each race, and the level of segregation that remains unexplained when

---

<sup>4</sup> He cited an unpublished report from the US Bureau of the Census and the Commission on Civil Rights, and it is not clear how detailed the occupational classification was.

both groups have the same characteristics, in line with previous research on the employment and earnings gaps.

### 3.1 Measuring segregation

We compare the employment distribution across  $J$  occupations of workers from a target or comparison group (i.e. black Africans, labeled as  $i = c$ ) and another one, the reference distribution (i.e. Whites, denoted by  $i = r$ ). The vector of relative frequencies is  $f^i = (f_1^i, \dots, f_J^i)$ , where  $f_j^i$  is the proportion of workers from group  $i$  in occupation  $j$ , when occupations are sorted in ascending values of the relative share of members of the reference group (i.e.,  $f_j^r / f_j^c$ ).  $F_j^i = \sum_{s=1}^j f_s^i$  indicates the corresponding cumulative distribution value. The objective is to assess the extent to which each population group tends to be concentrated in a different subset of occupations, and how this changed over time. For that, we first compare each year's segregation curve, and then quantify the amount of segregation at each moment using specific segregation indices.

The segregation curve  $F^r(p), p \in [0,1]$  plots the cumulative proportions of workers for the comparison ( $F_j^c$ ) and reference ( $F_j^r$ ) groups for the  $j$ th occupation with largest underrepresentation of the reference group, connected with linear segments. The 45° line indicates the case of no segregation (both groups have the same employment distribution across occupations). The segregation curve goes along the abscissa and then shifts to 1 at  $p = 1$  in the case of maximum segregation (both groups working in different occupations). If the segregation curves of two distributions (i.e. years) do not intersect, the one with the curve falling below exhibits higher segregation (upon agreement on only four basic properties; Hutchens, 2004).<sup>5</sup> A large set of segregation indices (including Gini, and Generalized Entropy and Atkinson families) will rank them consistently. However, if the curves do intersect, we cannot rank them without agreeing on additional properties, and those measures can produce different rankings, depending on the degree of sensitivity of the index to disequalizing movements at different points of the distribution. For the sake of simplicity, our results will rely on the computation of two indices of segregation.

The *Dissimilarity* index,  $D$ , can be defined as half the sum of discrepancies in the population shares of each group by occupation:

---

<sup>5</sup> A measure of segregation should verify Homogeneity, Symmetry, Principle of (disequalizing) movements between occupations, and Insensitivity to proportional divisions.

$$D(f^c, f^r) = \frac{1}{2} \sum_{j=1}^J |f_j^c - f_j^r| = \max_{j \in [1, J]} \{F_j^c - F_j^r\}, \quad (1)$$

Geometrically,  $D$  is the maximum vertical distance between the diagonal and the segregation curve. This occurs at the critical occupation  $q$ , defined so that the comparison group is overrepresented below and underrepresented above:  $D(f^c, f^r) = F_q^c - F_q^r$ , where  $q = \max_{j \in [1, J]} \{j \mid f_j^c \geq f_j^r\}$ .  $D$  can be interpreted as the proportion of workers of any group that should change occupation (from those in which their group is overrepresented to those in which it is underrepresented) to achieve full integration.

The *Gini* index can be defined as the area between the segregation curve and the diagonal (divided by its maximum,  $\frac{1}{2}$ ) and thus written as the weighted sum of these vertical distances computed at the midpoints between adjacent occupations:<sup>6</sup>

$$Gini(f^c, f^r) = 2 \sum_{j=1}^J (\hat{F}_j^c - \hat{F}_j^r) f_j^c, \quad (2)$$

$$\text{where } \hat{F}_j^i = \frac{1}{2}(F_{j-1}^i + F_j^i) = F_{j-1}^i + \frac{1}{2}f_j^i, \text{ and } F_0^i = 0.$$

*Gini* ranks distributions consistently with non-intersecting segregation curves. However,  $D$  is consistent only in a weak sense (it will never rank two distributions in the reverse order), because it is insensitive to any disequalizing movement that occurs between occupations above or below  $q$ .  $D$  is the *Gini* between two sets of occupations (those dominated by each race). The main contribution of *Gini* is that it also takes into account segregation within these two large sets of occupations. Both indices vary between 0 (no segregation) and 1 (full segregation) and are symmetric in population groups (it is irrelevant which group is the comparison and which the reference).

### 3.2 Segregation into low-paying occupations

Gradín (2016) adapted the previous approach to measure the extent to which one group, black Africans, tends to be systematically overrepresented in low-paying jobs. This implies stratification in occupations (also referred to as vertical or ordinal segregation). The approach basically consists in re-ranking the distribution of occupations by a measure of its quality (e.g.

---

<sup>6</sup> Note that  $\hat{F}_j^i$  are used instead of  $F_j^i$  due to the fact that we connect the segregation curve by linear segments while define the cdfs as step functions.



the average earnings  $w_j$ ). We call  $g_j^i$  and  $G_j^i$  the relative frequency and cumulative frequency of workers from group  $i$  in occupation  $j$  in this re-ranked distribution.

The concentration curve  $G^r(p), p \in [0,1]$  plots the cumulative proportion of workers from both groups (with occupations sorted by  $w_j$ ):  $G_j^c$  in the horizontal axis and  $G_j^r$  in the vertical one (connected by linear segments). The target group  $c$  is segregated into low-paying occupations (compared with group  $r$ ) if the proportion of workers from this group is larger (or equal) below any reasonable low-pay threshold. This means that the concentration curve falls below the diagonal ( $G_j^c \geq G_j^r$ ) over the target range, and that there is first-order stochastic dominance of  $r$  over  $c$ .<sup>7</sup> If there is no segregation or if the labor market is segmented but with both groups in occupations providing a similar pay, the labor market is not stratified. The labor market is stratified when workers from one group are segregated into occupations that systematically tend to pay less.

The values of the concentration curve are bounded from below by the segregation curve (when all segregation is into low-paying jobs) and from above by its mirror image above the diagonal (when the segregation of the comparison group is into high-paying occupations). The actual values of the curve depend on the correlation between the employment distribution using the two alternative ranks of occupations (sorted by earnings and by racial ratios). If segregation is pay-neutral, the concentration curve will go along the diagonal. Whenever the concentration curves of two distributions (i.e. years) do not overlap, we can say that in the one with the curve falling above the comparison group exhibits less segregation into low-paying occupations.

We use the concentration versions of the Gini and Dissimilarity indices, obtained by replacing  $f$  by  $g$  in the second part of expression **(1)** and in **(2)**, to quantify segregation into low-pay and to rank distributions accordingly when the concentration curves overlap. The Gini concentration index is twice the area (positive or negative) between the diagonal and the concentration curve, and corresponds to the index of vertical segregation proposed by Blackburn and Jarman (1997) based on Somers' (1962) measure of statistical association. The Dissimilarity concentration index is the maximum vertical distance (positive or negative) between the diagonal and the concentration curve, and measures the proportion of workers of each group that should change occupation in order to eliminate segregation into low-paying (high-paying) occupations for any possible low-pay threshold. Each concentration index is bounded between the corresponding segregation index (when all segregation of the comparison group is into low-paying jobs) and its

---

<sup>7</sup> By replacing  $\geq$  by  $\leq$  or by  $=$  one can similarly define segregation into high-paying occupations and neutral-pay segregation.

negative value (all segregation is into high-paying jobs). Thus, the index falls in the range between -1 and 1, with the extremes requiring full segregation. A positive (negative) sign indicates predominant segregation of the comparison group into low-paying (high-paying) occupations. If the distribution is pay neutral, the concentration indices will be zero. We will compute standard errors for segregation and concentration indices using bootstraps.

We can also define concentration ratios as the proportion of observed segregation of the comparison group that is low-paying (or high-paying), by normalizing each concentration index by its maximum value (the segregation index), with the sign still indicating whether the comparison group tends to be segregated into low- or high-paying occupations,  $r_S = \frac{S(g^c, g^r)}{S(f^c, f^r)}$ ,  $S = Gini, D$ . In particular,  $r_{Gini}$  is the *Gini correlation coefficient* between groups' ratio and average earnings across occupations, computed among members of group  $c$ .<sup>8</sup> These concentration indices (and ratios) are symmetric in their absolute values. Exchanging group labels (who is the reference and who is the comparison group) will just change their sign.

### 3.3 Measuring conditional segregation

The observed level of segregation might be the result of the distribution of relevant characteristics differing across population groups. We follow here Gradín (2013b, 2014), who adapted the DiNardo et al.'s (1996) procedure for the decomposition of the interdistributional wage differentials to the measurement of segregation. To disentangle what part is driven by differences in observable characteristics (explained or compositional effect) and what part is conditional segregation of workers with similar characteristics on the basis of race, we construct a counterfactual distribution  $F_Y^c$ . In this counterfactual, individual observations of the comparison group (black Africans) are reweighted to reproduce the same distribution of characteristics of the reference (whites). The reweighting factor is the odds of being white conditional on characteristics using a logit regression.<sup>9</sup> In this counterfactual, both races will exhibit the same distribution of types (workers with any given combination of characteristics,

---

<sup>8</sup> This is a measure of association that uses the Gini covariance as a measure of variability (instead of the conventional covariance), whose properties are a mixture of Pearson's and Spearman's correlations (see Schechtman and Yitzhaki, 1987, 1999, Yitzhaki and Olkin, 1991).

<sup>9</sup> That is, the reweighting factor is  $\Psi_x = \frac{f^r(x)}{f^c(x)} = \frac{N^c Pr(r|x)}{N^r Pr(c|x)}$ , where  $Pr(r|x) = \frac{\exp(x\beta)}{1+\exp(x\beta)}$  and  $Pr(c|x) = 1 - Pr(r|x)$ ; while  $x$  is the vector of characteristics and  $\beta$  the estimated coefficients.  $\frac{N^c}{N^r}$  is a constant indicating the groups' population ratio.

such as holding a university degree, living in urban Western Cape, ...), but each race keeps its own distribution across occupations conditional on type.<sup>10</sup>

This flexible semi-parametric approach allows us to obtain the aggregate decomposition of any unconditional segregation index  $S(f^c, f^r)$  into explained and unexplained terms:

$$S(f^c, f^r) = S^E + S^U = [S(f^c, f^r) - S(f^y, f^r)] + S(f^y, f^r). \quad (3)$$

Where  $S^E = [S(f^c, f^r) - S(f^y, f^r)]$  is the level of segregation explained by both population groups having different distributions of characteristics (types of workers).  $S^U = S(f^y, f^r)$  is the unexplained or conditional segregation that remains after equalizing the distribution of types in both groups, depending on how much the labor market segregates (on the basis of race) people with similar observed characteristics. The identification of the unexplained term with discrimination in the labor market, however, has to be cautious, like in the analysis of wage or employment discrimination, because it may also reflect racial differences in unobserved characteristics (e.g. job preferences, quality of education, unobservable skills). Similarly, the explained part could also reflect anticipated discrimination in the labor market by the disadvantaged group (leading, for instance, to lower investment in human capital or influencing their migration patterns), apart from reflecting pre-labor market discrimination (such as in the access to education).

A detailed decomposition of the explained segregation term will allow to identify the main determinant factors.<sup>11</sup> Given the non-linear nature of the approach, this is not straightforward, however. Starting with the case in which all estimated coefficients in the logit regression are set to zero, we produce a sequence of reweighting factors consecutively switching the coefficients of each set of characteristics to its estimated value, finishing when all coefficients are changed. The contribution of each set of characteristics would be the change in segregation after their associated coefficients were switched on, but this procedure has a path-dependency problem. To avoid that, we obtain the contribution of each factor after averaging over all possible

---

<sup>10</sup> Another counterfactual is possible in which blacks keep their own distribution of characteristics, but are given the conditional employment distribution of whites (obtained by reweighting the reference distribution instead). However, given that blacks exhibit characteristics (such as lower attained education) that constraint their occupational opportunities, the other counterfactual seems to be more reasonable as one expects black Africans to eventually converge with whites in education and other attributes (and not otherwise).

<sup>11</sup> With this methodology it is not possible to obtain a detailed decomposition of the unexplained effect (which in any case would be subject to a serious identification problem).

sequences (Gradín, 2014, using a Shapley decomposition: Chantreuil and Trannoy, 2013; Shorrocks, 2013).

The same exercise is done with segregation into low-paying occupations, after replacing  $f$  by  $g$  in (3).

#### 4. Data

Censuses are conducted in South Africa since 1911, but those prior to the 1994 democratic elections (the last one in 1991) are problematic, especially regarding the African population. The 1996 Census was the first one covering the entire country and treating all populations groups equally (e.g. StatsSA, 2007). Our main empirical analysis uses microdata samples from post-apartheid 1996 and 2001 Censuses, and the 2007 Community Survey from Statistics South Africa, harmonized by the Minnesota Population Center in its Integrated Public Use Microdata Series (IPUMS-I, Minnesota Population Center, 2015). Unfortunately, the most recent 2011 Census did not code the information about occupation. The use of census data guarantees larger samples from which to analyze segregation across a more detailed classification of occupations, while providing the required information related to workers' characteristics.

Race is considered using the classification that comes from the apartheid. Whites are those with European ancestry (mostly Dutch and British), making 9% of the population in 2007 (16% of workers). Black Africans make the largest population group (80% of the population and 69% of workers in 2007) and are mostly the descendants of the immigration of Bantu farmers from the Great Lakes region in East Africa into eastern areas of South Africa since the third century. This racial category comprises different ethnicities from the Nguni (e.g. Xhosa, Zulu), Sotho-Tswana and other minor linguistic branches. The other nonwhite categories (Coloured and Asians) are not analyzed here.

For the sake of inter-temporal comparability, the final sample is composed of white and African individuals not living in group quarters<sup>12</sup>, 15-65 years old, who are employed, not in the Armed Forces.<sup>13</sup> This implies a total of 1,414,812 individual observations with the following distribution

---

<sup>12</sup> We included in the sample those with group quarter status unknown in 1996. IPUMS-I reports that 17 districts in Easter Cape were not organized into households in that census.

<sup>13</sup> The universe for occupational variables is employed persons at least 15 years old in private households in 1996, employed or economically active persons at least 10 years old in 2001, and persons aged 15 to 74 years old with a job last week, not in institutions in 2007. Employment status is defined for 15-65 year-old people in 2001 and 2007 (15 or older in 1996) and refers to the time of the census (1996) or the reference week (2001 and 2007). For comparability issues across these three datasets, see Yu (2009).

by year and race: 630,350 (166,560 whites, 463,790 blacks) in 1996; 590,227 (139,085 whites, 451,142 blacks) in 2001; and 194,235 in 2007 (33,268 whites, 160,967 blacks).

Our main results use the IPUMS harmonized 3-digit International Standard Classification of Occupations (ISCO-1988)<sup>14</sup>, with 125 categories, including one for those with occupation not classified elsewhere or unknown, which is problematic given its large importance, especially in 2007 (16% compared with around 7% in the previous years). For robustness, we also produced results for IPUMS 1 and 2-digit harmonized classifications (with 10 and 37 categories respectively), and for all three classifications with those reporting unknown occupation removed from the sample.

Earnings for each occupation will be approximated using contemporary average income using person's annual income in Rands for the twelve months prior to the census.<sup>15</sup> Given the strong stratification of the South African labor market, it is not straightforward whose income should we consider to rank occupations. We present our main results using the average calculated over the entire population, although we also estimated the alternative using the black population only.<sup>16</sup>

Workers' characteristics used to estimate conditional segregation were defined as follows. Location includes area (urban or rural) and province (Western Cape, Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, North West, Gauteng, Mpumalanga, and Limpopo). Educational attainment distinguishes no schooling, some primary, primary (6 years), lower secondary, secondary, university, other education, and unknown education. Immigration is measured by immigrant status (no immigrant, national immigrant, immigrant from abroad) and years residing in current dwelling. Other demographic variables include: age intervals (15-24, 25-34, 35-44, 45-54, and 55-64 years old), sex, marital status (single, never married, or unknown; married or in consensual union; separated, divorced, or spouse absent; widowed), household head, spouse, and disabled statuses.

The empirical analysis is also based on the South Africa - Post Apartheid Labour Market Series (PALMS v3.1, Kerr, Lam, and Wittenberg, 2016) 1994-2015, from DataFirst portal at University of Cape Town, to have more detailed information over time and for the sake of robustness.

---

<sup>14</sup> It differs from the original unharmonized classification (161 categories) because some small occupations have been aggregated by IPUMS into one single category within the same group.

<sup>15</sup> Values recoded by IPUMS to the midpoints of the broad intervals given in the original data, with the top interval coded to its lowest possible value. The average was preferred in this case to the median due to the high probability of ties when income is reported in intervals.

<sup>16</sup> The Spearman (rank) correlation between black and white average incomes across occupations was 46% in 1996, 59% in 2001 and 46% in 2007.

PALMS combines different Statistics South Africa surveys: the annual October Household Surveys (OHS 1994-1999), the biannual Labour Force Surveys (LFS 2000-2007), and the Quarterly Labour Force Surveys (QLFS 2008-15). The sample consists of 1,017,093 observations (855,882 blacks, 161,211 whites), with 46,232 observations/year on average (38,904 blacks and 7,328 whites), but with great variability across years: from a minimum of 11,040 in 1996 (8,891 blacks and 2,149 whites) to a maximum of 83,227 in 2008 (71,036 blacks and 12,191 whites).<sup>17</sup> We also used ISCO-1988 occupations at 1, 2 and 3-digit classifications, even if the smaller sample sizes, compared with census data, impose some cautions about the more detailed results. PALMS also have an estimation of real earnings that will be used to rank occupations, after some adjustments.<sup>18</sup>

## **5. Segregation and stratification trends after the apartheid**

### **5.1 Trends in unconditional occupational segregation**

We start the analysis using the segregation curves to check if it is possible to identify a clear and robust trend in occupational segregation by race with census data (Figure 1.a). The 2007 curve entirely falls above the corresponding curve in 1996, what means that upon agreement on only four basic principles, one can say that there was a unambiguous decline in segregation that will be confirmed by most indices of segregation. However, the story is different by sub-periods.

The 2001 segregation curve generally falls below that of 1996, except at the top decile of occupations with the largest overrepresentation of whites. This means that segregation increased around occupations already disproportionately filled by black Africans between 1996 and 2001, although there was some desegregation in occupations with larger shares of whites. These intersecting curves imply that segregation increased by all consistent indices unless they put a larger weight on predominantly white occupations. We, however, find a robust reduction in segregation between 2001 and 2007, no matter what index we use, because the 2007 curve entirely falls above that of 2001.

---

<sup>17</sup> For intertemporal consistency, we deleted 2,090 observations with unknown occupation (most from 1996 and 1997). Observations are weighted using the cross Entropy weight derived by DataFirst from 2008 ASSA demographic model. For details of these data, see Kerr and Wittenberg (2016).

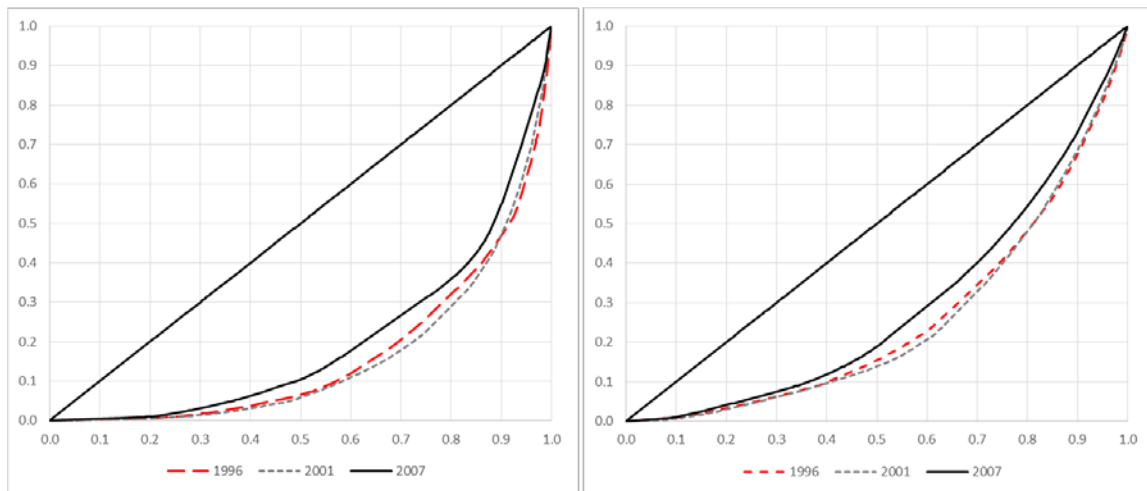
<sup>18</sup> PALMS constructed an inflation adjusted labor earnings variable. There are some outliers and a large number of observations with missing earnings. The latter affects all occupations in various years (1996, 2008-9, and 2015), and some occupations using detailed classifications in others. For that, we used the median (instead of the mean) and imputed it for occupations lacking such information using earnings in the previous available year. The series may be affected by changes over time, especially from OHS to LFS in 1999-2000 (see Wittenberg, 2014).

The use of indices allows to quantify the intensity of segregation in each year (Figures 2.a-b; Tables 2 and 3). The increase of segregation between 1996 and 2001 was about 2 percent (Gini) or 6 percent ( $D$ ), while the desegregation between 2001 and 2007 was about 13 or 16 percent, respectively. The net reduction for the entire 1996-2007 period was of near 11 percent with both indices, but the remaining level of segregation in 2007 was still large, with a Gini of 0.599, 74% of which was between occupations dominated by each race ( $D=0.442$ ). This trend in segregation is similar with the 1 and 2-digit classifications. If we remove observations with occupation unknown or not classified elsewhere (instead of considering this group as an independent category), then the reduction in segregation (3-digit) is smaller: about 8% between 2001 and 2007, 3-7% for the overall period.

**Figure 1. White-African segregation curves**

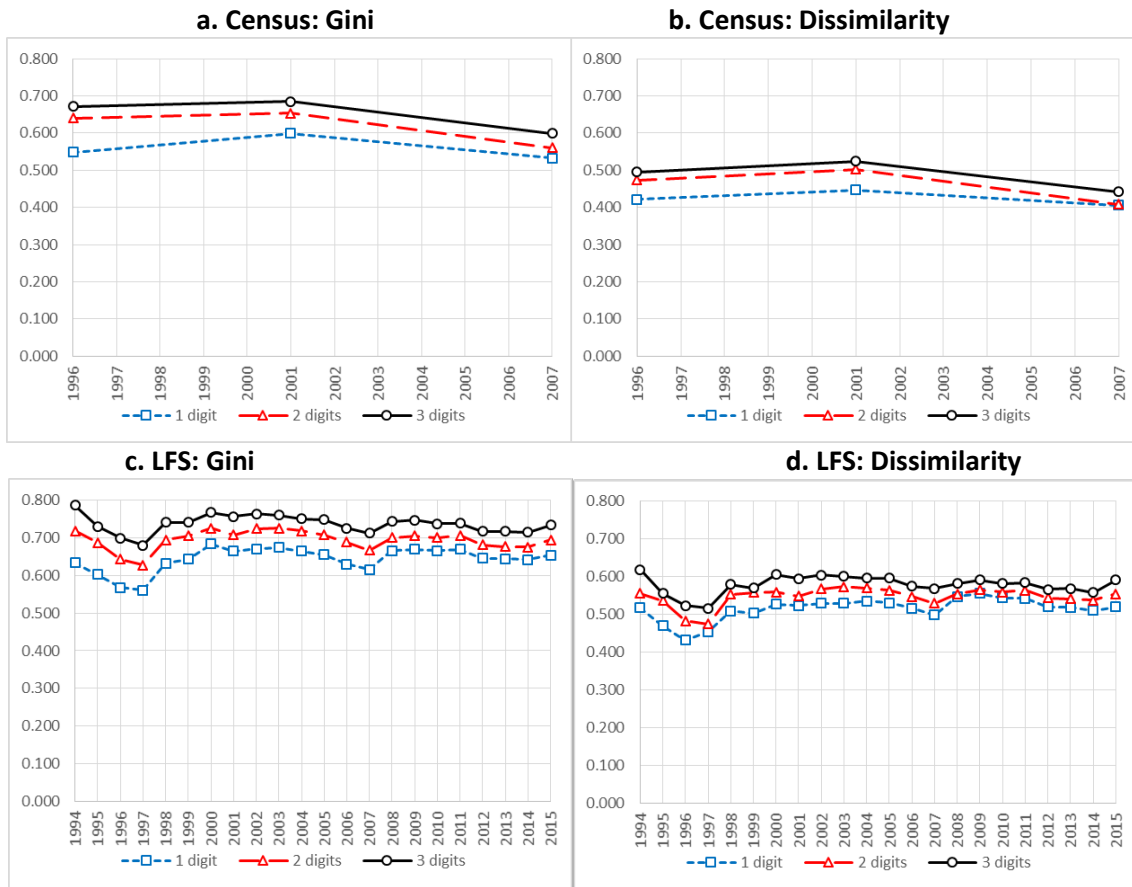
**a. Observed**

**b. Unexplained**



Source: Own construction based on IPUMS-International (3-digit classification).

**Figure 2. White-African segregation indices**



Source: Own construction based on IPUMS-International (census) and PALMS (LFS).

The labor force surveys reflect an even more pessimistic trend in segregation for 1994-2015 (Figures 2.c-d). We can distinguish an initial intense decline in segregation between 1994 and 1997, right after the end of the apartheid. This decline was followed by various oscillations according to the business cycle.<sup>19</sup> It is, however, discouraging to find out that segregation in 2015 was still similar or only slightly below the level achieved right at the end of the apartheid, and substantially above that in 1997 (at least 8% and 15% with Gini and *D*).<sup>20</sup>

## 5.2 The segregation of black Africans into lower-paying occupations

We now address the issue of the quality of occupations held by black Africans, by looking at the concentration curves and indices. Using census data, Figure 3.a shows that black Africans are disproportionately over-represented in lower-paying jobs because each year's curve falls below the diagonal (implying first-order stochastic dominance along the entire occupational

<sup>19</sup> The correlation between unemployment rates and segregation (3 digits) after 1997 is 67-72% (Gini and Dissimilarity).

<sup>20</sup> The results with labor force surveys corroborate the 1996-2001 increase and the 2001-2007 decrease found using census data, but with different intensities, such that the 1996-2007 period shows a net increase (between 2-8% with Gini, 9-15% with *D*, depending on the classification used).



distribution). Between 1996 and 2001 there was an increase in the segregation of blacks into low-paying occupation for almost the entire range of earnings, although the concentration curves cross at the 97<sup>th</sup> percentile of black workers (this implies small improvement of blacks at high-paying occupations). The situation improved between 2001 and 2007 (the latter curve is always above the former). For the entire period, 1996-2007, the curves cross twice (at the 83<sup>rd</sup> and 90<sup>th</sup> percentiles), showing a general improvement of the situation of black Africans at the bottom and top of the distribution, but with some deterioration in the middle.

From Figures 4.a-c, we can infer that almost all segregation of black Africans with respect to whites is into low-paying occupations because the concentration curve lies very close to the corresponding segregation curve every year. This strong racial stratification is confirmed by the corresponding Gini and *D* concentration ratios, close to 100%. Thus, the proportion of black Africans in an occupation is a very good predictor of how low it pays on average.<sup>21</sup> This correlation was intensified over time, from 90% (92.5%) in 1996 to 95% (96%) with Gini (*D*) because between 1996 and 2001 concentration indices increased more intensely than their segregation counterparts, and decreased at a similar rate between 2001 and 2007 (Figure 5.a). black Africans faced less segregation in 2007 as compared with 1996, but with a higher fraction of it being into low-paying occupations. The unnormalized concentration indices increased by 6% (Gini) and 10% (*D*) between 1996 and 2001, and decreased by 11% and 15% between 2001 and 2007. The reductions for the entire period were of only 6% and 7%.<sup>22</sup>

The results using labor force data (Figure 6 for 3 digits) show that unnormalized levels of segregation into low-paying occupations follow a similar trend as the one described for segregation. However, the degree of stratification is even higher in 2015 than right after the end of the apartheid, and about 23-29% higher than its minimum achieved in 1996, with oscillations along the business cycle in between.<sup>23</sup> These results also confirm that most segregation of black Africans is into low-pay, with a tendency to increase the fraction over time (from 81% in 1994

---

<sup>21</sup> This may be influenced by the fact that the presence of blacks in an occupation by itself pushes down the average income. The use of the African income structure (instead of that of the entire population) to measure how much an occupation pays maintains the qualitative results, even if the share of segregation that can be considered into low-paying occupations is smaller: 79%, 84%, and 84% in 1996, 2001 and 2007 (Gini), and 80%, 84%, 94% (*D*).

<sup>22</sup> Like in the case of segregation, the improvement of blacks was smaller if we remove workers with unknown occupation, resulting in no improvement between 1996 and 2007 with *D*, and a small 3% reduction with Gini.

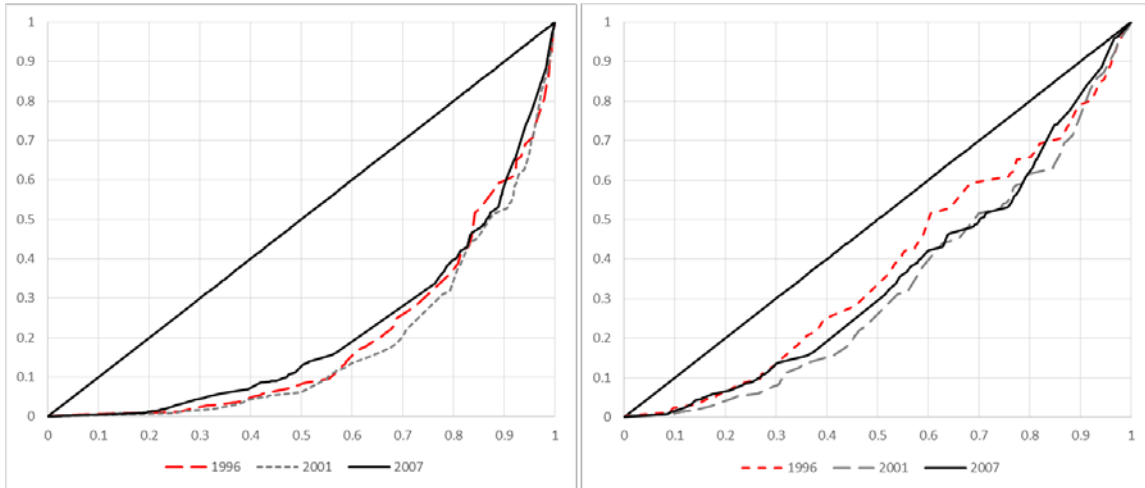
<sup>23</sup> The correlation between stratification and unemployment rates was 74-76% (3-digit Gini and *D*) between 1994-2015.

to 89-95% in 2015). A similar picture is obtained with the 2-digit classification (reported in Table A1 in the Appendix).

**Figure 3. White-African concentration curves**

**a. Observed**

**b. Unexplained**



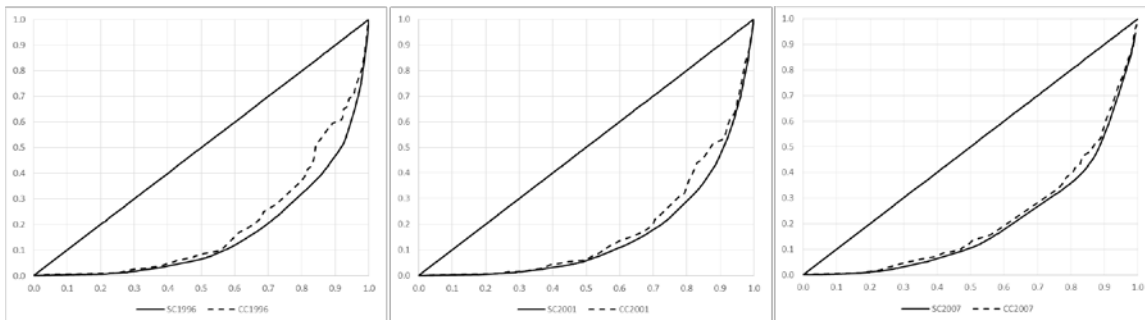
Source: Own construction based on IPUMS-International (3-digit classification).

**Figure 4. White-African segregation and concentration curves**

**a. 1996**

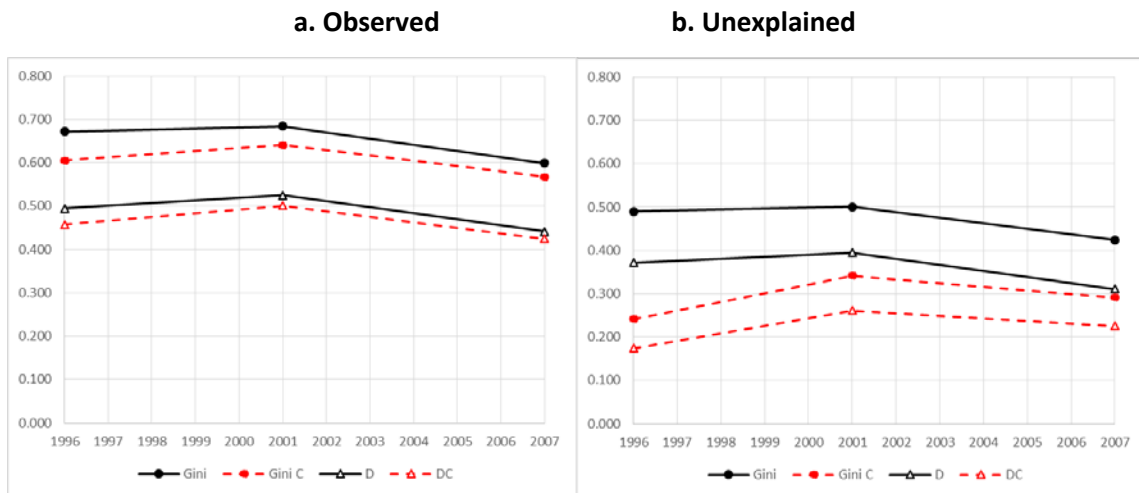
**b. 2001**

**c. 2007**



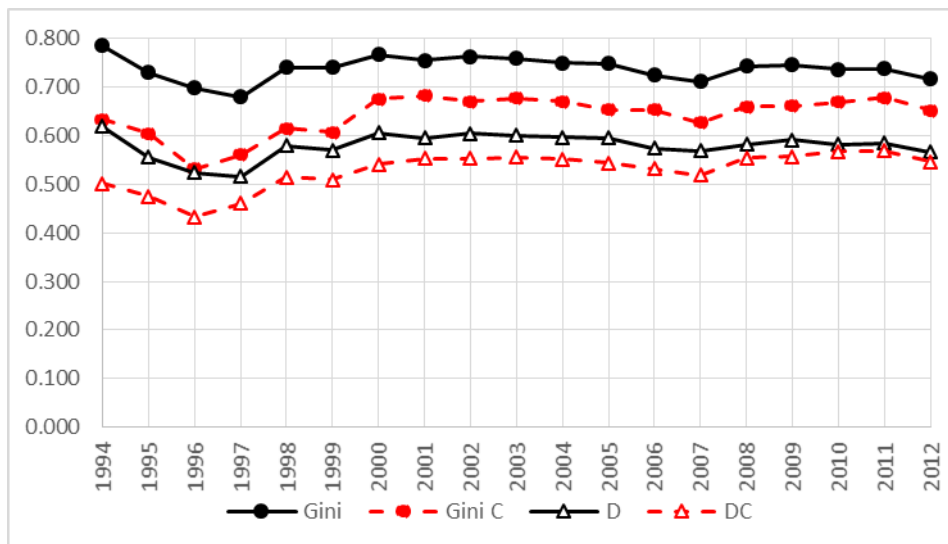
Source: Own construction based on IPUMS-International (3-digit classification).

**Figure 5. White-African segregation into low-paying occupations**



Note: Gini and Dissimilarity segregation and concentration (C) indices.  
 Source: Own construction based on IPUMS-International (3-digit classification).

**Figure 6. White-African segregation into low-paying occupations**



Note: Gini and Dissimilarity segregation and concentration (C) indices.  
 Source: Own construction based on PALMS (3-digit classification).

### 5.3 Conditional racial segregation and stratification

The segregation of African workers across occupations, and their over-representation into low-paying occupations, could be to some extent the result of inequalities of other kind (geographical, demographic, educational, etc.) that occurred previous to the entrance in the labor market, whether the result of previous or anticipated discrimination or not. As reported in Table 1, black Africans, compared with whites, are under-represented in urban areas, in richest provinces such as Gauteng (which includes Johannesburg and Pretoria) or Western Cape (including Cape Town). They are also under-represented among high skilled workers (with

secondary or higher education completed) and immigrants, and they tend to be younger and unmarried in larger proportions (African spouses are less likely to be employed). To see how much segregation is due to the particular distribution of characteristics across races, we compare observed segregation and segregation in the counterfactual situation in which black Africans are given the same distribution of characteristics of whites (Tables 2 and 3 using census data).

Only 29% of segregation in 2007 is directly associated with differences in observed characteristics between black Africans and whites (i.e. 0.175 with Gini, and 0.130 with  $D$ ).<sup>24</sup> More precisely, about 26% of segregation is explained by differences in attained education, and another 4% by the different geographical distribution of workers of each race. There is virtually no effect associated with differences across demographic variables or the immigration profiles. This means that a large 71% of segregation, i.e. Gini=0.424 and  $D$ =0.311, remains after equalizing the distribution of characteristics for black and white workers (73% with Gini, 75% with  $D$  in 1996 and 2001).

Racial inequality in the distribution of characteristics helps to better explain the segregation of black Africans into low-paying occupations, near one half in 2007 (49% Gini; 47%  $D$ ), with education playing the most fundamental role (44% Gini; 42%  $D$ ), the rest being associated with differences in demographic variables and location.<sup>25</sup>

These large unexplained terms in segregation and stratification are the result of differences in the conditional occupational distributions, with similar workers of each race working in different occupations. This may be the result of differences across unobservables, such as the lower quality of education received by blacks, of differences in preferences, or of direct discrimination by race in hiring or promotion practices. It is also interesting to note the differential roles that the explained and unexplained terms played in the trends described above.

There is some overlapping between the 1996 and 2001 unexplained segregation curves (Figure 1.b), while the level of the unexplained segregation indices increased, accounting for most (79-85%) of the overall increase in segregation during this period. The 2007 curve of unexplained segregation falls always above that of 2001, which implied a large decline in unexplained segregation indices, accounting for the decrease in segregation with  $D$ . It also accounted for near 90% of the decrease with Gini, although in this case there was also a substantial reduction

---

<sup>24</sup> Auxiliary logit regressions are reported in Tale A3 in the Appendix.

<sup>25</sup> The explained terms are smaller if we use the alternative counterfactual and give black Africans the conditional employment distribution of whites (by reweighting whites' distribution to reproduce the characteristics of blacks): 16-18% (segregation) and 24-26% (segregation into low-paying jobs).

associated with the relative improvement in education of blacks (that must have reduced segregation within white or black dominated occupations and for that reason did not affect  $D$ ). For the entire period, the observed reduction in segregation was mostly driven by the unexplained part (with non-overlapping curves).

In the case of unexplained stratification, concentration curves are less informative because they overlap in all periods. The indices, however, produce clear results. The increase in the level of segregation into low-paying occupations between 1996 and 2001 was also entirely driven by large increases in the unexplained term, only partially compensated by reductions in the explained one (mostly due to education upgrading among black Africans). Between 2001 and 2007, both components (explained and unexplained) were reduced, but the reduction in this type of segregation was mostly driven by the unexplained part (70%) in the case of Gini, and the explained part in the case of  $D$  (56%). As a result, we do find evidence that the improvement in the level of education of black Africans helped to push down racial stratification of occupations in both periods (a total reduction of about 34-40% was explained by this characteristic). While during the first period this was partially offset by a higher unexplained term, this was reversed in the second one.

**Table 1. Workers' characteristics by race**

	Black Africans			Whites		
	1996	2001	2007	1996	2001	2007
<b>Rural</b>	35.5	32.9	32.7	8.6	8.0	7.8
<b>Urban</b>	64.5	67.1	67.3	91.4	92.0	92.2
<b>Western Cape</b>	4.6	5.3	6.4	18.2	18.7	20.6
<b>Eastern Cape</b>	10.0	8.7	11.3	7.2	6.7	6.3
<b>Northern Cape</b>	1.2	1.1	1.1	2.4	2.2	1.8
<b>Free State</b>	8.8	8.0	6.4	6.6	4.9	5.6
<b>KwaZulu-Natal</b>	18.0	17.7	17.7	13.0	11.0	9.8
<b>North West</b>	10.8	10.3	9.3	4.5	5.0	4.2
<b>Gauteng</b>	28.5	31.1	30.5	40.4	44.2	43.5
<b>Mpumalanga</b>	9.0	8.6	8.8	5.5	4.4	5.9
<b>Limpopo</b>	9.1	9.3	8.5	2.5	2.9	2.3
<b>No schooling</b>	16.3	14.0	7.2	0.6	0.6	0.2
<b>Some primary</b>	12.5	12.4	11.8	0.2	0.3	0.2
<b>Primary (6 years)</b>	25.6	20.7	18.4	1.5	2.4	2.3
<b>Lower secondary</b>	21.0	21.1	23.1	15.5	15.0	11.0
<b>Secondary</b>	19.4	28.7	33.8	58.5	64.3	62.1
<b>University</b>	1.5	3.1	4.6	12.6	17.4	23.6
<b>Other education</b>	3.1	0.0	1.2	7.1	0.0	0.6
<b>Unknown education</b>	0.6	0.0	0.0	4.0	0.0	0.0
<b>15-24 years old</b>	10.1	10.0	12.8	14.3	11.8	10.9
<b>25-34 years old</b>	35.8	33.8	31.3	29.7	28.3	23.3
<b>35-44 years old</b>	31.0	31.7	28.8	27.4	28.4	27.4
<b>45-54 years old</b>	16.4	18.3	19.6	20.1	21.4	23.8
<b>55-65 years old</b>	6.6	6.2	7.5	8.6	10.2	14.6
<b>Female</b>	41.0	42.0	43.6	42.8	44.2	45.5
<b>Single/never married/unknown</b>	37.4	36.3	42.1	20.4	19.3	21.1
<b>Married/in union</b>	56.0	56.9	51.1	70.4	72.0	71.4
<b>Separated/divorced/spouse absent</b>	3.7	3.6	3.0	7.4	7.0	5.7
<b>Widowed</b>	2.9	3.2	3.8	1.8	1.7	1.7
<b>Head</b>	57.8	58.8	57.2	55.1	52.1	51.1
<b>Spouse</b>	16.2	16.4	15.0	28.4	28.9	29.8
<b>Disabled</b>	7.3	3.3	2.0	2.5	1.9	1.5
<b>Native</b>	88.8	93.6	93.6	87.2	89.8	90.8
<b>National immigrant</b>	9.5	5.6	5.4	11.4	9.1	7.7
<b>Immigrant from abroad</b>	1.7	0.7	1.0	1.3	1.0	1.5

Source: Own construction based on IPUMS-International.

**Table 2. Segregation/concentration (Gini) with standard errors (se)**

1996	Segregation	se	% observed	Concentration	se	% observed	Ratio	% observed
<b>Observed</b>	0.672	0.001	100	0.606	0.001		0.901	100
<b>Unexplained</b>	0.490	0.003	72.9	0.242	0.003	39.9	0.493	54.7
<b>Explained (total)</b>	0.182	0.002	27.1	0.364	0.003	60.1	0.408	45.3
<b>Location</b>	0.015	0.001	2.2	-0.014	0.001	-2.3	-0.040	-4.4
<b>Education</b>	0.169	0.002	25.2	0.373	0.002	61.6	0.432	47.9
<b>Demographics</b>	-0.001	0.001	-0.2	0.005	0.001	0.9	0.017	1.8
<b>Immigration</b>	0.000	0.000	0.0	0.000	0.000	-0.1	-0.001	-0.1
<b>2001</b>								
<b>Observed</b>	0.685	0.001	100	0.641	0.001		0.936	100
<b>Unexplained</b>	0.501	0.002	73.1	0.342	0.003	53.3	0.683	72.9
<b>Explained (total)</b>	0.184	0.002	26.9	0.299	0.002	46.7	0.253	27.1
<b>Location</b>	0.019	0.001	2.8	-0.006	0.001	-1.0	-0.038	-4.1
<b>Education</b>	0.168	0.001	24.5	0.293	0.001	45.7	0.260	27.8
<b>Demographics</b>	-0.003	0.001	-0.5	0.013	0.001	2.0	0.032	3.4
<b>Immigration</b>	0.001	0.000	0.1	0.000	0.000	0.0	-0.001	-0.1
<b>2007</b>								
<b>Observed</b>	0.599	0.003	100	0.567	0.003		0.948	100
<b>Unexplained</b>	0.424	0.005	70.8	0.292	0.006	51.4	0.688	72.6
<b>Explained (total)</b>	0.175	0.004	29.2	0.276	0.005	48.6	0.260	27.4
<b>Location</b>	0.021	0.002	3.6	0.005	0.002	0.9	-0.026	-2.7
<b>Education</b>	0.155	0.003	25.9	0.247	0.003	43.5	0.228	24
<b>Demographics</b>	-0.003	0.002	-0.6	0.023	0.002	4.0	0.059	6.2
<b>Immigration</b>	0.002	0.000	0.3	0.001	0.000	0.1	-0.001	-0.1

Note: Bootstrap standard errors (200 replications).

Source: Own construction based on IPUMS-International (3-digit classification).

**Table 3. Segregation/concentration (D) with standard errors (se)**

1996	Segregation	se	% observed	Concentration	se	% observed	Ratio	% observed
<b>Observed</b>	0.495	0.001	100	0.458	0.001	100	0.925	100
<b>Unexplained</b>	0.371	0.002	75.0	0.174	0.002	38.0	0.468	50.7
<b>Explained (total)</b>	0.124	0.002	25.0	0.284	0.002	62.0	0.456	49.3
<b>Location</b>	0.009	0.001	1.9	-0.010	0.001	-2.1	-0.033	-3.5
<b>Education</b>	0.117	0.002	23.6	0.293	0.001	64.0	0.476	51.5
<b>Demographics</b>	-0.002	0.001	-0.4	0.001	0.001	0.2	0.014	1.5
<b>Immigration</b>	0.000	0.000	0.0	0.000	0.000	-0.1	-0.001	-0.1
<b>2001</b>								
<b>Observed</b>	0.525	0.001	100	0.502	0.001	100	0.956	103
<b>Unexplained</b>	0.395	0.002	75.2	0.261	0.002	52.0	0.661	71.5
<b>Explained (total)</b>	0.130	0.002	24.8	0.241	0.002	48.0	0.295	31.9
<b>Location</b>	0.015	0.001	3.0	-0.008	0.001	-1.7	-0.047	-5.1
<b>Education</b>	0.117	0.001	22.3	0.240	0.001	47.9	0.312	33.7
<b>Demographics</b>	-0.003	0.001	-0.6	0.009	0.001	1.9	0.033	3.5
<b>Immigration</b>	0.001	0.000	0.1	-0.001	0.000	-0.1	-0.002	-0.2
<b>2007</b>								
<b>Observed</b>	0.442	0.003	100	0.424	0.003	100	0.961	104
<b>Unexplained</b>	0.311	0.004	70.5	0.225	0.005	53.1	0.724	78.3
<b>Explained (total)</b>	0.130	0.004	29.5	0.199	0.005	46.9	0.237	25.6
<b>Location</b>	0.018	0.001	4.0	0.011	0.001	2.6	-0.008	-0.9
<b>Education</b>	0.115	0.003	26.0	0.176	0.003	41.6	0.204	22.1
<b>Demographics</b>	-0.003	0.002	-0.8	0.011	0.002	2.6	0.042	4.5
<b>Immigration</b>	0.001	0.000	0.3	0.001	0.000	0.2	0.000	0.0

Note: Bootstrap standard errors (200 replications).

Source: Own construction based on IPUMS-International (3-digit classification).



## 6. Concluding remarks

Discriminatory legislation and social practices in pre-democratic South Africa lead to a labor market strongly stratified by race, with whites holding the most skilled and best-paying jobs. Lessons from other societies, such as the US or Latin American countries, indicate that racial discrimination and segregation are more difficult to eradicate than removing all discriminatory legislation. Discrimination may persist before the entrance in the labor market in the form of lower amount and quality of education for nonwhites or the negative influence of ghettos or family background, compromising their economic opportunities. This may later be aggravated by direct or subtler discriminatory practices in hiring or promotion, whether based on prejudices or on information problems (statistical discrimination).

With all the necessary cautions that data limitations impose, we have not found strong evidence supporting that the distribution of occupations has been effectively either desegregated or de-stratified in post-apartheid South Africa. There is limited evidence of such a process only over short periods, during the first years of democracy in mid'1990s or during early 2000s, but not the required sustained path over time. It seems, in fact, that the current situation is not better than it was in mid'1990s. Not only has segregation remained high, but the nature of that segregation still implies a strong racial stratification with an over-representation of black Africans in the lowest-paying jobs.

Inequality in the distribution of worker's characteristics, especially attained education, explains less than a third of segregation and about a half of stratification. The improvement in the characteristics of black Africans over time had a positive impact on segregation and stratification trends, but is not being effective enough to revert the inherited situation. A large unexplained term remains. Black Africans tend to work in different and lower-paying occupations compared to whites with similar observed characteristics. Changes over time on this unexplained part generally drove the segregation and stratification trends, which are strongly connected with the business cycle.

## References

- Allanson, P., J. Atkins, and T. Hinks (2000), "A Multilateral Decomposition of Racial Wage Differentials in the 1994 South African Labour Market", *The Journal of Development Studies*, 37(1): 93-120.
- Allanson, P., J. Atkins, and T. Hinks (2002), "No End to the Racial Wage Hierarchy in South Africa?", *Review of Development Economics*, 6(3): 442-59.
- Banerjee, A., S. Galiani, J. Levinsohn, Z. McLaren, and I. Woolard (2008), "Why has Unemployment Risen in the New South Africa?", *Economics of Transition*, 16(4): 715-740.

- Beinart, W. and S. Dubow (Eds.) (1995), *Segregation and Apartheid in Twentieth-Century South Africa*, Routledge, London, UK.
- Blackburn R.M. and J. Jarman (1997), "Occupational Gender Segregation", *Social Research Update*, 16.
- Brookes, M. and T. Hinks (2004), "The Racial Employment Gap in South Africa", *South African Journal of Economics*, 72(3): 573-580.
- Campbell, P.R. (1987), "South Africa 1960 to 1986, a Statistical View of Racial Differences", US Bureau of the Census, Center for International Research.
- Chantreuil, F. and A. Trannoy (2013), "Inequality Decomposition Values: the Trade-Off between Marginality and Consistency," *Journal of Economic Inequality*, 11(1): 83-98.
- DiNardo, J., N.M. Fortin, T. Lemieux (1996), "Labor Market Institutions and the Distribution of Wages, 1973–1992: A Semiparametric Approach", *Econometrica*, 64: 1001-1044.
- Duncan, O.D., and B. Duncan (1955), "A Methodological Analysis of Segregation Indexes", *American Sociological Review*, 20(2): 210-217.
- Feinstein, C.H., (2005), *An Economic History of South Africa: Conquest, Discrimination and Development*, Cambridge University Press, Cambridge, UK.
- Gradín, C. (2013a), "Race, Poverty, and Deprivation in South Africa", *Journal of African Economies*, 22(2): 187-238.
- Gradín, C. (2013b), "Conditional Occupational Segregation of Minorities in the U.S.", *Journal of Economic Inequality*, 11(4): 473-493.
- Gradín (2014), "Race and Income Distribution: Evidence from the USA, Brazil and South Africa", *Review of Development Economics*, 18(1): 73-92.
- Gradín, C. (2016), "Segregation of Women into Low-Paying Occupations in the US", Mimeo, Universidade de Vigo.
- Hutchens, R.M (2004), "One Measure of Segregation", *International Economic Review*, 45(2): 555-578.
- Jahn, J., C.F. Schmid, and C. Schrag (1947), "The Measurement of Ecological Segregation", *American Sociological Review*, 12(3): 293-303.
- Kerr, A., D. Lam and M. Wittenberg (2016), *Post-Apartheid Labour Market Series* [dataset], Version 3.1, Cape Town: DataFirst [producer and distributor].
- Kerr, A. and M. Wittenberg (2016), "A Guide to Version 3.1 of the Post-Apartheid Labour Market Series (PALMS)", DataFirst Portal, University of Cape Town.
- Keswell, M. (2010), "Education and Racial Inequality in Post Apartheid South Africa", in Attewell, P. and K.S. Newman, *Growing Gaps: Educational Inequality around the World*, Oxford University Press, Oxford, UK.
- Keswell, M., S. Girdwood, and M. Leibbrandt (2013), "Educational Inheritance and the Distribution of Occupations: Evidence from South Africa", *Review of Income and Wealth*, 59: S111-S137.
- Kingdon, G.G. and J. Knight (2004), "Race and the Incidence of Unemployment in South Africa", *Review of Development Economics*, 8(2): 198–222.
- Kingdon, G.G. and J. Knight (2007), "Unemployment in South Africa, 1995–2003: Causes, Problems and Policies", *Journal of African Economies*, 16(5): 813-84.
- Leibbrandt, M., I. Woolard, H. McEwen, and C. Koep (2010), "Employment and Inequality Outcomes in South Africa", Southern Africa Labour and Development Research Unit and School of Economics, University of Cape Town.
- Mariotti, M. (2012), "Labour Markets During Apartheid in South Africa", *Economic History Review*, 65(3): 1100–1122.
- Minnesota Population Center (2015), *Integrated Public Use Microdata Series, International: Version 6.4* [Machine-readable database], Minneapolis: University of Minnesota.
- Parashar, S. (2008), "Marginalized by Race and Place: Occupational Sex Segregation in Post-Apartheid South Africa", Thesis, University of Maryland.

- Rodrik, D. (2008), "Understanding South Africa's Economic Puzzles", *Economics of Transition*, 16(4): 769–797.
- Rospabé, S. (2002), "How Did Labour Market Racial Discrimination Evolve After The End of Apartheid?", *South African Journal of Economics*, 70(1): 185–217.
- Schechtman, E. and S. Yitzhaki (1987), "A Measure of Association Based on Gini's Mean Difference", *Communications in Statistics*, 16(1): 207–231.
- Schechtman, E. and S. Yitzhaki (1999), "On the Proper Bounds of the Gini correlation", *Economics Letters*, 63: 133–138.
- Shorrocks, A.F. (2013), "Decomposition Procedures for Distributional Analysis: a Unified Framework Based on the Shapley value", *Journal of Economic Inequality*, 11(1): 99-126.
- Somers, R.H. (1962), "A New Asymmetric Measure of Association for Ordinal Variables", *American Sociological Review*, 27(6): 799-811.
- StatsSA (2007), "Using the 2001 Census: Approaches to Analysing Data", Statistics South Africa, Pretoria.
- Treiman, D.J. (2007), "The Legacy of Apartheid: Racial Inequalities in the New South Africa", in A.F Heath and S.-Y. Cheung, *Unequal Chances: Ethnic Minorities in Western Labour Markets*, 401-447, Oxford University Press, Oxford, UK.
- Treiman, D.J., M. McKeever, and E. Fodor (1996), "Racial Differences in Occupational Status and Income in South Africa, 1980 and 1991", *Demography*, 33(1): 111-132
- van der Berg, S. (2007), "Apartheid's Enduring Legacy: Inequalities in Education", *Journal of African Economies*, 16(5): 849–880.
- Wittenberg, M. (2007), "Dissecting Post-Apartheid Labour Market Developments: Decomposing a Discrete Choice Model while Dealing with Unobservables", Economic Research Southern Africa (ERSA) Working Paper, 46, University of Cape Town.
- Wittenberg, M. (2014), "Analysis of Employment, Real Wage, and Productivity Trends in South Africa since 1994", Conditions of Work and employment Series, 45, International Labour Office, Geneva, Switzerland.
- Yitzhaki, S. and I. Olkin (1991), "Concentration Indices and Concentration Curves", in K. Mosler and M. Scarsini (eds.), *Stochastic Orders and Decisions under Risk*, Institute of Mathematical Statistics, Lecture-Notes Monograph Series, 1991: 380-392.
- Yu, D. (2009), "The Comparability of Census 1996, Census 2001 and Community Survey 2007", Stellenbosch Economic Working Papers, 21/09, Department of Economics and Bureau For Economic Research, University of Stellenbosch.

**Table A1. Occupational segregation by race**

Digits	Segregation indices						Concentration indices and ratios											
	Gini			Dissimilarity			Gini						Dissimilarity					
	1	2	3	1	2	3	1	ratio	2	ratio	3	ratio	1	ratio	2	ratio	3	ratio
<b>1994</b>	0.634	0.718	0.785	0.517	0.556	0.620	0.590	93.2%	0.614	85.5%	0.633	80.7%	0.499	96.5%	0.504	90.6%	0.501	80.9%
<b>1995</b>	0.603	0.687	0.730	0.470	0.537	0.556	0.578	95.8%	0.588	85.5%	0.604	82.8%	0.466	99.3%	0.447	83.3%	0.475	85.5%
<b>1996</b>	0.567	0.643	0.699	0.432	0.483	0.524	0.552	97.4%	0.547	85.0%	0.532	76.2%	0.428	99.0%	0.420	86.8%	0.432	82.6%
<b>1997</b>	0.560	0.627	0.679	0.454	0.475	0.516	0.525	93.7%	0.553	88.3%	0.561	82.6%	0.454	100.0%	0.446	94.0%	0.461	89.3%
<b>1998</b>	0.631	0.694	0.740	0.510	0.553	0.579	0.576	91.3%	0.623	89.8%	0.615	83.1%	0.510	100.0%	0.541	97.8%	0.515	88.9%
<b>1999</b>	0.643	0.704	0.741	0.503	0.558	0.570	0.604	94.0%	0.590	83.7%	0.607	81.9%	0.503	100.0%	0.498	89.2%	0.510	89.4%
<b>2000</b>	0.683	0.725	0.767	0.527	0.560	0.606	0.620	90.8%	0.655	90.3%	0.676	88.1%	0.527	100.0%	0.553	98.8%	0.541	89.2%
<b>2001</b>	0.665	0.707	0.756	0.524	0.548	0.595	0.617	92.9%	0.641	90.7%	0.683	90.4%	0.524	100.0%	0.545	99.4%	0.553	93.0%
<b>2002</b>	0.670	0.724	0.763	0.529	0.568	0.605	0.604	90.2%	0.642	88.7%	0.670	87.7%	0.529	100.0%	0.555	97.8%	0.553	91.4%
<b>2003</b>	0.673	0.725	0.759	0.530	0.574	0.601	0.628	93.3%	0.653	90.0%	0.677	89.2%	0.530	100.0%	0.547	95.3%	0.556	92.5%
<b>2004</b>	0.664	0.717	0.750	0.536	0.570	0.596	0.627	94.4%	0.645	89.9%	0.671	89.5%	0.536	100.0%	0.547	96.0%	0.551	92.4%
<b>2005</b>	0.655	0.707	0.748	0.531	0.564	0.596	0.611	93.3%	0.632	89.5%	0.653	87.3%	0.531	100.0%	0.528	93.6%	0.544	91.3%
<b>2006</b>	0.630	0.688	0.725	0.516	0.547	0.574	0.603	95.7%	0.639	92.8%	0.654	90.2%	0.516	100.0%	0.523	95.6%	0.532	92.6%
<b>2007</b>	0.615	0.666	0.712	0.499	0.530	0.569	0.589	95.7%	0.607	91.1%	0.627	88.1%	0.499	100.0%	0.506	95.5%	0.518	91.1%
<b>2008</b>	0.665	0.700	0.743	0.547	0.554	0.582	0.636	95.7%	0.646	92.3%	0.660	88.9%	0.538	98.4%	0.530	95.7%	0.555	95.3%
<b>2009</b>	0.669	0.705	0.746	0.555	0.565	0.592	0.632	94.5%	0.645	91.6%	0.661	88.6%	0.542	97.6%	0.536	94.8%	0.558	94.3%
<b>2010</b>	0.665	0.699	0.737	0.544	0.559	0.582	0.635	95.5%	0.656	93.7%	0.670	90.9%	0.544	100.0%	0.553	98.9%	0.567	97.4%
<b>2011</b>	0.669	0.706	0.738	0.543	0.564	0.585	0.644	96.3%	0.664	94.1%	0.678	92.0%	0.535	98.6%	0.550	97.4%	0.569	97.3%
<b>2012</b>	0.646	0.681	0.717	0.520	0.544	0.566	0.628	97.4%	0.625	91.8%	0.652	90.9%	0.513	98.8%	0.537	98.8%	0.547	96.6%
<b>2013</b>	0.644	0.676	0.717	0.519	0.542	0.569	0.615	95.5%	0.625	92.5%	0.641	89.5%	0.512	98.8%	0.529	97.6%	0.534	93.8%
<b>2014</b>	0.641	0.675	0.714	0.511	0.538	0.558	0.609	95.0%	0.628	93.0%	0.636	89.0%	0.504	98.6%	0.526	97.7%	0.532	95.3%
<b>2015</b>	0.653	0.694	0.734	0.520	0.555	0.591	0.618	94.7%	0.642	92.5%	0.655	89.2%	0.515	98.9%	0.540	97.2%	0.559	94.6%

Source: Own construction based on PALMS.

**Table A2. Occupational segregation by race (standard errors)**

Digits	Segregation						Concentration					
	Gini			Dissimilarity			Gini			Dissimilarity		
	1	2	3	1	2	3	1	2	3	1	2	3
<b>1994</b>	0.007	0.006	0.005	0.007	0.007	0.006	0.007	0.007	0.007	0.007	0.007	0.007
<b>1995</b>	0.007	0.007	0.006	0.007	0.008	0.007	0.007	0.007	0.007	0.007	0.007	0.007
<b>1996</b>	0.011	0.010	0.009	0.011	0.010	0.010	0.011	0.012	0.013	0.011	0.012	0.012
<b>1997</b>	0.008	0.008	0.007	0.009	0.008	0.008	0.009	0.009	0.009	0.009	0.009	0.008
<b>1998</b>	0.010	0.009	0.008	0.011	0.010	0.009	0.011	0.010	0.011	0.011	0.010	0.010
<b>1999</b>	0.009	0.008	0.008	0.010	0.009	0.009	0.009	0.009	0.009	0.010	0.010	0.010
<b>2000</b>	0.007	0.007	0.006	0.008	0.007	0.007	0.008	0.007	0.007	0.009	0.009	0.009
<b>2001</b>	0.006	0.006	0.005	0.007	0.007	0.006	0.006	0.006	0.006	0.007	0.006	0.006
<b>2002</b>	0.005	0.005	0.005	0.006	0.006	0.005	0.006	0.006	0.005	0.006	0.006	0.006
<b>2003</b>	0.006	0.005	0.005	0.006	0.006	0.005	0.006	0.006	0.005	0.006	0.006	0.006
<b>2004</b>	0.006	0.006	0.006	0.007	0.007	0.006	0.007	0.007	0.007	0.007	0.007	0.007
<b>2005</b>	0.008	0.007	0.006	0.008	0.008	0.007	0.008	0.008	0.008	0.008	0.008	0.008
<b>2006</b>	0.008	0.008	0.007	0.009	0.008	0.008	0.009	0.008	0.008	0.009	0.009	0.008
<b>2007</b>	0.010	0.010	0.008	0.013	0.010	0.009	0.012	0.012	0.011	0.013	0.012	0.012
<b>2008</b>	0.004	0.004	0.004	0.005	0.004	0.004	0.005	0.004	0.004	0.005	0.005	0.005
<b>2009</b>	0.005	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
<b>2010</b>	0.005	0.004	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
<b>2011</b>	0.004	0.004	0.004	0.005	0.005	0.005	0.004	0.004	0.004	0.005	0.005	0.005
<b>2012</b>	0.005	0.005	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
<b>2013</b>	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.006	0.005
<b>2014</b>	0.005	0.005	0.005	0.006	0.005	0.006	0.006	0.005	0.005	0.006	0.005	0.005
<b>2015</b>	0.005	0.005	0.004	0.006	0.005	0.005	0.005	0.005	0.005	0.006	0.006	0.005

Source: Own construction based on PALMS.

**Table A3. Logit regressions (probability of being white)**

	1996		2001		2007	
<b>Number of obs.</b>	630,350		590,227		194,235	
<b>Wald chi2(28)</b>	118,613		97,458		25,124	
<b>Prob. &gt; chi2</b>	0		0		0	
<b>Pseudo R2</b>	0.423		0.341		0.321	
<b>-Dependent variable</b>						
<b>White</b>						
<b>-Explanatory variables:</b>	<b>Coef.</b>	<b>Std. Err.</b>	<b>Coef.</b>	<b>Std. Err.</b>	<b>Coef.</b>	<b>Std. Err.</b>
<b>Urban</b>	0.918	0.012	0.915	0.014	0.937	0.025
<b>Eastern Cape</b>	-1.703	0.018	-1.458	0.017	-1.451	0.036
<b>Northern Cape</b>	-0.158	0.033	-0.080	0.031	-0.395	0.050
<b>Free State</b>	-1.399	0.018	-1.446	0.019	-1.143	0.036
<b>KwaZulu-Natal</b>	-1.619	0.016	-1.560	0.015	-1.460	0.030
<b>North West</b>	-1.983	0.020	-1.648	0.020	-1.588	0.041
<b>Gauteng</b>	-1.299	0.013	-1.111	0.012	-0.969	0.024
<b>Mpumalanga</b>	-1.497	0.020	-1.490	0.019	-1.202	0.036
<b>Limpopo</b>	-2.533	0.025	-2.026	0.025	-2.102	0.049
<b>Some primary completed</b>	-1.188	0.070	-0.637	0.060	-0.334	0.181
<b>Primary (6 years) completed</b>	0.197	0.039	0.737	0.041	1.546	0.141
<b>Lower secondary general completed</b>	2.910	0.034	2.745	0.038	3.145	0.137
<b>Secondary, general track completed</b>	4.450	0.033	4.012	0.037	4.595	0.137
<b>University completed</b>	5.388	0.037	4.841	0.039	5.435	0.138
<b>Unknown/missing</b>	4.153	0.036	-	-	3.010	0.161
<b>Other education</b>	5.180	0.042	-	-	-	-
<b>Aged 25-34</b>	-1.313	0.014	-0.969	0.014	-0.729	0.028
<b>Aged 35-44</b>	-1.143	0.015	-0.779	0.015	-0.404	0.030
<b>Aged 45-54</b>	-0.386	0.017	-0.071	0.017	0.264	0.033
<b>Aged 55-65</b>	0.031	0.021	0.687	0.021	1.211	0.039
<b>Female</b>	-0.312	0.010	-0.177	0.009	-0.242	0.020
<b>Married/in union</b>	1.013	0.011	1.124	0.012	1.037	0.025
<b>Separated/divorced/spouse absent</b>	1.780	0.020	1.713	0.020	1.461	0.043
<b>Widowed</b>	0.802	0.031	0.546	0.030	0.368	0.059
<b>Household Head</b>	0.138	0.012	-0.285	0.012	-0.440	0.025
<b>Spouse</b>	0.729	0.015	0.203	0.015	0.164	0.033
<b>Disable</b>	1.019	0.021	0.172	0.028	-0.250	0.068
<b>Immigrant within South Africa</b>	0.085	0.013	0.181	0.015	0.278	0.031
<b>Immigrant from abroad</b>	-0.052	0.034	0.210	0.044	0.353	0.077
<b>Intercept</b>	-4.693	0.043	-4.136	0.049	-5.057	0.158

Note: Omitted categories: Rural, Western Cape, No schooling, Aged 16-24, Non-immigrant, Male, No disability, other than spouse or household head.

Source: Own construction based on IPIMUS-International.