# Higher Residual Wage Dispersion for White Workers in South Africa: Composition Effects or Higher Demand for Skill?

*Richard U. Agesa***
*Jacqueline Agesa*****
*Geoffrey Bongani*******

June, 2007

**JEL Classification:** O15, O55, J31
**Keywords:** Wages; Wage inequality; Labor market; South Africa

***** Corresponding author: Division of Finance and Economics, Marshall University, One John Marshall Drive, Huntington, WV 25755; Tel: (304) 696-2606 (voice-mail); Fax: (304) 696-3662; E-mail: agesa@marshall.edu

****** Division of Finance and Economics, Marshall University, One John Marshall Drive, Huntington, WV 25755; Tel: (304) 696-2607 (voice-mail); Fax: (304) 696-3662; E-mail: agesaj@marshall.edu

******* Department of Economics, University of Pretoria, Pretoria 0002, Republic of South Africa; Email: s24123341@tuks.co.za

Financial support for this project, including travel arrangements that made possible collaboration between the authors in Huntington and the University of Pretoria was provided by Economic Research Southern Africa (ERSA).

# Higher Residual Wage Dispersion for White Workers in South Africa: Composition Effects or Higher Demand for Skill?


## Abstract

*Apartheid polices in South Africa ensured that whites received more and better quality education than Africans. It is hence conceivable, consistent with human capital theory, that whites would have a higher return to measured skills and higher dispersion of unmeasured skills (residual wage dispersion). We test this hypothesis using the September 2005 Labor Force Survey. Our findings are consistent with the theory: whites have higher returns to measured skills as well as higher dispersion of unmeasured skills. We find that the latter is attributed to higher returns to unmeasured skills and contend the higher returns to unmeasured skills for white workers stems from higher returns to their measured skills, driven by relatively higher demand for skilled workers (whites) fueled by the current surge in economic growth in South Africa. Curiously, we find higher wage inequality amongst white workers, a surprising finding given privileges whites enjoyed under apartheid.*

**Higher Residual Wage Dispersion for White Workers in South Africa: Composition Effects or Higher Demand for Skill?**

## 1. Introduction

The sanctuary of apartheid provided whites in South Africa a shelter to engage in overt racial wage discrimination. In 1994, however, apartheid was dismantled and overt racial wage discrimination was outlawed. Since then, a flurry of research has emerged that examines racial wage differences for Africans (blacks), whites, coloreds, and Indians (Asians) in the post-apartheid era. This research takes on several, though not necessarily exclusive, strands: one strand examines whether the racial wage gap has waned (I. Woolard and C. Woolard, 2006). A second controls for the quality of education as a determinant of racial wage differences (Chamberlain and van der Berg, 2002). A third uses the canonical Oaxaca/Blinder (1973) decomposition to attribute racial wage gaps to either racial differences in productivity attributes or unobservable attributes (discrimination) (Allanson and Atkins, 2005; Brookes and Hinks, 2004; Allanson *et al*, 2002; Moll, 2000; Allanson, Atkins and Hinks, 2000; Sherer, 2000; Bhorat, 2000; Mwabu and Schultz, 2000; Hinks, 1999). A fourth (Burger and Jafta, 2006) extends the third by decomposing racial wage differences into components due to explained, unexplained, and unobservable components. The findings of these studies, though mixed, offer a resounding theme: the African/white racial wage gap, the focus of the current study, persists and may be attributed to racial differences in productivity attributes.

This paper contends that issues surrounding African/white wage differentials in South Africa deserve further inquiry. In particular, while we do not intend to reduce the significance of the above studies, we add a new dimension to the fourth strand of the above literature by invoking a postulate consistent with human capital theory: the notion that if individuals who

invest more in education have higher marginal returns to education than others, then the labor market price or return to schooling will be higher at higher levels of education; moreover, years of schooling are an imperfect proxy for true educational inputs and that the error term (residual wage dispersion) includes unmeasured aspects of educational inputs such as school quality (Mincer, 1974).[1] Consequently, residual wage dispersion would be larger at higher levels of schooling. For example, if the return to years of schooling is greater for individuals (or groups) with relatively more human capital skills, then it is reasonable to expect that the dispersion of unmeasured human capital skills, such as school quality, would be higher for such individuals (or groups) as well.

This phenomenon matches the South African setting very well. In particular, the apartheid Bantu Education Act of 1952 (Shepherd, 1955) ensured that whites received more and better quality of education than Africans, and it is hence conceivable whites would have relatively higher marginal returns to their measured human capital skills as well as a higher dispersion of their unmeasured human capital attributes (residual wage dispersion). We employ a semiparametric procedure analogous to that of Lemieux 2002, to test this hypothesis using the September 2005 Labor Force Survey (LFS).

Such an analysis is important for two reasons: first, racial differences in prices of both measured and unmeasured human capital skills may disadvantage Africans and it is useful to examine the impact of each on racial differences in wage dispersion. This would allow us, within the context of human capital theory, to possibly provide a theoretical explanation for the underlying sources of racial (African/white) wage differences in South Africa. Second, the

[1] Henceforth, 'price' and 'return' to human capital will interchangeably be used to mean the same thing.

analysis would allow us to explain the persistence of the African/white wage gap, despite policy interventions (such as affirmative action) designed to mitigate the racial wage gap since the fall of apartheid in 1994.

The empirical technique first decomposes the African/white wage gap into components due to differences regression coefficients, the distribution of covariates, and residuals and applies kernel density methods to provide clear visual representations of the African/white wage gap throughout the entire conditional distribution. Second, and more importantly, we decompose the variance of wages and covariates into components due to measured human capital skills (predicted variance) and unmeasured human capital skills (residual variance) to examine the impact of each on racial differences in wage dispersion. Differences in residual wage dispersion may be attributed to either differences in prices of unmeasured human capital skills (Juhn, Murphy and Pierce, 1993; hereafter JMP) or composition effects, the notion that workers are more experienced and possess more human capital (Lemieux, 2002). For this reason, we model the African/white difference in residual wage dispersion as a function of racial differences in unmeasured skills and racial differences in skill prices to identify possible sources of racial differences in residual wage dispersion. To the best of our knowledge, no previous study has performed such an analysis using data from South Africa.

Non-trivial findings suggest a considerable African/white wage gap in South Africa. White workers earn more than Africans by 1.4931 percentage points. We also find that the largest component of the wage gap is attributed to racial differences in the return to human capital attributes. We suggest that the underlying source driving the larger return to measured human capital skills for white workers may be a high growth in the demand for higher skilled workers relative to the increase in supply fueled by the current surge in economic growth in

South Africa. And since white workers possess relatively more skills, it is conceivable that the relative growth in the demand for higher skilled workers (whites) would exceed that for African workers, causing the returns to human capital skills to be higher for white workers.

More importantly, we find relatively higher residual wage dispersion for white workers, with much of the difference in dispersion attributed to racial differences in prices of unmeasured human capital skills. This is a significant finding and is consistent with human capital theory: whites possess relatively more human capital skills and have a higher return to their human capital attributes; consequently, we would expect residual wage dispersion to be higher for white workers. Furthermore, we find that most of the difference in the variance of covariates between African and white workers (between-group inequality) is explainable by differences in human capital skills between the two races.

These findings may have important policy implications: to the extent that racial differences in residual wage dispersion are due to racial differences in unmeasured human capital attributes, an effective policy to mitigate the African/white racial wage gap in South Africa would be to aggressively increase the quantity and quality of education for African workers. Finally, we find greater wage inequality (90/10 wage gap) amongst white workers, a surprising finding given privileges whites enjoyed under the apartheid system.

In what follows, section 2 describes the data, section 3 explains the methodology, section 4 discusses the findings, section 5 concludes.

**2. Data and descriptive statistics**

We use data from the September 2005 LFS conducted by Statistics South Africa. For each race, African and white, we divide the data into $j$ cells with age and education categories

as our basic covariates, that is $x_{ir} = [x_{i1},...,x_{ijr},...,x_{iJr}]$. In the September 2005 LFS age is coded in 5-year band categories. From here, we generate 10 age categories in 5 year bands for workers aged from 15 to 64. Education is also coded in categories and we generate 4 dummy variables: individuals with no education, individuals with general education (grades 1 through 13), individuals with further education (grades 14 through 20), and individuals with higher education (university education). Sample sizes for each age/education category for African and white workers are reported in Tables 1 and 2 respectively.

From the 10 age and 4 education categories for each race, we generate 40 age/education dummy variable cells as reported in Tables 1 and 2. The sample of African workers in Table 1 has 51,719 observations. However, a caveat is in order here: from Table 1 some cells have zero or possess too few observations to warrant any statistical analysis. The use of such small sample sizes may generate inconsistent and inefficient estimators. For this reason, we drop from the sample cells that have zero, or fewer than 20 observations. After these adjustments, we end up with 38 age/education dummy variable cells for African workers with 51, 712 observations.

The sample of white workers is considerably smaller with 4,686 observations as reported in Table 2. Moreover, the sample of white workers is characterized by a significant feature: few (if any) white workers possess no education. In other words, there are few white workers with no formal education, that is education category one (column 1). At first glance, it may seem that the data for white workers is truncated at the lower end of the education distribution. However, historical factors may have resulted in an imbalance in education for African and white workers. In particular, the Bantu Education Act of 1952 ensured that whites received more and better quality of education than Africans. Under this act, whites enjoyed compulsory education from

age seven to sixteen, and while education was compulsory for Africans from age seven to thirteen, school enrollment was strictly enforced and encouraged for whites but rarely enforced for Africans. As a result, almost all whites received education while the same was not true for Africans, resulting in an education imbalance between the two races as is reported in columns 1 of Tables 1 and 2.

Therefore, an important underlying characteristic of the data is few (if any) white workers possess no schooling. For this reason, we drop from the sample education category one for white workers. And like in the sample for African workers, we drop from the sample of white workers cells with zero or less than 20 observations. After these adjustments we have 28 age/education dummy variable cells for white workers with 4,666 observations.

**3. Empirical Strategies**

Our empirical strategy mimics that of Lemieux (2002) which analyzes group differences in the overall distribution of wages. The advantage of the Lemieux framework is that since relative prices of human capital skills may differ by race, it is useful to decompose the variance of wages and variance of covariates (for each race) into components due to measured human capital skills (predicted variance) and unmeasured human capital skills (residual variance), thus allowing us to gauge the relative influence of each component on racial differences in wage dispersion throughout the entire condition distribution. Moreover, the technique lends itself to visual representation of wage distributions. Our first task is to compute the conditional mean wage for each race.

**3a. African and white mean log wages**

The sample average of each age-education dummy variable cell $x_{ijr}$ ($x$ indexes each age-education dummy variable category, $i$ is each individual, $j$ is each cell, and $r$ indexes race, African or white) is the sample-weighted proportion of the sample in cell $j_r$. We label this sample-weighted proportion as $\theta_{jr}$.

(1) $$\bar{x}_{jr} = \sum_i \omega_i x_{ijr} = \sum_{x_{ijt}=1} \omega_i = \theta_{jr}$$

where $\omega_i$ are sample-weights taken from the September 2005 LFS. The Ordinary Least Squares (OLS) estimates $b_r$ (with no intercept) are the means of log hourly wages $w_{ir}$ in cell $j$:

(2) $$b_r^{/} = [b_{1r},...,b_{jr},...,b_{Jr}] = [\overline{w}_{jr},...,\overline{w}_{jr},...,\overline{w}_{Jr}],$$

where $\overline{w}_{jr} = (\frac{1}{\theta_{jr}}) \sum_{x_{ijr}=1} \omega_i w_{ir}$.

The conditional mean hourly log wage for each race is computed as:

(3) $$\ln \overline{w}_r = \sum_i \omega_{ir} w_{ir} = \sum_j \theta_{jr} \overline{w}_{jr} = \sum_j \bar{x}_{jr} b_{jr} = \bar{x}_r b_r$$

where $\ln \overline{w}_r$ is the mean hourly real log wage across the entire sample of individuals with reported wages for each race. The covariates are age-education dummy variable cells (with no intercept) for African and white workers defined in Tables 1 and 2 respectively.

**3b. African and white variance of wages**

The variance of wages for each race is computed as:

(4) $$V_r = \sum_j \theta_{jr} (\ln \overline{w}_{jr} - \overline{w}_r)^2 + \sum_j \theta_{jr} \sigma_{jr}^2$$

where $V_r$ is the variance of wages for each race, $\overline{w}_r$ is the average log wage across the entire sample for each race, $\ln \overline{w}_{jr}$ is the average log wage in characteristic cell $j$ for each race, $\theta_{jr}$ is

the weighted-sample proportion in cell *j* for each race, and $\sigma_{jr}^2$ is the residual variance for each race in each cell. The sums, for each race, are defined over age-education dummy variable cells as reported in Tables 1 and 2.

For each race, the first term on the right hand side of equation 4 is the variance of cell means across cells and is the predicted or the between-cell variance of wages (weighted by cell-size), which constitutes the weighted sum of squared deviations of skills across cells. The second term is the residual or the mean of within-cell variance (weighted by cell-size), which is the weighted sum of the residual variance over skills in each cell.

In the human capital interpretation (Mincer, 1974), differences in human capital skills, $x$, of workers would affect the distribution of wages in two ways: first, if whites possess relatively more human capital skills, then we would expect a changes or higher predicted variance of wages, $\sum_j \theta_w (\ln \overline{w}_{jw} - \overline{w}_w)^2$, for white workers because the sample proportion, $\theta_{jw}$, would be relatively higher for white workers. Second, since the residual variance, $\sum_j \theta_{jr} \sigma_{jr}^2$, is a weighted sum of cell specific variances, it too should be higher for white workers due to differences in sample proportions, $\theta_{jr}$, that is the proportion of whites with relatively more human capital is higher. Moreover, in our dummy variable model, the sample proportions equal the sample average of human capital covariates $x$ (equation 1).

**3c. Counterfactual African wages and variances**

Our objective, however, is to decompose the entire conditional distribution of wages and not just focus exclusively on what happens at the conditional mean. To do so, we consider two possibilities. First, it is conceivable that prices or returns to human capital skills may differ for

each race. In particular, the returns to human capital may be higher for white workers, perhaps due to past apartheid polices. To account for this possibility, we compute a counterfactual mean wage ($\ln \overline{w}_A^a$) that would have prevailed for African workers if prices of human capital skills for African workers were the same as for white workers. This counterfactual wage is computed as follows:

(5) $$\ln \overline{w}_A^a = \overline{x}_{jA} b_w + u_{jw}$$

where *A* is African workers, *w* is white workers, and $u_{jw}$ is the vector of residuals. Similarly, the counterfactual variance of wages for African workers $(V_A^a)$, i.e. the variance of African wages that would have prevailed if prices of human capital were the same as for white workers is computed as:

(6) $$V_A^a = \sum_j \theta_{jA} (\ln \overline{w}_{jw} - \overline{x}_A b_w)^2 + \sum_j \theta_{jA} \sigma_{jA}^2$$

This counterfactual variance only affects the first sum since the only changes that occur are those in the parameters $b_w$. The difference between $V_A$ in equation 4 and $V_A^a$ in equation 6 captures differences in prices of covariates (age-education characteristics) between African and white workers in the September 2005 LFS.

The second possibility is that the distribution of wages may differ for white and African workers due to differences in the skill composition of the workforce, i.e. white workers (due to past apartheid polices) are more educated. To account for this possibility, we compute another counterfactual African wage, $\ln \overline{w}_A^b$, that would have prevailed if African workers possessed the same price of human capital and the same distribution of human capital as white workers computed as follows:

(7) $$\ln \overline{w}_A^b = \overline{x}_{jw} b_w + u_{jA}$$

Likewise, the counterfactual variance $(V_A^b)$ of wages, i.e. the variance of African wages with the returns to human capital and the distribution of human capital that matches that of white workers is computed as:

(8) $$V_A^b = \sum_j \theta_{jw} (\ln \overline{w}_{jw} - \overline{x}_w b_w)^2 + \sum_j \theta_{jw} \sigma_{jA}^2$$

This counterfactual variance affects both the first and the second sum since the proportion of the sample within each cell i.e. $(\theta_{jw})$ changes.

It is worthy of mention that the counterfactual mean wage $\ln \overline{w}_{jA}^b$ in equation 7 (the African counterfactual wage that would have prevailed if both the price and the distribution of human capital were the same as for white workers) also happens to be the same mean wage for white workers in equation 3. However, the counterfactual variance $V_A^b$ for African workers in equation 8 is not the same as $V_w$ in equation 4, the variance of wages for white workers, due to differences in the residual variances $\sigma_{jA}^2$ in equation 8 and $\sigma_{jw}^2$ in equation 4. Furthermore, drawing from Lemieux (2006), we assume that the returns to characteristics for the two racial groups are independent of the distribution of covariates. This assumption allows us to match the returns for white workers with the distribution of wages for African workers and generate counterfactual earnings distributions specified in equations 5 and 7.

**3d. Decomposing the African/white variance of wages**

From the above counterfactual exercises, we decompose the variance of wages between African and white workers as follows:

(9) $$V_A - V_w = (V_A - V_A^a) + (V_A^a - V_A^b) + (V_A^b - V_w)$$

The first term on the right-hand side represents differences in the variance explainable by racial differences in the price of human capital. The second term represents the additional difference in variance due to racial differences in the distribution of human capital, and the last term represents the difference in variance due to racial differences in the within-cell variance (or residual variance $u_i$). The sum of the first two represent how much of the racial difference in wage variance is due to between-group inequality changes, and the last represents how much of the racial difference in wage variance is due to within-group inequality changes.

**3e. Kernel density estimation**

Kernel density estimates (or smoothed histograms) provide visual representations of weighted raw and counterfactual wages. We use the kernel density estimator introduced by Rosenblatt (1956) and Parzen (1962) in which sample weights are attached to each observation. The weighting function is specified as:

$$\hat{f}_h(w) = \sum_{j=1}^{n} \frac{\omega_j}{h} K\left(\frac{w - W}{h}\right), \tag{10}$$

Where $\hat{f}_h$ is the kernel density estimate of a univariate density *f* based on a random sample $w_i$ $W_1,...,W_n$ of size *n* with weights $\omega_1,...,\omega_n$ ($\sum \omega_i = 1$). The weighting function $K(\cdot)$ is the kernel function and the kernel function used in this paper is Gaussian with weights $\omega_j$ taken from the September 2005 LFS. The smoothing parameter *h* is the bandwidth. An important decision is the choice of the bandwidth since too small a bandwidth leads to jagged density estimates; too large a bandwidth over-smooths the data. An appropriate bandwidth would provide a reasonably smooth density estimate. For this reason, we use the plug-in bandwidth method calculated by

Sheather and Jones (1991) selector. From our data, the optimal bandwidths for the weighted actual and counterfactual log earnings equations range from 0.0480 and 0.1857.

**4. Results and discussions**

Table 3 reports several descriptive and counterfactual log wage statistics for African and white workers. Sample weights from the September 2005 LFS are used throughout Table 3. From column 1, the mean hourly log wage is about 1.4931 percentage points higher for white workers. Visual representations of the African/white wage distributions using kernel density estimates are reported figure 1; both densities represent log wage distributions estimated using equation 3. The two densities are bell-shaped, with the kernel density for white workers showing a decisive rightward translation. The gap between the two densities reflects the 1.4931 racial earnings difference. The gap, however, is more pronounced in the middle of the distributions but converges at the top of the distributions, suggesting small wage variation between African and white workers at the very top of the skill distributions.

The kernel wage density for African workers has a coefficient of skewness of -0.0078 while the coefficient of skewness for the white kernel density is -0.4351. Although both are negative, suggesting that the mean wage is to the left of the median wage in both wage distributions, the differences between the lowest wage and the mean wage is much larger for white workers; that is, the lowest wage is relatively further away from the mean wage for white workers. This finding is consistent with the notion that wage inequality is higher for white workers, a notion further supported by the larger 90/10 wage gap (an alternative measure of wage inequality) for white workers in column 5 of Table 3. Furthermore, the variance of wages is higher for African workers by 0.2299 percentage points, column 2. This finding is not surprising and lends credence to the notion that, because there are hardly any whites with no

formal education, the distribution of covariates may be relatively more compact for white workers.

Our findings also suggest that the largest component of the racial wage gap is attributed to racial differences in the return to human capital skills, which constitute -1.3545 of the -1.4931 racial age gap (Column 1). This finding is not surprising and suggests that racial differences in human capital skills (brought about by previous apartheid polices) resulted in whites acquiring more skills, and consistent with human capital theory, the returns to human capital are higher for relatively more skilled white workers. We attribute the greater returns to human capital skills for white workers to a relatively greater demand for skilled workers relative to the supply (white workers possess more skills), fueled by the current surge in economic growth in South Africa.

The results from various counterfactual exercises are reported in rows 6, 7, and 8 of Table 1. From row 7, racial differences in covariates account for the remainder (-0.1386) of the racial wage gap, suggesting that whites possess more human capital skills than African workers. This finding is not surprising and is consistent with a priori expectations: previous apartheid polices (in particular the Bantu Education Act of 1952) ensured that whites received more education than Africans, culminating in racial differences in access to schooling.

More importantly, the variance of differences in covariates (0.1022) is positive and there are two offsetting reasons for this result. First, the predicted variance is positive but lower for white workers by 0.9276 percentage points, column 3. This result is contrary to our expectations: we would expect higher predicted variance for white workers because the sample proportion of whites with more skills would be higher. However, as suggested earlier, apartheid policies ensured that there were no whites without formal education. Consequently, the distribution of covariates may be relatively more compact for white workers, resulting in a lower

predicted (or between-group) variance for white workers. This result is consistent with our earlier finding of lower variance for white workers, hence lending further support to the notion that the distribution of covariates may be relatively compact for white workers.

Second, the residual variance of covariates is lower for Africans by -0.8254 percentage points, column 4, suggesting a higher residual dispersion of unmeasured attributes for white workers. Again, this finding is consistent with a priori expectations: the sample proportion of whites $\theta_{jw}$ with more skills is higher for white workers, and we would expect (consistent with human capital theory) whites to have a higher dispersion of unmeasured human capital attributes as well. Indeed, the residual variance, $\sum_j \theta_{jw}\sigma_{jw}^2$, of wages is a weighted sum of cell specific variances and it too should be higher for white workers due to differences in sample proportions, $\theta_{jw}$; that is the proportion of whites with relatively more skills is higher.

To see this point clearly, from section 3, this last effect (higher residual variance for white workers) is given by $\sum_j (\theta_{jA} - \theta_{jw})\sigma_{jA}^2$ which represents differences in the contribution of covariates or *X's* (or sample proportions $\theta_{jr}$) to both the between and within variance of covariates. In this setting, $\theta_{jA}$ and $\theta_{jw}$ represent the sample proportions of African and white workers respectively in each cell. In our sample, there are systematically fewer African workers in 'high dispersion' cells $\sigma_{jA}^2$ but more white workers in high dispersion cells $\sigma_{jw}^2$. In other words, white workers have more experience and are more educated and as predicted by human capital theory we would expect the dispersion of unmeasured human capital attributes to be higher for white workers. Indeed, the magnitude of the higher residual wage dispersion for

whites increases dramatically (0. 2502 percentage points) from 0.7451 in row 5 to 0.9953 in row 8 in the African-white gap in residual variance of wages.

*What is the source of the higher residual wage dispersion for white workers? Composition effects or increase in skill prices?*

The literature postulates that there are two possible sources for higher residual dispersion of wages: composition effects (Lemieux, 2002), and differences in prices of unmeasured (JMP, 1993). In the South African setting, composition effects emanate from the notion that white workers possess more experience and are also more educated. And since the return to measured human capital skills is higher for more skilled workers, we would also expect the dispersion of unmeasured human capital skills (such as school quality) to also be relatively higher for the more skilled white workers.[2]

On the other hand, the notion that racial differences in prices of unmeasured human capital skills is consistent with a higher return to human capital should be interpreted as evidence that the price of unmeasured human capital skills would be higher as well. In particular, white workers have a higher return to human capital would consequently receive a higher return to their unmeasured human capital attributes (JMP, 1993).

Consistent with Lemiuex, 2002, we test whether composition effects or differences in skill prices drive the higher residual wage dispersion for white workers by generating a counterfactual distribution of residuals for African workers that would have prevailed if African workers possessed the same price of measured human capital skills and the same skill pricing function ($p$) of unmeasured human capital attributes as white workers. This counterfactual wage

[2] Studies suggest a higher quality of education for South African whites (Moll, 1998).

residual is specified as $u_{iA}^{b} = p_w(\eta_{iA})$, where the skill pricing function $p$ takes a non-linear pricing scheme defined as $u_{ir} = p_r(\eta_{ir}) + \varepsilon_{ir}$, where $p_r(.)$ the price (or return) of unmeasured human capital and is a monotonic and continuous function, $\eta_{ir}$ is unmeasured human capital, and $\varepsilon_{ir}$ is the random error component not linked with skills and productivity attributes and is equal to zero.

We then re-estimate counterfactual wage equation 7 for African workers as if prices of measured human capital skills and the pricing function of unmeasured skills would have been the same as for white workers, specified as follows:

(13) $$\ln \overline{w}_A^b = \overline{x}_{jw} b_w + u_{iA}^b$$

The resulting counterfactual distribution should account for all differences between Africans and whites except for those involving a change in the rank of the residual in the overall distribution of residuals.[3] A word of caution is necessary here: combining the counterfactual predicted wage residual with the counterfactual predicted wage for African workers does not yield the same counterfactual distribution as the actual distribution for white workers because knowledge of the marginal distribution of $xb$ and $u$ is not generally enough to characterize the marginal distribution of log wages (Lemiuex, 2002) . In other words, an infinite number of marginal distributions of log wages is compatible with given marginal distributions of $xb$ and $u$ . Nonetheless, equation 13 would provide reasonable estimates of the counterfactual distribution of residuals for African workers that would have prevailed if African workers possessed the

[3] This is because the distribution of residuals for Africans and whites differ such that they cannot be captured by a common transformation function.

same price of measured human capital skills and the same skill pricing function ($p$) of unmeasured human capital attributes as white workers.

The results of the decomposition of wage residuals of counterfactual earnings equation 13 are reported in Table 4. The findings are interesting: whereas the largest component of the total variance of wages is explained by racial differences in covariates in panel A, the largest component of the change in the variance of wage residuals is explained by the skill-pricing function $p$, which constitutes 0.3841 of the 0.3076 change in the variance of wage residuals in panel B. This finding is consistent with human capital theory; that is, racial differences in the price of unmeasured human capital attributes drive much of the racial difference in residual wage dispersion. We contend, consistent with JMP (1993), that the higher price of unmeasured skills for white workers is a consequence of a higher return to their measured human capital skills. We attribute the higher prices of unmeasured human capital attributes to the current surge in economic growth in South Africa.

Kernel density estimates other counterfactual exercises are reported in figures 2 and 3. Figure 2 reports the counterfactual kernel wage density that would have prevailed if Africans had the same returns to human capital as white workers (equation 5). On the other hand, figure 3 reports the counterfactual kernel wage density that would have prevailed if African workers possessed not only the same returns to human capital skills but also possessed the same distribution of covariates as white workers (equation 7). In both figures, the counterfactual kernel densities have similar shapes and are consistent with the findings in Table 3: the counterfactual wages have a decisive rightward translation (relative to African wages), suggesting that whites have higher returns to human capital and possess more human capital attributes than African workers respectively.

**Racial differences in wage inequality**

Wage inequality, defined as the difference between the 90$^{th}$ and 10$^{th}$ percentile, is higher amongst white workers (column 5) in Table 3. Our findings also suggest that the largest component of racial wage inequality is attributed to racial differences in covariates (column 5). This finding is hardly surprising given that white workers possess relatively more human capital skills.

We also find that the 50/10 and 90/50 wage gaps are higher for white workers, lending further support to the notion that wage variation is higher amongst white workers. These results offer a curious phenomenon: there may be a large variation in earnings between high skilled and low skilled whites, suggesting higher returns for whites at the top of the wage distribution. More importantly, these results are consistent with our earlier finings of a relatively larger coefficient of skewness for the kernel density of white wages -0.4351, suggesting that the mean wage is to the left of the median wage and that the spread between the mean wage and the lowest wage is relatively larger for white workers.

**5. Implications and conclusions**

Apartheid polices in South Africa ensured that whites received more and better quality of education than Africans. The Bantu Education Act of 1952 perpetuated the concept of racial "purity" which rationalized keeping African education inferior. Under this act, Africans "were educated for their opportunities in life," and that there was no place for them "above the level of certain forms of labor." It is hence conceivable, consistent with human capital theory, that whites would have a higher return to measured human capital attributes as well as a higher dispersion of unmeasured human capital attributes (residual wage dispersion). Human capital theory further postulates that residual wage dispersion is due to unmeasured differences in human capital

investment and should differ when the price or return to human capital investment differs. For example, when the return to years of schooling differs by group, it is reasonable to expect that that the return to unmeasured human capital attributes, such as school quality, would differ for each group as well (Mincer, 1974).

We test this notion using data from the September 2005 Labor Force Survey. Specifically, we employ a semiparametric procedure that mimics that of Lemieux 2002 to decompose the African/white wage gap into components due to regression coefficients, covariates, and residuals and further model racial differences in residuals as a function of differences in unmeasured skills and skill prices. No previous study has, to the best of our knowledge, performed such exercise using data from South Africa; such an analysis is important because, despite the collapse of apartheid and recent policy interventions to mitigate the African/white wage gap, we may provide insights for the persistence of the wage gap.

Our findings are consistent with a priori expectations: whites earn higher wages, have higher returns to their measured human capital attributes, and have higher residual wage dispersion. We also find that the higher residual wage dispersion for white workers is attributed to higher prices of their unmeasured human capital skills. These findings are consistent with human capital theory: the higher return to unmeasured human capital skills for white workers is a consequence of the higher return to their measured human capital skills.

We contend that the fundamental source of the higher return to measured human capital attributes for white workers may be a greater demand for relatively higher skilled workers precipitated by strong economic growth in South Africa. Indeed, the supply of skilled African and white workers has increased in the post-apartheid era; however, whites have historically possessed more schooling, and with the current rise in economic growth, it is likely that the

demand for higher skilled workers (whites) would surge relative to the demand for African workers. This would drive up prices of measured human capital skills for white workers and consistent with human capital theory, increase the dispersion and prices of unmeasured human capital skills (residual wage dispersion) for white workers as well. Moreover, we find that most of the difference in the variance of covariates between African and white workers (between-group inequality) is explainable by differences in human capital skills between the two races. These findings may have significant policy implications: to reduce the African/white racial wage gap, an appropriate policy would be to increase the quantity and quality of education for African workers.

Curiously, we find greater wage inequality (the 90/10 wage gap) amongst white workers; we also find that the 90/50 and 50/10 wage gaps are higher for white workers in South Africa. These findings suggest that, despite a compact distribution of covariates for white workers, there is a large earnings difference between whites at the top and at the bottom of the wage distribution. These are indeed surprising findings, given the privileges whites enjoyed under the apartheid system.

**References**

Allanson, Paul and Atkins, Jonathan P., "The Evolution of the Racial Wage Hierarchy in Post-apartheid South Africa," *Journal of Development Studies*, vol. 41, no. 6, (August 2005: 1023-50.

Allanson, Paul and Atkins, Jonathan P and Hinks, Timothy, "No End to the Racial Wage Hierarchy in South Africa," *Review of Development Economics*, vol. 6, no. 3, (October 2002): 442-459.

Allanson, Paul and Atkins, Jonathan P and Hinks, Timothy, "A Multilateral Decomposition of Racial Wage differentials in the 1994 South African Labor Market," *Journal of Development Studies*, vol. 37, no. 1, (October 2000): 93-120.

Blinder, Alan, "Wage Discrimination: Reduced Forms and Structural Estimation," *Journal of Human Resources*, Vol. 8, No. 4, (Fall, 1973): 436-55

Bhorat, Harron, "Wage Premia and Wage Differentials in the South African Labour Market," Development Policy Research Unit, University of Cape Town (September 2000).

Berger, Rlof and Jafta Rachel, "Returns to Race: Labor Market Discrimination ion Post-Apartheid South Africa," Stellenbosch Economic Working Papers: 04/2006, Bureau of Economic Research and the Department of Economics, University of Stellenbosch, South Africa.

Brookes, Mich and Hinks, Timothy, "The Racial Employment Gap in South Africa," *South African Journal of Economics*, Vol. 72, No. 3, (September 2004): 573-597.

Chamberlin Doubell and van Der Ber, Servaas, "Earnings Functions, Labour Market Discrimination and Quality of Education in South Africa", Stellenbosch Economic Working Papers: 2/2002, Bureau of Economic Research and the Department of

Economics, University of Stellenbosch, South Africa.

Hinks, Timothy, "Racial Wage Discrimination and the End of Apartheid in South Africa: A Multilateral Approach." Mimeo, Middlesex University Business School, London, U.K., 1999.

Chinhui, Juhn and Kevin M. Murphy and Brooks, Pierce, "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, vol. 101, no. 3, (June, 1993): 410-442.

Lemieux, Thomas, "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill," *American Economic Review*, Vol. 96, No. 3, (June, 2006): 461-498.

Lemieux, Thomas, "Decomposing Changes in Wage Distributions: A Unified Approach," *Canadian Journal of Economics*, Vol. 35, No. 4, (November, 2002): 646-688.

Mincer, J, *Schooling Experience, and Earnings*, New York: National Bureau of Economic Research, Columbia University Press, 1974.

Moll, Peter G. "Discrimination is Declining in South Africa, but Inequality is Not." *Journal for Studies in Economics and Econometrics*, vol. 24, No. 3 (November, 2000): 91-108.

Moll, Peter G. "Primary schooling, Cognitive Skills and Wages in South Africa." *Economica*, vol. 65, No. 258 (May 1998): 263-284.

Mwabu, Germano and Schultz, T.P., "Wage Premiums for Education and Location for South African Workers by Gender and Race." *Economic Development and cultural Change*, vol. 48, No. 2, (2000): 307-34.

Oaxaca, Ronald, "Male-Female Wage Differentials in Urban Labor Markets," *International Economic Review*, Vol. 14, No. 3, (October, 1973): 693-709.

Parzen, Emanuel, "On Estimation of a Probability Function and Mode," *The Annals of*

*Mathematical Statistics*, Vol. 33, No. 3, (September, 1962): 1065-1076.

Rosenblatt, Murray, "Remarks on Some Non-Parametric Estimates of a Density Function," *The Annals of Mathematical Statistics*, Vol. 33, No. 3, (September, 1956): 832-837.

Sheather, S., and Jones, M, "A Reliable Data-based Bandwidth Selection Method for Kernel Density Estimation," *Journal of Royal Statistical Society*, *series B (Methodological)*, Vol. 53, No. 3, (1991): 683-690

Shepherd, R.H.W., "South African Bantu Education Act," *African Affairs,* Vol. 54, No. 215 (April 1955): 138-142

Sherer, George, "Intergroup Economic Inequality in South Africa: The Post-apartheid Era," *American Economic Review,* Vol. 90, No. 2, (May 2000): 317-21.

Woolard, Ingrid and Woolard, Chris., Earnings Inequality in South Africa 1995-2003, HSRC Press, Cape Town, South Africa (2006).

## Table 1

## Age-education categories: African

| Age group | Education Category 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|
| 15≤age≤19 | 81 | 10,444 | 37 | 0 | 10,562 |
| | 1.73 | 23.56 | 1.75 | 0.00 | 20.42 |
| 20≤age≤24 | 136 | 8,183 | 265 | 34 | 8,618 |
| | 2.91 | 18.46 | 12.50 | 5.73 | 16.66 |
| 25≤age≤29 | 211 | 5,816 | 430 | 83 | 6,540 |
| | 4.51 | 13.12 | 20.28 | 14.00 | 12.65 |
| 30≤age≤34 | 284 | 4,687 | 425 | 97 | 5,493 |
| | 6.07 | 10.57 | 20.05 | 16.36 | 10.62 |
| 35≤age≤39 | 386 | 3,962 | 364 | 131 | 4,843 |
| | 8.25 | 8.94 | 17.17 | 22.09 | 9.36 |
| 40≤age≤44 | 575 | 3,453 | 241 | 96 | 4,365 |
| | 12.29 | 7.79 | 11.37 | 16.19 | 8.44 |
| 45≤age≤49 | 684 | 2,839 | 180 | 76 | 3,779 |
| | 14.62 | 6.40 | 8.49 | 12.82 | 7.31 |
| 50≤age≤54 | 813 | 2,135 | 97 | 52 | 3,097 |
| | 17.38 | 4.82 | 4.58 | 8.77 | 5.99 |
| 55≤age≤59 | 764 | 1,661 | 55 | 17 | 2,497 |
| | 16.33 | 3.75 | 2.59 | 2.87 | 4.83 |
| 60≤age≤64 | 744 | 1,148 | 26 | 7 | 1,925 |
| | 15.90 | 2.59 | 1.23 | 1.18 | 3.72 |
| Total | 4,678 | 44,328 | 2,120 | 593 | 51,719 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

**Notes:**
**For education categories:**
1: Individuals with no education
2: Individuals with general education (grades 1 through 13)
3: Individuals with further education (grades 14 through 20)
4: Individuals with higher education (university education)

## Table 2

## Age-education categories: Whites

**Education Category**

| **Age group** | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|
| 15≤age≤19 | 3 | 466 | 5 | 4 | 478 |
| | 27.27 | 14.86 | 0.54 | 0.65 | 10.20 |
| 20≤age≤24 | 2 | 321 | 76 | 43 | 442 |
| | 18.18 | 10.24 | 8.22 | 7.00 | 9.43 |
| 25≤age≤29 | 1 | 255 | 78 | 46 | 380 |
| | 9.09 | 8.13 | 8.43 | 7.49 | 8.11 |
| 30≤age≤34 | 0 | 305 | 121 | 64 | 490 |
| | 0.00 | 9.73 | 13.08 | 10.42 | 10.46 |
| 35≤age≤39 | 1 | 325 | 119 | 86 | 531 |
| | 9.09 | 10.36 | 12.86 | 14.01 | 11.33 |
| 40≤age≤44 | 0 | 284 | 109 | 92 | 485 |
| | 0.00 | 9.06 | 11.78 | 14.98 | 10.35 |
| 45≤age≤49 | 1 | 312 | 121 | 100 | 534 |
| | 9.09 | 9.95 | 13.08 | 16.29 | 11.40 |
| 50≤age≤54 | 2 | 301 | 120 | 72 | 495 |
| | 18.18 | 9.60 | 12.97 | 11.73 | 10.56 |
| 55≤age≤59 | 1 | 319 | 101 | 52 | 473 |
| | 9.09 | 10.17 | 10.92 | 8.47 | 10.09 |
| 60≤age≤64 | 0 | 248 | 75 | 55 | 378 |
| | 0.00 | 7.91 | 8.11 | 8.96 | 8.07 |
| Total | 11 | 3,136 | 925 | 614 | 4,686 |
| | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

**Notes:**
**For education categories:**
1: Individuals with no education
2: Individuals with general education (grades 1 through 13)
3: Individuals with further education (grades 14 through 20)
4: Individuals with higher education (university education)

## Table 3

**African/white log wage distributions in South Africa**

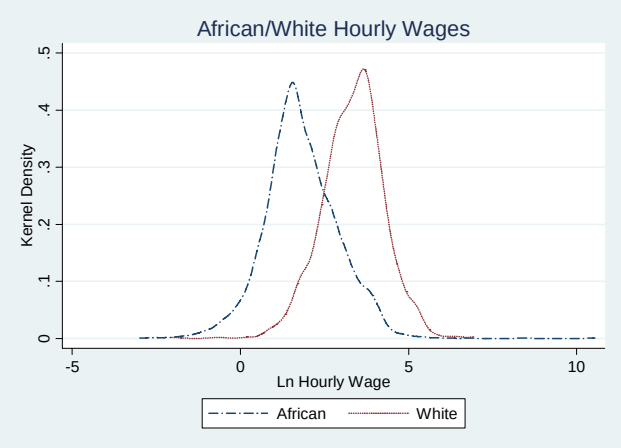
| | **Mean** | **Variance** | | | **Wage gap by percentiles** | | |
|---|---|---|---|---|---|---|---|
| | | Total | $xb$ | Residual | 90-10 | 50-10 | 90-50 |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| **1**. African | 1.9420 | 1.1404 | 0.2254 | 0.9150 | 0.3908 | 0.2286 | 0.1622 |
| **2**. African with white $b's$ | 3.2965 | 2.0080 | 1.0930 | 0.9150 | 0.6625 | 0.4245 | 0.2380 |
| **3**. African with white $b$'s and $x$'s | 3.4351 | 1.9058 | 0.1654 | 1.7404 | 1.0816 | 0.3202 | 0.7614 |
| **4**. White | 3.4351 | 0.9105 | 0.1654 | 0.7451 | 1.1038 | 0.4810 | 0.6228 |
| **5**.African-white difference (**1-4**) | -1.4931 | 0.2299 | 0.0600 | 0.1699 | -0.7130 | -0.2524 | -0.4606 |
| *Effect of:* | | | | | | | |
| **6**. $b$ (**1**–**2**) | -1.3545 | -0.8676 | -0.8676 | -- | -0.2717 | -0.1959 | -0.0758 |
| **7**. $x$ (**2-3**) | -0.1386 | 0.1022 | 0.9276 | -0.8254 | -0.4191 | 0.1043 | -0.5234 |
| **8**. Residual **(3–4)** | -- | 0.9953 | -- | 0.9953 | -0.0222 | -0.1608 | 0.1386 |

*Notes:*

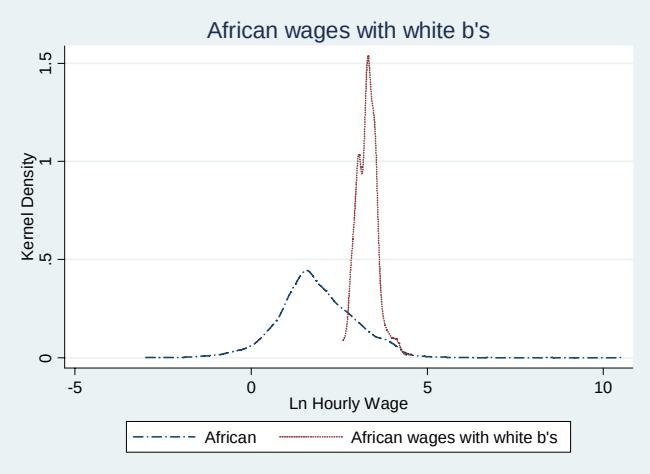
1. (Row 1, column 1): African wages with African *x's* and *b's,* $\ln \overline{w}_A$ (equation 3)
2. (Row 2, column 1): African counterfactual wage with African *x's* but white *b's,* $\ln \overline{w}_A^a$ (equation 5)
3. (Row 3, column 1): African counterfactual wage with white *x's* and *b's,* $\ln \overline{w}_A^b$ (equation 7)
4. (Row 4, column 1): White wages with white *x's* and *b's,* $\ln \overline{w}_w$ (equation 3)

1. (Row 1, column 2): Variance of African wages, $V_A$ (equation 4)
2. (Row 2, column 2): Variance of African counterfactual wage, $V_A^a$ (equation 6)
3. (Row 3, column 2): Variance of African counterfactual wage, $V_A^b$ (equation 8)
4. (Row 4, column 2): Variance of White wages, $V_w$ (equation 4)

## Figure 1




## Figure 2




## Figure 3

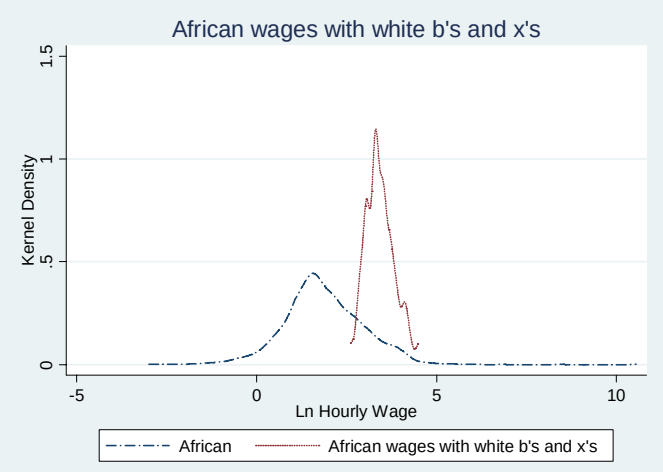


**Table 4**

**A. Variance of wages:**

| | |
|---|---|
| Total Change: | 0.2299 |
| Effect of: | |
| *b* | -0.8676 |
| *x* | 0.7899 |
| *p* | 0.3076 |
| unexplained change: | 0.0000 |

**B. Variance of wage residuals:**

| | |
|---|---|
| Total change: | 0.3076 |
| Effect of: | |
| *x* | -0.0765 |
| *p* | 0.3841 |
| unexplained change: | 0.0000 |