Abstract
This paper investigates progress in reducing the high level of racial stratification of occupations after apartheid in South Africa. Empirical analysis, using census microdata and Labour Force Surveys, does not provide compelling evidence of sustained or significant desegregation. Occupations remain highly segmented by race, with blacks disproportionately holding low-paying jobs (compared with whites), although segregation and segmentation also affect in a different way the other population groups (Indians/Asians and Coloureds). Less than a third of the occupational segregation and about half of the segmentation of Africans (with respect to whites) are related to their characteristics, especially their lower educational achievement, a gap that has been reduced over time. Segregation and stratification, however, remain when blacks and whites with similar characteristics are compared.

KEYWORDS
occupational segregation, stratification, low-paying, apartheid, South Africa, race

INTRODUCTION
The lives of South Africans have been dominated by racial segregation since the first Europeans arrived at the Cape in 1652, beginning the largest European settlement on the continent. The segregation of blacks, along with that of Coloured and Indians/Asians, was intensified during apartheid,
the political and social regime enforced by the National Party after it took office in 1948 until the first general democratic elections in 1994.

The ultimate aim of white rulers was to force nonwhites to provide seasonal, cheap, and abundant labor for farms, mines, and other sectors, while keeping economic and political power in their own hands. Segregation in South Africa stood out for the range and extent of its discriminatory legislation, which affected every possible sphere of life (e.g., work, education, health, transport, recreation, politics, sexual relationships). Among this legislation, the ‘color bar’ resulted in job reservation for whites that excluded blacks from skilled and semi-skilled jobs, also depriving them of 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 & Dubow, 1995). Core elements of this segregation, such as the exclusion of blacks from skilled work (especially if it involved 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 deracialized South Africa started after the end of the apartheid regime under the rule of the African National Congress. This involved the formal dismantlement of all the remaining segregative legislation, along with the introduction of anti-discriminatory and affirmative policies to reverse its effects (i.e., 1995 Labour Relations Act, 1998 Employment Equity Act, 2000 Promotion of Equality and Prevention of Unfair Discrimination Act, 2003 Black Economic Empowerment Act). Deeply rooted inequalities along racial lines, however, proved more difficult to remove, especially in the context of a sluggish economy, the result of the shrinkage of the nonmineral tradable sector from the early 1990s on (Rodrik, 2008), with a chronically high level of unemployment.

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

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, & Schrag, 1947; Duncan & Duncan, 1955). We also analyze the vertical or ordinal dimension of segregation, measuring the extent to which the labor market is stratified by race, with blacks being systematically segregated into low-paying occupations, using concentration curves and indices when occupations are sorted by average earnings (Gradín, 2013a, 2017). Additionally, we attempt to identify the driving factors of this segregation at each moment in time by measuring the level conditional on workers’ characteristics using a counterfactual distribution in which blacks are given the characteristics of whites (Gradín, 2013a). More precisely, we analyze if segregation is driven by workers’ endowments, such as the lower level of education of blacks or their over-representation in rural areas and the poorest provinces of the country. Alternatively, segregation might result from the labor market being intrinsically segregative among workers with similar characteristics based on their race. Both sources of segregation might be the result of discrimination (actual or anticipated), but their distinction helps us 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 main empirical results, the sixth section deals with the multigroup case and the last section concludes.
As a legacy of colonialism and apartheid, the labor market in South Africa is largely stratified by race. Some of these racial inequalities have been extensively addressed in the previous literature, trying to measure how much progress, if any, was accomplished after the first democratic elections. Most of the research has focused so far on the magnitude of racial gaps in labor market outcomes and the extent to which they can 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., Kingdon & Knight, 2007; Banerjee, Galiani, Levinsohn, McLaren, & Woolard, 2008). This occurred in a context characterized by labor market inflexibility and a small informal sector compared with other developing countries (e.g., Kingdon & 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 characteristics of workers from 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 to 1999 and by Brookes and Hinks (2004) for 1995 to 2002. Paradoxically, changes in the characteristics of black South African men after apartheid have made them more employable over time, but at the same time their propensity to be employed has declined (Wittenberg, 2007).

Once workers enter 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. Some sort of desegregation started before the end of apartheid (Mariotti, 2012) with the increasing access of blacks to semi-skilled occupations that had previously been reserved for whites between 1970 and 1980, the year of the formal abolition of job reservation, although this was prompted by 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, McKeever, and Fodor (1996) estimated that the gap in occupational status in 1980 and 1991 (measured by the scale from the International Socioeconomic Index of Occupations) could be largely explained by the distinctive 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 had increased to 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 for continuing segregation, Keswell, Girdwood, and Leibbrandt (2013) highlighted the importance of education, showing that black female children who inherited the same level of educational opportunity as their parents were 6% to 10% more likely to be at the bottom of the occupational distribution than if they were exposed to better educational opportunities. Regarding the consequences, Gradín (2013b) has shown that the higher presence of blacks in more skilled occupations (along with their improved education) has contributed to reducing the racial poverty gap after apartheid (which was to a substantial extent explained...
by the cumulative disadvantaged characteristics of blacks, 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 based on a detailed classification of occupations, the main purpose of this paper. This is important because any difference between population groups within broadly defined skill categories remains hidden in other analyses. 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 to professional and technical occupations. More recently, in the context of an analysis of sex segregation, Parashar (2008) reported a Dissimilarity index of white–black segregation (two-digit classification of occupations) of 0.572 in 2001 using census data. She also reported that this segregation was greater 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 United States, stands out for displaying higher segregation by race than by gender. There is, thus, no clear picture of the long-term trend in segregation and the change in its nature after the end of apartheid.

More attention has focused on the huge earnings gap found between black and white South Africans. For example, Allanson, Atkins, and Hinks (2000) reported that one third of that gap in 1994 remained unexplained after controlling for differences in productivity, and Allanson, Atkins, and Hinks (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, which accounted for virtually none of the differential in 1993 but about 40% of it by 2002. This finding points to the increasing importance of differences in the quality of the education received by each population group, while the gap in the number 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. Burger, Jafta, and Von Fintel (2016) have claimed that 2003 (the year the Black Economic Empowerment Act was approved) was a turning point, when black–white discrimination started to decline continuously thereafter.

3 | METHODOLOGY

The conventional framework for measuring segregation of two groups across occupations uses the segregation curve and indices such as Dissimilarity or Gini (Jahn et al., 1947; Duncan & Duncan, 1955). Gradín (2017) expanded this framework to consider 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 (2013a) 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 both groups have the same characteristics, in line with previous research on the employment, earnings, and occupational attainment gaps.

3.1 | Measuring segregation

We compare the employment distribution across $J$ occupations of workers from a comparison group (blacks) and a reference group (whites), respectively $i = c,r$, described by the vector $f^i = (f^i_1, ..., f^i_J)$, where $f^i_j$ is the proportion of workers from group $i$ in occupation $j$, when
occupations are sorted by the ascending values of the relative share of members of whites (i.e., $f_j^w/f_j^t$). $F_j^t = \sum_{s=1}^{j} f_j^t$ indicates the corresponding cumulative distribution value.\(^1\)

The segregation curve $F(p, p \in [0, 1])$ plots the cumulative proportions of whites $F^w$ (vertical axis) and blacks $F^c$ (horizontal axis) for the $j$th occupation with largest under-representation of whites, 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). 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:

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

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 over-represented below and under-represented above: $D(f^c, f^t) = F_q^c - F_q^t$, where $q = \max_{j \in [1, J]} \{ j \mid f_j^c \geq f_j^t \}$. $D$ can be interpreted as the proportion of workers of any group that should change occupations (from those in which their group is over-represented to those in which it is under-represented) 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:

$$Gini(f^c, f^t) = 2 \sum_{j=1}^{J} \left( \tilde{F}_j^c - \tilde{F}_j^t \right) f_j^c, \quad (2)$$

where $\tilde{F}_j^t = \frac{1}{2} \left( F_{j-1}^t + F_j^t \right) = F_{j-1}^t + \frac{1}{2} f_j^t$, and $F_0^t = 0$.

Gini ranks distributions consistently with nonintersecting 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 considers 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).

Although there are various alternative ways to compute multigroup segregation (e.g., Silber, 1992; Reardon & Firebaugh, 2002; Alonso-Villar & Del Río, 2010), we offer here an aggregate index that is just the weighted average of all six possible pairwise comparisons among the four groups, with weights being the product of the corresponding two population shares (divided by the

---

\(^1\) Hutchens, M. (2004). "Measuring Segregation: The Distribution of Blacks and Whites by Occupation, 1970-2000." www.census.gov/prod/2004pubs/mar04com5.pdf
corresponding sum for all pairs). If \( K \) is the total number of groups and \( p^k \) is the population share of group \( k \), the average multigroup segregation index is given by:

\[
SM(f) = \frac{1}{\sum_{k=1}^{K-1} \sum_{j=i+1}^{K} p^k p^j} \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} S(f^i, f^j) p^i p^j. \tag{3}
\]

3.2 Segregation into low-paying occupations

Gradín (2017) adapted the previous approach to measure the extent to which one group, blacks, tends to be systematically over-represented in low-paying jobs. This low-pay segregation implies stratification in occupations (also referred to as vertical or ordinal segregation). The approach basically consists in reranking the distribution of occupations by a measure of their quality (e.g., the average earnings \( w_j \)). We call \( g^i_j \) and \( G^i_j \) the relative frequency and cumulative frequency of workers from group \( i \) in occupation \( j \) in this reranked distribution. 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 changes 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 concentration curve \( G^r(p), p \in [0, 1] \) plots the cumulative proportion of workers from both groups (with occupations sorted by \( w_j \)): \( G^i_j \) in the horizontal axis and \( G^r_j \) in the vertical axis (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 \( (G^c_j \geq G^r_j) \) over the target range, and that there is first-order stochastic dominance of \( r \) over \( c \). If there is no segregation or if the labor market is segmented but with both groups in occupations providing 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 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 using \( g \) instead of \( f \) in the geometrical interpretations, to quantify low-pay segregation and to rank distributions accordingly when the concentration curves overlap. The Gini concentration index,

\[
G_{\text{Gini}}(g^c, g^r) = 2 \sum_{j=1}^{J} \left( \hat{G}^c_j - \hat{G}^r_j \right) g^r_j, \tag{4}
\]

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,
\[ D(g^c, g^r) = G_c^s - G_r^s; \text{where}\left| G_c^s - G_r^s \right| = \max_{j \in [1,J]} \left\{ \left| G_j^c - G_j^r \right| \right\}, \tag{5} \]
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 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 negative value (when 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-pay (high-pay) 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, \( r_{\text{Gini}} = \frac{S(g^c, g^r)}{S(f^c, f^r)} \), \( S = \text{Gini}, \ D \). In particular, \( r_{\text{Gini}} \) is the Gini correlation coefficient between groups’ ratio and average earnings across occupations, computed among members of group \( c \). These concentration indices (and ratios) are symmetric in their absolute values. Exchanging group labels (which is the reference, and which is the comparison group) will just change their sign.

The absolute value of low-pay segregation indicates the level of stratification (regardless of which group is over-represented in low-paying occupation). An average multigroup stratification index can be obtained also as the weighted sum of all pairwise comparisons:
\[ S_M(g) = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \frac{1}{p^i p^j} \sum_{i=1}^{K} \sum_{j=i+1}^{K} S(g^j, g^i) |p^i p^j|. \] \tag{6}

### 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 Gradin (2013a, 2014), who adapted DiNardo, Fortin, and Lemieux’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 based on race, we construct a counterfactual distribution \( f^c \). In this counterfactual, individual observations of the comparison group (blacks) 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. In this counterfactual, both races will exhibit the same distribution of types (workers with any given combination of characteristics, such as holding a university degree, living in urban Western Cape, etc.), but each race keeps its own distribution across occupations conditional on type.

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:
\[ S(f', f') = S^E + S^U = [S(f', f') - S(f', f')] + S(f', f'), \]

where \( S^E = [S(f', f') - S(f', f')] \) is the level of segregation explained by both population groups having different distributions of characteristics (types of workers). \( S^U = S(f', f') \) is the unexplained term that compares the counterfactual (rewighted sample of blacks) with whites. That is, it indicates the level of conditional segregation that remains after equalizing the distribution of types in both groups, depending on how much the labor market segregates (based on race) people with similar observed characteristics. The identification of the unexplained term with discrimination in the labor market, however, must be cautious, as 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 discrimination in the labor market anticipated 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 us to identify the main determinant factors. Given the nonlinear 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 sequences (Gradín, 2014, using a Shapley decomposition: Chantreuil & Trannoy, 2013; Shorrocks, 2013).

The same exercise is done with segregation into low-paying occupations, after replacing \( f \) with \( g \) in (7).

4 | DATA

Censuses have been conducted in South Africa since 1911, but those prior to the 1994 democratic elections (the last one in 1991) are problematic, especially regarding the black population. The 1996 census was the first one covering the entire country and treating all population groups equally (e.g., StatsSA, 2007). Our main empirical analysis uses microdata samples from the 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—International, 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 apartheid. Whites are those with European ancestry (mostly Dutch and British), making up 9% of the population in 2007 (16% of workers). Blacks (or Africans) are the largest population group (80% of the population and 69% of workers in 2007) and are mostly the descendants of Bantu farmers who have migrated 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.

For the sake of inter-temporal comparability, the final sample is composed of white and black individuals not living in group quarters, 15 to 64 years old, who are employed, and not in the
Armed Forces. This implies a total of 1,414,812 individual observations with the following distribution 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 (33,268 whites, 160,967 blacks) in 2007. The other nonwhite categories (Coloured and Indian/Asian) account for 11% and 4% of workers and will be considered in the analysis of multigroup segregation and stratification.

Our main results use the IPUMS harmonized three-digit International Standard Classification of Occupations (ISCO-1988) with 125 categories, including one for those with occupations not classified elsewhere or unknown, which is problematic given its considerable importance, especially in 2007 (16% compared with around 7% in the previous years). For robustness, we also produced results for IPUMS one- and two-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 a person’s annual income in Rands for the 12 months prior to the census. Given the strong stratification of the South African labor market, it is not straightforward whose income we should consider ranking 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.

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), gender, 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, & Wittenberg, 2016) 1994 to 2015, from the DataFirst portal at University of Cape Town, to have more detailed information over time and for the sake of robustness. 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, 2008–2015). The 15 to 64 year-old sample consists of 997,658 observations (841,975 blacks, 155,683 whites), with 45,348 observations per year on average (38,272 blacks and 7,077 whites), but with great variability across years: from a minimum of 10,897 in 1996 (8,786 blacks and 2,111 whites) to a maximum of 81,985 in 2008 (70,178 blacks and 11,807 whites). We used ISCO-1988 occupations at the three-digit classification, even if the smaller sample sizes, compared with census data, impose some cautions about the more detailed results (for that reason, results using the one- and two-digit classifications are reported in the online Appendix—for access see Supporting Information at the end of the paper). There are some additional problems, like the under-representation of some informal workers, especially during the first years, or some inconsistencies in the codification of jobs over time (see Gradín, 2018). PALMS also has an estimation of real earnings that will be used to rank occupations, after some adjustments. The conditional analysis with PALMS will use a more restricted set of worker characteristics (i.e., omitting immigration, disability, and head/spouse status).

Given the problems that both monetary variables in IPUMS and census present (e.g., a substantial number of zeroes, being reported in intervals), it is important to note that with this approach, we are only using the information about the ranking of occupations, not the point estimates of earnings used in other analyses (e.g., wage discrimination). In this line, it is noteworthy that there
is a high weighted correlation between the occupational rankings produced by IPUMS and PALMS monetary variables (e.g., about 92% in 2007), and between these and the rankings produced by education (e.g., 91% in 2007 census for the proportion of workers with secondary school or higher). Thus, we can read the results of stratification as indicating segregation into low-paying or into low-skilled jobs indistinctly.

5

SEGREGATION AND STRATIFICATION TRENDS AFTER APARTHEID: BLACKS AND WHITES

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 1a). The 2007 curve entirely falls above the corresponding curve in 1996, which means that upon agreement on only four basic principles, one can say that there was an unambiguous decline in segregation that will be confirmed by most indices of segregation. However, the story is different by subperiods. The 2001 segregation curve generally falls below that of 1996, except at the top decile of occupations with the largest over-representation of whites. This means that segregation increased around occupations already disproportionally filled by blacks 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.
FIGURE 2  Racial occupational segregation indices (blacks and whites) observed (a) and unexplained or conditional on characteristics (b)

Source. Own construction based on IPUMS-International (census) and PALMS (LFS) [Colour figure can be viewed at wileyonlinelibrary.com]
The use of indices allows us to quantify the intensity of segregation in each year (Figure 2a, Table 1, and Table A6 in the online Appendix). The increase in segregation between 1996 and 2001 was about 2% (Gini) or 6% ($D$), while the decrease in segregation between 2001 and 2007 was about 13% or 16%, respectively. The net reduction for the entire 1996 to 2007 period was of nearly 11% 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 one- and two-digit classifications.

The LFS data reflect an even more pessimistic trend in segregation for 1994 to 2015 (Figure 2a, Table 2). We can distinguish an initial intense decline in segregation between 1994 and 1997, right after the end of apartheid. Note, however, that this initial period in PALMS is the most problematic for the under-representation of some low-skilled workers (see Gradín, 2018). This decline was followed by various oscillations according to the business cycle. It is, however,
| Year | Segregation | Low-pay segregation |
|------|-------------|---------------------|
|      | Uncond. | Conditional | All | Locat. | Educ. | Demo. | Uncond. | Conditional | All | Locat. | Educ. | Demo. |
| 1994 | 0.785   | 0.624       | 0.161 | 0.028 | 0.146 | 0.013 | 0.634   | 0.294       | 0.340 | −0.016 | 0.376 | −0.020 |
| 1995 | 0.730   | 0.559       | 0.170 | 0.022 | 0.157 | 0.009 | 0.605   | 0.233       | 0.372 | 0.002  | 0.380 | −0.010 |
| 1996 | 0.698   | 0.550       | 0.148 | 0.014 | 0.145 | 0.011 | 0.584   | 0.223       | 0.361 | −0.002 | 0.368 | −0.005 |
| 1997 | 0.681   | 0.503       | 0.178 | 0.015 | 0.157 | −0.006| 0.562   | 0.213       | 0.349 | −0.006 | 0.355 | 0.000 |
| 1998 | 0.740   | 0.549       | 0.191 | 0.027 | 0.166 | 0.002 | 0.613   | 0.250       | 0.363 | 0.008  | 0.359 | −0.004 |
| 1999 | 0.740   | 0.536       | 0.204 | 0.024 | 0.193 | 0.013 | 0.609   | 0.257       | 0.352 | −0.021 | 0.382 | −0.009 |
| 2000 | 0.765   | 0.547       | 0.218 | 0.030 | 0.195 | 0.007 | 0.675   | 0.229       | 0.446 | 0.010  | 0.416 | 0.020 |
| 2001 | 0.756   | 0.540       | 0.216 | 0.026 | 0.194 | 0.005 | 0.682   | 0.303       | 0.379 | 0.005  | 0.363 | 0.011 |
| 2002 | 0.762   | 0.550       | 0.212 | 0.030 | 0.193 | 0.011 | 0.660   | 0.285       | 0.375 | 0.003  | 0.367 | 0.004 |
| 2003 | 0.759   | 0.527       | 0.231 | 0.025 | 0.203 | −0.004| 0.677   | 0.295       | 0.382 | 0.000  | 0.364 | 0.018 |
| 2004 | 0.749   | 0.527       | 0.222 | 0.010 | 0.209 | −0.003| 0.670   | 0.324       | 0.345 | −0.016 | 0.338 | 0.023 |
| 2005 | 0.748   | 0.557       | 0.191 | 0.012 | 0.174 | −0.004| 0.650   | 0.274       | 0.376 | −0.002 | 0.363 | 0.015 |
| 2006 | 0.725   | 0.538       | 0.186 | 0.004 | 0.175 | −0.007| 0.654   | 0.333       | 0.321 | −0.014 | 0.313 | 0.022 |
| 2007 | 0.710   | 0.527       | 0.183 | 0.001 | 0.177 | −0.005| 0.625   | 0.304       | 0.321 | −0.022 | 0.316 | 0.027 |
| 2008 | 0.742   | 0.493       | 0.250 | 0.053 | 0.188 | −0.009| 0.660   | 0.277       | 0.383 | 0.039  | 0.299 | 0.046 |
| 2009 | 0.745   | 0.509       | 0.235 | 0.046 | 0.187 | −0.003| 0.660   | 0.275       | 0.385 | 0.035  | 0.303 | 0.047 |
| 2010 | 0.737   | 0.508       | 0.228 | 0.044 | 0.178 | −0.006| 0.670   | 0.276       | 0.394 | 0.042  | 0.303 | 0.050 |
| 2011 | 0.737   | 0.507       | 0.229 | 0.046 | 0.178 | −0.006| 0.678   | 0.305       | 0.373 | 0.045  | 0.284 | 0.044 |
| 2012 | 0.716   | 0.498       | 0.218 | 0.054 | 0.154 | −0.010| 0.650   | 0.299       | 0.352 | 0.057  | 0.257 | 0.038 |
| 2013 | 0.716   | 0.480       | 0.236 | 0.063 | 0.162 | −0.012| 0.640   | 0.293       | 0.347 | 0.056  | 0.251 | 0.040 |
| 2014 | 0.715   | 0.505       | 0.210 | 0.054 | 0.150 | −0.006| 0.637   | 0.297       | 0.340 | 0.056  | 0.244 | 0.041 |
| 2015 | 0.732   | 0.505       | 0.227 | 0.033 | 0.185 | −0.009| 0.653   | 0.288       | 0.366 | 0.004  | 0.313 | 0.049 |

*Note.* Locat. = Location [province and rural/urban (not available in 2005–07)]; Educ. = Education; Demo. = Demographics [age, sex, marital status].  
*Source.* Own construction based on PALMS.
discouraging to find out that the level of segregation in 2015 was still similar to or only slightly below that achieved right at the end of apartheid, and substantially above that in 1997 (at least 7% and 14% with Gini and $D$).\(^8\) Comparable results are obtained with the two-digit classification (Tables A2 and A3 in the online Appendix).

We analyze the robustness of the trend in segregation to removing observations with occupation unknown or not classified elsewhere, instead of considering this group as an independent category. The trend is very similar (cf. Figure 3, Table A4). The reduction in segregation (three-digit) using census data is just a bit smaller: about 8% between 2001 and 2007, 3% to 7% for the overall period. An almost identical result is obtained if we, instead, implicitly impute the occupation of these workers based on their observed characteristics, using a similar reweighting procedure as the one used to estimate conditional segregation.\(^9\) This is an indication of little impact of workers sorting based on observables on segregation.

5.2 The segregation of blacks into lower-paying occupations

We now address the issue of the quality of occupations held by blacks, by looking at the concentration curves and indices. Using census data, Figure 4a shows that blacks 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 distribution). Between 1996 and 2001 there was an increase in the segregation of blacks into low-paying occupations for almost the entire range of earnings, although the concentration curves cross at the 97th percentile of black workers (this implies a small improvement for blacks in high-paying occupations). The situation improved between 2001 and 2007 (the latter curve is always above the former). For the entire period, 1996 to 2007, the curves cross...
FIGURE 6  Segregation of blacks into low-paying occupations (blacks and whites) observed (a) and unexplained or conditional on characteristics (b)

Source. Own construction based on IPUMS-International (three-digit classification) [Colour figure can be viewed at wileyonlinelibrary.com]
twice (at the 83rd and 90th percentiles), showing a general improvement in the situation of blacks at the bottom and top of the distribution, but with some deterioration in the middle.

From Figures 5a to 5c, we can infer that almost all segregation of blacks 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 blacks in an occupation is a very good predictor of how low it pays on average. This correlation intensifies 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 between 2001 and 2007 they decreased at a similar rate (Figure 6a, Table 1). Blacks faced less segregation in 2007 than in 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%, respectively, between 2001 and 2007. The reductions for the entire period were of only 6% and 7%, respectively.

The results using labor force data (Figure 6a, Table 2) show that unnormalized levels of segregation into low-paying occupations turned out to be also quite persistent in the long run. The degree of stratification is higher in 2015 than in the 1990s, with oscillations along the business
cycle in between (exhibiting short periods of decline, e.g., between 1994 and 1997, 2000 and 2002, 2003 and 2007, or 2011 and 2014 using Gini).

These results using PALMS also confirm that most segregation of blacks is into low-paying occupations, with the fraction tending to increase over time, from 81% in 1994 with both indices, to 89% (Gini) and 95% (D) in 2015. A similar picture is obtained with the two-digit classification (reported in Tables A2 and A3 in the online Appendix).

As in the case of segregation, the improvement for blacks is 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 (cf. Figure 7, Table A4). A similar trend can be found when the sample of workers with known occupation of each race and year is reweighted to reproduce the characteristics of the corresponding entire sample. In the case of LFS, the impact of workers with unknown occupation is small and concentrated in earlier years.

5.3 Conditional racial segregation and stratification

The segregation of black African workers across occupations, and their over-representation in low-paying occupations, could be to some extent the result of inequalities of other kinds (geographical, demographic, educational, etc.) that occurred before the entrance into the labor market, whether the result of previous or anticipated discrimination or not (see Table A1 in the online Appendix). Blacks, compared with whites, are under-represented in urban areas, in the 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 (black spouses are less likely to be employed). To see how much segregation is due to the different distribution of characteristics across races, we compare observed segregation and segregation in the counterfactual situation in which blacks are given the same distribution of characteristics as whites using census data first.

Only 29% of segregation in 2007 is directly associated with differences in observed characteristics between blacks and whites (i.e., 0.175 with Gini and 0.130 with D). 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 immigration profiles. This means that a large 71% of segregation—that is, 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 blacks into low-paying occupations, nearly 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. The racial gap in attained education has also been identified as an important factor in explaining the racial wage gap (e.g., 36% in Keswell, 2010), part of which is precisely through the stratification of occupations.

These large unexplained terms in segregation and stratification are the result of differences in the conditional occupational distributions, with similar workers working in different occupations based on their race. This may be the result of direct discrimination by race in hiring or promotion practices, but it could also be the result of other unobservables, such as differences in preferences by race, or of the lower quality of education received by blacks (in line with the increasing role of differences in the returns to education in explaining the earnings gap).
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 1b), while the level of the unexplained segregation indices increased (Figure 2b), 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 implies a large decline in unexplained segregation indices, accounting for the decrease in segregation with $D$. It also accounts for nearly 90% of the decrease with Gini, although in this case there was also a substantial reduction associated with the relative improvement in education of blacks (which 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 nonoverlapping curves).

In the case of unexplained stratification, concentration curves are less informative because they overlap in all periods. The indices, however, produce clear results. There is an increase in the level of low-pay segregation conditional on worker characteristics between 1996 and 2007 (Figure 5b). The increase in the level of low-pay segregation between 1996 and 2001 was entirely driven by large increases in the unexplained term, only partially compensated by reductions in the explained term (mostly because of education upgrading among blacks). 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 (56%)

**FIGURE 8** Multigroup segregation and stratification (Gini)

*Note.* Stratification is the absolute value of LP segregation. Multigroup Gini is the weighted average of all pairwise comparisons (weights are the product of the corresponding population shares divided by the sum).

*Source.* Own construction based on IPUMS-International and PALMS [Colour figure can be viewed at wileyonlinelibrary.com]
in the case of D. As a result, we do find evidence that the improvement in the level of education of blacks helped to push down racial stratification of occupations in both periods: a total reduction of about 34% to 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 period.

The same conditional analysis using LFS data instead (Figures 2b and 5b, Table 2) shows comparable results. About 31% of segregation in 2015 is related with differences in workers’ characteristics (education explaining the largest part, 25% of segregation). The level of unexplained segregation in 2015 is smaller than in 1994, but similar to the one already observed in 1997. Similarly, 56% of low-pay segregation in 2015 is related with differences in workers’ characteristics (education explaining the largest part, 48% of low-pay segregation). This level was similar in 1994, but has increased since 1997. This trend contrasts with the reduction in the level of low-pay segregation explained by education. This means that while the better education of blacks contributed to reduce the level stratification between 2000 and 2014, this trend was compensated by the strong increase in the unexplained term between 2000 and 2006, paradoxically a period in which an important piece of affirmative action was enacted (the 2003 Black Economic Empowerment Act).

6 | MULTIGROUP SEGREGATION AND STRATIFICATION

The previous analysis has focused on the trend of segregation and stratification between blacks and whites. These are not only the two main population groups, but also represent the two ends of the socioeconomic scale. South Africa is a more diverse society, though, with the coexistence of two other population groups: Coloureds and Indians/Asians. For that reason, Figure 8 adds to the estimation of black–white segregation and stratification discussed above, the corresponding multigroup levels.

Given that segregation and stratification is highest between blacks and whites than between any other pair of groups, the multigroup levels are lower than the values showed for the two main groups. The total reduction in multigroup segregation over time was larger than between whites and blacks, 17%, with virtually no reduction in stratification (the reduction during the second period just compensated the increase during the first one).

### TABLE 3 Multigroup segregation and stratification (Gini)

|                  | White vs. Indian/Asian | White vs. Coloured | Indian/Asian vs. Black | Coloured vs. Black | Indian/Asian vs. Coloured | Multigroup: Unweighted Average | Multigroup: Weighted Average* |
|------------------|------------------------|-------------------|------------------------|-------------------|--------------------------|-------------------------------|-----------------------------|
| **Segregation**  |                        |                   |                        |                   |                          |                               |                             |
| 1996             | 0.672                  | 0.370             | 0.618                  | 0.621             | 0.362                    | 0.526                         | 0.528                       | 0.561                       |
| 2001             | 0.685                  | 0.279             | 0.605                  | 0.620             | 0.300                    | 0.517                         | 0.501                       | 0.540                       |
| 2007             | 0.599                  | 0.228             | 0.505                  | 0.556             | 0.250                    | 0.449                         | 0.431                       | 0.467                       |
| **Stratification** |                       |                   |                        |                   |                          |                               |                             |
| 1996             | 0.606                  | 0.209             | 0.568                  | 0.461             | 0.027                    | 0.419                         | 0.382                       | 0.399                       |
| 2001             | 0.641                  | 0.163             | 0.583                  | 0.517             | 0.048                    | 0.453                         | 0.401                       | 0.428                       |
| 2007             | 0.567                  | 0.112             | 0.473                  | 0.473             | 0.127                    | 0.371                         | 0.354                       | 0.398                       |

*Weights are the product of the population shares of the two groups divided by the sum of the products of all pairs.

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

Note. Stratification is the absolute value of LP segregation (the first group in the pair is the group segregated into low-pay occupations).
Looking into the details (cf. Table 3), there was a reduction in the level of segregation for all pairs of groups. The improvement in integration was however small. The reduction was largest among groups that are closer in socioeconomic terms (38% between whites and Indians/Asians; 31% between blacks and Coloureds), and smaller between most distant groups (about 10% between blacks and whites or Indians/Asians). The reduction was intermediate between Coloureds and whites or Asians (respectively, 18% and 15%).

In the case of stratification, there was a stronger reduction between nonblack groups, especially between whites and Indians/Asians. The limited improvement of blacks with respect to whites contrasts with the increasing stratification with Asians and, especially, Coloureds. The former group, though, showed a significant improvement with respect to whites and to Indians/Asians during the second period.

7 | CONCLUDING REMARKS

Discriminatory legislation and social practices in pre-democratic South Africa led 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 United States or Latin American countries, indicate that removing all discriminatory legislation is not enough to eradicate racial discrimination and segregation. Discrimination may persist before entrance into the labor market in the form of a lower amount and quality of education for nonwhites or the negative influence of ghettos or family background, compromising nonwhites’ 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 compelling evidence to supporting the idea that the distribution of occupations has been effectively either desegregated or destratified in post-apartheid South Africa. There is limited evidence of some improvement only over short periods, during the first years of democracy in the mid-1990s or during the early 2000s, but not of the required sustained path over time. It seems, in fact, that the situation in 2015 was not better than it was in the 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 blacks in the lowest-paying jobs. A multigroup analysis, however, has shown that other nonwhite population groups (Coloured and, especially, Indian/Asian) have more significantly improved their relative position in the labor market when they are compared with whites.

Inequality in the distribution of workers’ characteristics, especially in attained education, explains less than a third of segregation and about a half of stratification between blacks and whites. The improvement in the characteristics of blacks over time has had a positive impact on segregation and stratification trends but has not been effective enough to reverse the inherited situation and have been counterbalanced by other forces. A large unexplained term remains or increases over time. Blacks tend to work in different and lower-paying occupations to whites with similar observed characteristics. Changes over time in this unexplained part generally drove the segregation and stratification trends. This happened, paradoxically, in a period in which a large set of affirmative action policies have been implemented after dismantling the apartheid segregative legislation. However, evaluating the impact of each of these policies on the overall labor market or on specific groups along other factors (structural transformation, the business cycle, etc.) goes beyond the scope of this paper, focused on assessing the long-term trend.
ACKNOWLEDGMENTS

I acknowledge the suggestions of two anonymous reviewers, as well as the support received from UNU-WIDER and from Agencia Estatal de Investigación/Fondo Europeo de Desarrollo Regional (FEDER) (ECO2016-76506-C4-2-R).

ENDNOTES

1 Volij (2018) and Mora and Silber (2018) provide reviews of the theoretical and empirical approaches in segregation.

2 With this method, 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).

3 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 to 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).

4 For inter-temporal consistency, we deleted 1,981 observations with unknown occupation because they mostly affected only two years (1996 and 1997), as well as 3,785 observations in 1999 labeled as “not applicable.” For robustness, we also report the results when workers with unknown occupation are considered in as independent occupational category. Observations are weighted using the cross-entropy weight derived by DataFirst from the Actuarial Society of South Africa (ASSA) 2008 demographic model. For details of these data, see Kerr and Wittenberg (2016).

5 PALMS constructed an inflation-adjusted labor earnings variable. There are some outliers and many observations with missing earnings. The latter affects all occupations in various years (1996, 2008–2009, and 2015), and some occupations using detailed classifications in others. For that, we used the median (instead of the mean), which is less sensitive to outliers, and imputed median earnings for occupations lacking such information using those in the previous available year. The series may be affected by changes over time, especially from OHS to LFS in 1999 to 2000 (see Wittenberg, 2014). The values for low-paying segregation for LFS 1996 will be interpolated based on the adjacent years because of the large number of occupations for which it was not possible to estimate the median earnings.

6 The differences between the 1996 (or 2001) and the 2007 curves are statistically significant at all percentiles (for statistical inference, see Table A1 in the online Appendix—see Supporting Information at the end of the paper).

7 The crossing at the 91st percentile is statistically significant.

8 The results with Labour Force Surveys corroborate the 1996 to 2001 increase and the 2001 to 2007 decrease found using census data, but with different intensities, such that the 1996 to 2007 period shows a net increase (2–8% with Gini, 9–15% with D, depending on the classification used). When workers with unknown occupations in Labour Force Surveys are included in an independent occupational category, segregation is smaller in 1996 to 1997 and 1999 than reported here (cf. Figure A1).

9 We created two samples for each race and year, one with workers with known occupation, the other one with all workers. In a first stage we estimated the probability of a worker belonging to the latter sample based on characteristics using a logit model. Then, we constructed new weights based on their predicted probabilities. Finally, segregation was computed using the sample with workers with known occupation, reweighted to reproduce the distribution of characteristics of the entire sample.

10 This may be influenced by the fact that the presence of blacks in an occupation pushes down the average income.

11 When workers with unknown occupations in Labour Force Surveys are included, low-pay segregation is much lower in 1996 and 1999 than reported here, with only a small differential in 1997 (cf. Figure A1).

12 Auxiliary logit regressions are reported in Table A6 in the online Appendix.

13 The explained terms are smaller if we use the alternative counterfactual and give blacks the conditional employment distribution of whites (by reweighting whites’ distribution to reproduce the characteristics of blacks): 16% to 18% (segregation) and 24% to 26% (segregation into low-paying jobs).
This is an area still to be explored. For example, no African country appears in a review of all correspondence studies on hiring discrimination conducted after 2005 (Baert, 2017).

ORCID
Carlos Gradín http://orcid.org/0000-0001-7744-0707

REFERENCES
Allanson, P., Atkins, J., & Hinks, T. (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., Atkins, J., & Hinks, T. (2002). No end to the racial wage hierarchy in South Africa? Review of Development Economics, 6(3), 442–459.
Alonso-Villar, O., & Del Rio, C. (2010). Local versus overall segregation measures. Mathematical Social Sciences, 60(1), 30–38.
Baert, S. (2017). Hiring discrimination: An overview of (almost) all correspondence experiments since 2005 (IZA Discussion Paper No. 10738). Institute of Labor Economics, Bonn, Germany.
Banerjee, A., Galiani, S., Levinsohn, J., McLaren, Z., & Woolard, I. (2008). Why has unemployment risen in the new South Africa? Economics of Transition, 16(4), 715–740.
Beinart, W., & Dubow, S. (Eds.). (1995). Segregation and apartheid in twentieth-century South Africa. London: Routledge.
Blackburn, R. M., & Jarman, J. (1997). Occupational gender segregation (Social Research Update 16—Spring 1997). Department of Sociology: University of Surrey, Guildford, U.K.
Brookes, M., & Hinks, T. (2004). The racial employment gap in South Africa. South African Journal of Economics, 72(3), 573–580.
Burger, R., Jafta, R., & Von Fintel, D. (2016). Affirmative action policies and the evolution of post-apartheid South Africa’s racial wage gap (UNU-WIDER Working Paper No. 66). UNU-WIDER, Helsinki, Finland.
Campbell, P. R. (1987). South Africa 1960 to 1986, a statistical view of racial differences. US Bureau of the Census, Center for International ResearchWashington, D.C.:.
Chantreuil, F., & Trannoy, A. (2013). Inequality decomposition values: The trade-off between marginality and consistency. Journal of Economic Inequality, 11(1), 83–98.
DiNardo, J., Fortin, N. M., & Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. Econometrica, 64(5), 1001–1044.
Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indexes. American Sociological Review, 20(2), 210–217.
Gradín, C. (2013a). Conditional occupational segregation of minorities in the U.S. Journal of Economic Inequality, 11(4), 473–493.
Gradín, C. (2013b). Race, poverty, and deprivation in South Africa. Journal of African Economies, 22(2), 187–238.
Gradín, C. (2014). Race and income distribution: Evidence from the USA, Brazil and South Africa. Review of Development Economics, 18(1), 73–92.
Gradín, C. (2017). Segregation of women into low-paying occupations in the US (UNU-WIDER Working Paper No. 89). UNU-WIDER, Helsinki, Finland.
Gradín, C. (2018). Occupational gender segregation in post-apartheid South Africa (UNU-WIDER unpublished manuscript). UNU-WIDER, Helsinki, Finland.
Hutchens, R. M. (2004). One measure of segregation. International Economic Review, 45(2), 555–578.
Jahn, J., Schmid, C. F., & Schrag, C. (1947). The measurement of ecological segregation. American Sociological Review, 12(3), 293–303.
Kerr, A., & Wittenberg, M. (2016). A guide to version 3.1 of the post-Apartheid labour market series (PALMS). DataFirst Portal, University of Cape Town, South Africa.
Kerr, A., Lam, D., & Wittenberg, M. (2016). Post-Apartheid labour market series (dataset), Version 3.1. DataFirst (producer and distributor), Cape Town, South Africa.
Keswell, M. (2010). Education and racial inequality in post-Apartheid South Africa. In P. Attewell & K. S. Newman (Eds.), Growing gaps: Educational inequality around the world. Oxford, U.K.: Oxford University Press.

Keswell, M., Girdwood, S., & Leibbrandt, M. (2013). Educational inheritance and the distribution of occupations: Evidence from South Africa. Review of Income and Wealth, 59(S1), S111–S137.

Kingdon, G. G., & Knight, J. (2004). Race and the incidence of unemployment in South Africa. Review of Development Economics, 8(2), 198–222.

Kingdon, G. G., & Knight, J. (2007). Unemployment in South Africa, 1995–2003: Causes, problems and policies. Journal of African Economies, 16(5), 813–884.

Mariotti, M. (2012). Labour markets during Apartheid in South Africa. Economic History Review, 65(3), 1100–1022.

Minnesota Population Center. (2015). Integrated public use microdata series, International. Version 6.4 (machine-readable database). University of Minnesota, Minneapolis, MN.

Mora, R., & Silber, J. (2018). Segregation: Empirical findings. In C. D’Ambrosio (Ed.), Handbook of research on economic and social well-being (pp. 504–530). Northampton, MA: Edward Elgar.

Parashar, S. (2008). Marginalized by race and place: Occupational sex segregation in post-Apartheid South Africa (Thesis). University of Maryland, College Park, MD.

Reardon, S. F., & Firebaugh, G. (2002). Measures of multi-group segregation. Sociological Methodology, 32(1), 33–67.

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.

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.

Silber, J. (1992). Occupational segregation indices in the multidimensional case: A note. Economic Record, 68(3), 276–277.

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. Pretoria: Statistics South Africa.

Treiman, D. J. (2007). The legacy of Apartheid: Racial inequalities in the new South Africa. In A. F. Heath & S.-Y. Cheung (Eds.), Unequal chances: Ethnic minorities in western labour markets. Oxford: Oxford University Press.

Treiman, D. J., McKeever, M., & Fodor, E. (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.

Volij, O. (2018). Segregation: Theoretical approaches. In C. D’Ambrosio (Ed.), Handbook of research on economic and social well-being (pp. 80–503). Northampton, MA: Edward Elgar.

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 No. 46). Cape Town: University of Cape Town, South Africa.

Wittenberg, M. (2014). Analysis of employment, real wage, and productivity trends in South Africa since 1994 (Conditions of Work and Employment Series No. 45). Geneva, Switzerland: International Labour Office.

Yu, D. (2009). The comparability of census 1996, census 2001 and Community Survey 2007 (Stellenbosch Economic Working Papers No. 21/09). Department of Economics and Bureau for Economic Research, University of Stellenbosch, South Africa.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section at the end of the article.

---

**How to cite this article:** Gradín C. Occupational segregation by race in South Africa after apartheid. Rev Dev Econ. 2019;23:553–576. [https://doi.org/10.1111/rode.12551](https://doi.org/10.1111/rode.12551)