Race and the Incidence of Unemployment in South Africa*

Geeta Gandhi Kingdon
John Knight

RRH: RACE AND UNEMPLOYMENT IN SOUTH AFRICA

LRH: Geeta Kingdon and John Knight

Abstract

South Africa's unemployment rate is one of the highest in the world, and it has important distributional implications. The paper examines both entry into and duration of unemployment using data for the mid-1990s. A probit model of unemployment shows an important role for race, education, age, gender, home-ownership, location, and numerous other variables, all of which have plausible explanations. The large race gap in unemployment is explored further by means of a decomposition analysis akin to that normally used to analyze wage discrimination. There remains a substantial residual which might represent unobserved characteristics, such as quality of education, or discrimination.

*Kingdon, Knight: University of Oxford, Manor Road, Oxford, OX1 3UQ, U.K. Tel: 1865-271065, Fax: 271094. This paper has benefited from comments made by participants at the Development Policy Research Unit conference, South Africa, November 2001, and from the comments of seminar participants at the Centre for the Study of African Economies, University of Oxford. The research was supported by a DFID grant.

JEL Classification Numbers: J64, J71

Address of Contact Author: Geeta Kingdon, Department of Economics, University of Oxford, Manor Road, OX1 3UQ, U.K. Tel: 44-1865-271065; Fax: 44-1865-271094; email: geeta.kingdon@economics.ox.ac.uk

Race and the Incidence of Unemployment in South Africa

1. Introduction

The unemployment rate in South Africa is one of the highest in the world - 36% in 1999 by the broad definition. Even according to the conventional (narrow) definition, which applies a job-search test, one in every four adults who wanted work and actively looked for it was unemployed. Moreover, the unemployment rates for different groups reveal great disparity in the incidence of unemployment by race, gender, age, education, and region. Given the importance of employment income in total household income in South Africa (Bhorat *et al*., 2001), the varying incidence of unemployment across different groups has important implications for the distribution of income and for the incidence of poverty.

In this paper we firstly paint a picture of the distribution of unemployment in South Africa, asking the question 'who are the unemployed?' and identifying the characteristics that make a person more likely to be unemployed. This is done by means of both descriptive statistics and the estimation of a probit equation of unemployment. The probit permits us to measure the influence of a given factor or characteristic on unemployment probability holding other factors constant. It is particularly of interest to examine whether potentially policy-amenable variables such as education and location affect the probability of unemployment in different ways for the different races.

Secondly, we focus on the racial distribution of unemployment, exploring the extent to which the race gap in the probability of employment is due to the black group's inferior employment-enhancing characteristics and to employment discrimination in the labor market. While there is much research in South Africa investigating racial wage discrimination (Knight and McGrath, 1977 and 1987; Rospabe, 1997; Jensen, 1999; Moll, 2000; Erichsen and Wakeford, 2001; Allanson *et al*., 2001), a fuller picture of how the different races fare in the labor market needs to take account of employment discrimination as well, especially since access to employment is a strong predictor of income. Knight and McGrath (1977, 1987) showed that job discrimination was an important source of racial wage differences, but no estimates have as yet been made of the extent of employment discrimination.

2. Data

We use two cross-section datasets: one being the October Household Survey of 1994, carried out by the Central Statistical Office, now known as Statistics South Africa (or simply StatsSA), and the other derived from an integrated household survey carried out in 1993 by the South African Labor Research Unit (SALDRU). The October Household Survey of 1994 (OHS94) is a nationally representative household survey covering 33,000 households across 1,010 clusters in 266 districts. Sampling information is available from CSS(1995). The SALDRU survey is a nationally representative household survey patterned on the World Bank's Living Standards Measurement Study surveys. It yielded a dataset covering about 9000 households across 360 clusters. A detailed account of the sampling procedure is contained in SALDRU (1994).

Some important aspects of employment and unemployment have not been captured in the two surveys. For example, no information is available on the duration of unemployment in the SALDRU survey, and only limited information exists on the employment histories of those unemployed persons who actively looked for work in the week before the survey date. Although there is a question on duration of unemployment in the OHS94 questionnaire, this question is asked only of persons currently unemployed, so that completed duration of unemployment is not known for anyone.

Both datasets are cross-sections rather than panels. This limitation restricts the analysis to obtaining a snapshot of a person's unemployment situation at a point in time rather than over a period of time. This is unfortunate since a number of important questions about South African unemployment cannot be investigated, such as: Is high unemployment due mainly to a high rate of entry into unemployment or to its long duration, i.e. a low rate of exit from unemployment? Is employment probability duration-dependent? What is the completed duration of unemployment? Do probabilities of transition into employment from out-of-labor-force and unemployed states differ significantly? On the positive side, however, the strengths of the datasets lie in their nationally representative character and in their large sample size which permits reliable analysis at a high level of disaggregation.

3. The incidence of unemployment

Two definitions of unemployment are commonly utilized - the broad and the narrow. The narrowly defined unemployed are those who are currently not employed but who looked for work in the week (SALDRU) or four weeks (OHS94) prior to the survey visit. The broadly defined unemployed are the narrow unemployed plus those who say they want work but did not look for work in the past week (past four weeks).

As a number of recent studies have investigated the *extent* of unemployment in South Africa (Klasen and Woolard, 1999; StatsSA, 1998), only a short discussion of the major findings suffices here as a backdrop for our further analysis. Table 1 shows that the broadly measured unemployment rate in South Africa has risen from the already high figure of 31.5% in 1993 to 41.5% in 2001. Even the narrowly measured unemployment rate in the OHSs rose from 20% in 1994 to 29.5% over the seven-year period to 2001. These rates are extremely high by international standards (ILO, 2001). The great broad-narrow discrepancy in unemployment rates indicates that a large proportion of jobless persons who say they want work are not actively looking for work. Some analysts argue that many such persons are not labor force participants[1] but others have persuasively argued that the broad definition is the more relevant because tests suggest that non-searching persons are 'discouraged' workers in South Africa (Kingdon and

Knight, 2000; Poswell, 2002, p.6). The broad concept of unemployment is therefore the one that we analyze in this paper.

Table 2 shows the distribution of unemployment across groups. Unemployment varies dramatically by race: Africans face unemployment rates of 41% but the rate for whites is only 6%. Unemployment decreases monotonically by age, ranging from 51% for the youngest group to 17% for the eldest group. The incidence of unemployment also varies importantly by region, gender, and education. For example, people with higher education face an unemployment rate of 6% but those with primary education or less suffer a rate close to 40%. This pattern is at variance with that commonly observed in developing countries where graduate unemployment is prevalent. Women experience substantially higher unemployment than men. Rural unemployment rates are higher than urban rates, in contrast to the pattern in most developing countries. This is due to the segregation policies of the apartheid era which consigned millions of Africans to live in 'homelands' - predominantly rural areas of poor land quality and little employment opportunity. These homelands effectively became labor reserves from which permanent and even temporary movement to non-homeland areas was impeded by legislative and administrative rules. Thus, high unemployment in much of rural South Africa took the form of waiting in the homelands for a formal sector job opportunity to arise outside.

Nickell (1980) suggests that unemployment incidence should be separated into two components: the chances of entering unemployment and the duration for which individuals remain unemployed. His argument is that these two components are determined in different ways and may be affected by different factors. We adopt this strategy for South Africa.

<u>Entry to Unemployment</u>

Firstly, consider entry into unemployment. Table 3 shows that entry into unemployment in South Africa is mainly dominated by those who have never previously held a job, i.e. who enter unemployment when they enter the labor force, rather than by persons who held a job and then became unemployed. Only 38% of all unemployed persons entered unemployment from the employed state. The fact that a majority of the unemployed have never held a job before is one of the most striking features of unemployment in South Africa. It is worth examining this issue more carefully.

The incidence of direct entry into unemployment (without an intervening period of work) varies by several factors. For example, it differs substantially by race. African unemployed persons are more than twice as likely as whites never to have had work. While this could be partly due to the inferior employment-enhancing characteristics of Africans *vis a vis* whites, it could also be partly due to racial discrimination in employers' hiring practices. Rural dwellers are more likely than urban dwellers never to have worked, possibly because there are fewer job-opportunities in rural than urban areas and because the intensity of job-search is lower in the countryside owing to remoteness from centers of employment. Unemployed women are more likely than unemployed men never to have experienced a period in work. This may be due to women's lesser flexibility in terms of hours of work and the distance they are prepared to travel, or to their higher reservation wages, *ceteris paribus*, than men.

Age is an obvious important factor since the young are more likely to search rather than get 'locked-in' to an undesirable job. The young are also more able to afford unemployed job-search because they have fewer financial commitments than do older persons. Moreover, they may be more ignorant about what their skills can command in the labor market, i.e. may have higher reservation wages. It is more difficult to explain this phenomenon among the older age groups. For example, about 50% of all unemployed persons (61% of unemployed women and 37% of unemployed men) aged 36-64 had never worked before. This is either due to late entry into the labor force - a possible explanation for women who might delay entry until after child-bearing/rearing years - or to extraordinarily long duration of unemployment, which can be explained only by the lack of adequate jobs in the economy or by too narrow a concept of 'employment'. Whereas the *current* labor market status of individuals - whether they are regarded as unemployed or not - is carefully determined in the October Household Surveys through a series of comprehensive questions (Bhorat, 1999)[2], the simple single question on labor market history - i.e. whether the individual ever worked gainfully in the past - relied on the judgment of respondents some of whom might have been thinking of employment only as regular wage employment.

The validity of these explanations is confirmed in a multivariate context. Taking the sample of all unemployed persons, we fitted a probit of 'ever worked before' or 'entry into unemployment from employment' in the first three columns of Table 4. This shows that, holding other factors constant, Africans have a 35 percentage point lower probability than whites of having ever worked before for pay, profit, or family gain. Since the white probability of previous work is 70% (Table 3), the African probability is exactly half that of whites, after standardizing for observed characteristics. The monotonic effect and significance of age is confirmed, as is the effect of gender. The probability of having ever worked gainfully varies importantly by whether the individual is a household head and married; this is as might be expected *a priori*. Homeland residence is associated with a 16-19 percentage point lower probability of previous work compared with non-homeland residence.

The chances of entry into unemployment from the employed state can be separated into voluntary and involuntary entry. The last column of Table 3 shows that, on average, less than a quarter of those unemployed persons who previously worked had quit work 'voluntarily' rather than because of sacking, retrenchment, illness, or end of temporary job[3]. The fact that most people quit work involuntarily probably reflects low vacancy rates and high unemployment rates. However, there is considerable variation by group in Table 3. For example, groups that are more likely to quit work voluntarily are the young, the highly educated, women, and whites.

The cost of voluntary quitting into unemployment is a function of the cost of being unemployed, which depends on the income in and out of work and on the level of one's financial commitments. Thus, for example, the young would have lower income-loss from voluntarily quitting into unemployment insofar as they are often supported by their families while unemployed. The benefits of voluntary quitting into unemployment depend on the prospect for alternative wage and job opportunities. Since the alternative wage opportunities are relatively better the lower the level of firm-specific human capital which the individual possesses, younger workers would be more likely to expect that there are firms willing to pay starting wages similar to their current earnings. Workers liable to be sacked or made redundant include those whose productivity is on the wane but whose wages have not been commensurately reduced. Thus, we expect the old to have a low incidence of voluntary quits, and this is what we observe in the final column of Table 3.

If there is scarcity of educated labor or racial discrimination by employers, more educated people, or persons belonging to the favored racial group, will be more likely to quit voluntarily in search of better wage opportunities because their probability of re-employment is higher. This could explain why persons with higher education and whites have a higher propensity to quit voluntarily. Women are more than twice as likely as men to quit voluntarily into unemployment. Working women may quit voluntarily for child-bearing and -rearing and, being usually the secondary income earners in the family, are also more likely than men to give up their work in case of family emergencies or migration of spouse.

The last three columns of Table 4 present a binary probit of voluntary entry into unemployment from employment. It confirms the effect of age apparent in the descriptive statistics in Table 3. Persons aged 56-64 are significantly less likely to enter unemployment voluntarily than the young. However, while education appears to have a clear positive relationship with voluntary entry in Table 3, it has no significant effect in Table 4 once other factors are controlled. The effect on gender is strongly confirmed in the multivariate context. That rural non-homeland residence is associated with a 13 percentage point greater probability of voluntary entry into unemployment than rural homeland residence may be explained by the fact that wages are by far the lowest in rural non-homeland areas. Finally, race is very important, with Africans, coloreds, and Indians having respectively a 22, 17, and 12 percentage point lower probability of voluntary entry into unemployment than otherwise equivalent whites.

Duration of Unemployment

Now consider the duration of unemployment - the second aspect of unemployment incidence. The length of time for which an individual remains unemployed depends both on the rate at which he receives job offers and on the extent to which these offers are accepted (for instance, Nickell, 1980). It is clear that most identifiable variables have an impact on both the demand and supply sides of the labor market. For example, for certain types of job, e.g. manual work, younger people may be more likely to receive job offers than older people if they are seen as physically more capable. Younger workers may also be more likely to accept job offers insofar as they are more flexible and have less stringent ideas about what is suitable employment.

The question from which we have obtained data on duration of unemployment was ‘How long has (name) been seeking work?’. We interpret this to mean ‘how long has name been *wanting* work?’ rather than ‘how long has name been *searching for* work?’. This seems reasonable because the question was asked of all unemployed persons and not only of those who had taken active steps to find work. The available information is from persons who are currently unemployed, so that it represents *uncompleted* duration of an individual’s spell of unemployment.

The answers were recorded in categorized form (e.g. less than a month, 1-2 months, 2-6 months, 6-12 months, >12 months) rather than as a continuous variable. However, by assigning midpoints of the category, we have created a continuous variable, ‘number of months’ for the duration of unemployment. It is unfortunate that the last category specified in the OHS94 duration question was ‘>12 months’ since it turns out that more than 67% of the unemployed were in this category and it seems possible that many of these suffered unemployment for much longer periods of time than a year - in other words, there is a great loss of information and of variability in the duration variable because of this truncation. However, the October Household Survey of 1997 (OHS97) includes more detailed information, in particular duration categories ‘1 to 3 years’ and ‘greater than 3 years’. In OHS97, of all unemployed persons whose duration was a year or greater, 43.7% had duration between 1 and 3 years and 56.3% of greater than 3 years. We ascribe these proportions to persons in the OHS94 category ‘greater than one year’ in the last row of Table 5. For the other rows, the corresponding proportions for the various groups are applied. The resultant measure of months of unemployment (column 2) shows very long average uncompleted duration: 27 months. Duration is seen to rise monotonically with age and to fall monotonically with education.

The duration of unemployment varies substantially by group in Table 5. It is considerably lower for the younger, well-educated, and white groups than for their opposite numbers. This is consistent with the notion that, on the demand side of the labor market, groups that are deemed by employers to be more productive (younger and more educated) or more desirable (whites) will receive more job-offers. The younger group might also be more likely to accept the offers because of their greater adaptability than older people. Similarly, more educated people may have lower reservation wages (relative to their expected wages) than less well educated people because they are more knowledgeable about the worth of their skills. Thus, some of these groups are also more likely to accept job-offers than their opposite numbers and are likely to quit unemployment sooner.

4. Incidence of unemployment in a multivariate context

We wish to investigate the factors that affect the incidence of unemployment, i.e. to identify the characteristics of individuals that make them more likely to be unemployed. Tables 1 to 4 presented the racial, gender, regional, educational, and age distribution of unemployment in South Africa and showed certain patterns in the incidence of unemployment. However, analysis in a multivariate framework is required in order to isolate the effect of each variable holding the others constant. We utilize a standard discrete choice framework to model the states 'unemployed' and 'employed' as a binary probit.

While both supply-side (worker-related) and demand-side (employer-related) factors are responsible for an individual's labor market state (employed or unemployed), it is not possible in our model to distinguish between job-rationing reasons and worker preferences. For example, if being more educated increases the likelihood of being employed, this could be either because employers ration jobs by educational level of applicants in a labor surplus economy or because individuals who are more educated have more realistic reservation wages relative to their expected wages. The model is largely unable to distinguish between the constraints on and preferences for employment since their effects are not readily separable, though the inclusion of variables whose interpretation is unambiguous should help.

The only non-worker-related characteristics used in the model are (i) variables representing cost of job search, namely the condition of roads in the community; (ii) a proxy for the economic development of the community, capturing local employment opportunities and the local demand for labor; and (iii) a set of region dummies which aims to capture the effects of regional economic differences.

As unemployment is catastrophically high among Africans (41% by the broad definition in 1994) and only about half that rate among the next worst-off racial group (coloreds), we focus on the African group, though results for other minority groups are also presented, i.e. for the so-called coloreds, Asians, and whites. The sample contains only labor force participants and the base or reference category is the employed. Table 6 sets out the results of the pooled binary probit of unemployment for all races using the OHS94 data. Table 7 presents the binary probits of unemployment separately for African, colored, Indian and white persons. A likelihood ratio test of whether it is appropriate to pool the separate races into a single equation was easily rejected[4].

The effect of race on the probability of unemployment confirms the patterns noted earlier. The pooled model with the race dummy variables shows that, even after controlling for locational, demographic, and measured human capital characteristics such as age and education, Africans are 21, coloreds 15, and Indians 11 percentage points more likely to be unemployed than whites. The corresponding figures from a probit of unemployment fitted with SALDRU 1993 data (presented in Appendix 1) are 25, 20, and 15 percentage points respectively. It is possible that this difference reflects a reduction in racial discrimination in the employment practices of employers in the South African labor market in a period of rapid political change between 1993 and 1994.

The fact that even after the introduction of a battery of controls, non-whites suffered significantly greater chances of unemployment than whites in 1993-1994 suggests either racial discrimination in employers' hiring practices or prior discrimination in the schooling system whereby blacks suffered poorer quality schooling than whites, or both. Since quality of education received in the past was governed by race – there being four separate racially segregated school systems - we wanted to explore whether inferior quality schooling is responsible for blacks' higher unemployment. Case and Deaton (1999) have shown that variations in schooling outcomes such as school enrolment, years of education and achievement test scores are significantly explained by their proxy for school quality; it is plausible that labor market performance is also improved by school quality. We therefore estimated the probit models again for that subset of labor force participants in the SALDRU dataset for whom cognitive-skill scores were available. Despite doubts about the reliability of the test score data[5], we nevertheless experimented with including test scores as proxies for the quality of schooling. The presence of test scores (literacy, numeracy, or both together) made no significant difference to the estimated coefficients of the race dummy variables. On the available, weak, evidence we cannot conclude that racial differences in unemployment probability are partly due to racial differences in educational quality, though this is clearly plausible.

The separate probits of unemployment for African, colored, Indian, and white persons are presented in Table 7. It is conspicuous that in the African probit, most variables are statistically significant and the goodness-if-fit, as measured by the pseudo R-square, is better than in the probits for other race groups, particularly the whites. This is partly because there is a greater degree of variation in the dependent variable in the African sample.

The probability of unemployment decreases with age but at a diminishing rate[6]. Incumbents may be protected against competition from (young) entrants by labor market laws or institutions or by firm-specific human capital. Reservation wages may fall with age or with time spent in unemployment. Alternatively, younger people may have a greater chance of entry into unemployment because of their higher degree of job-mobility. There is support for the latter notion in Tables 3 and 4, which show that they are more likely to enter unemployment voluntarily. The higher degree of job-mobility among the young is likely to result from their low levels of firm-specific human capital, their relatively low current costs of unemployment, and their greater ease of finding another job (at least among those who ever held a job before) – as suggested by their lower unemployment duration figure in Table 5.

The incidence of unemployment decreases dramatically with education for all race groups. For example, possessing higher education reduces an African person's predicted probability of unemployment to nil: the marginal effect of the higher education dummy variable is about -39 percentage points whereas from raw data we know that the unemployment rate for Africans with no education (base category) is about 36%. Similarly, among coloreds: the marginal effect of higher education is nearly -16 percentage points and the (unstandardized) unemployment rate for uneducated coloreds is 14%. For Indians and whites with higher education, the predicted probability of unemployment is also nil. Among Africans, education begins to matter to unemployment from the junior secondary level onwards, i.e. Africans with 8-10 years of education have significantly lower chances of being unemployed than those with no education. For coloreds, Indians, and whites, education begins to matter only from the senior secondary level (11-12 years of schooling) onwards.

The role of housing tenure in predicting unemployment has been highlighted by Hughes and McCormick (1987), Nickell (1980) and Oswald (1997). Oswald (1997) reports cross-country and cross-region regressions in which private home ownership percentage is strongly positively associated with unemployment. This literature attributes increases in unemployment in certain OECD countries to the increased rate of home-ownership in these countries, the reason being that home-ownership (and council-housing) makes people immobile by increasing the cost of mobility. It is arguable that home-ownership can exert two opposing sorts of influence on the probability of unemployment. It may exert a *positive* effect either because homeowners are less mobile or because home-ownership may proxy household wealth and wealthier people may have higher reservation wages. Home-ownership may exert a *negative* effect if it is endogenous to unemployment (i.e. if unemployment determines the chances of owning your own home). Table 7 shows that for Africans the former influence dominates: home ownership increases the chances of unemployment by 5.4 percentage points. However, for Indians and whites the latter effect is more relevant, their home-ownership being associated with a lower probability of unemployment (the marginal effects being about -6 and -2 percentage points respectively). The district home-ownership rate has a large positive effect on the chances of unemployment for Africans only.

The number of dependants in the household (Numdep) could either increase unemployment probability - because of greater child-care responsibilities, particularly for women by making them less flexible labor force participants - or it could decrease unemployment probability - because of greater economic need and the consequent lower reservation wage. Thus, one cannot predict the sign of this variable *a priori*. For all four races, the child-care reason seems to dominate, making people significantly more likely to be unemployed. Gender-differentiated probits (not reported) show that the positive effect of Numdep on the probability of unemployment is about twice as strong for women as for men.

We had included certain household composition variables (such as marriage and headship status) in earlier versions of the unemployment probits. Both very significantly reduced the chances of unemployment in each race group. This is consistent with the notion of economic responsibility falling more heavily on household heads and married members. The negative effect of headship and marriage on the probability of unemployment may arise if these are taken as signaling qualities (say, greater maturity or trustworthiness) that employers use to ration jobs. Another explanation is that married and head persons' greater economic responsibility within the household means that they have lower (more realistic) reservation wages. However, our preferred specification excludes these variables on account of their strongly endogenous nature in an unemployment probit: people who are unemployed have lower chances of marrying and becoming heads of their own households.

Controlling for (former) homeland residence, the probability of unemployment is 16 percentage points higher for urban than rural Africans. Among coloreds and Indians, the unemployment chances of urban persons are 21 percentage points and 8 percentage points higher, respectively, than of their rural counterparts. This may be because urban-based job-search is considered more effective than rural-based job-search, as hypothesized in probabilistic models of labor migration. Among whites urban/rural residence has no significant impact.

Residence in a homeland still entails a substantially greater risk of unemployment than residence elsewhere. A black worker living in a homeland is about 21 percentage points more likely to be unemployed than a black worker living in a non-homeland region. This indicates that despite the considerable loosening of apartheid segregation laws by 1994, the former homeland regions still continued to act as labor reserves whose residents were at a great disadvantage in the labor market. Province dummies are included to see whether unemployment incidence varies substantially regionally. The base category is the major metropolitan area, based on Johannesburg (PWV, now known as Gauteng). Black workers in all provinces except northern and eastern Cape are significantly less likely to be unemployed than those in Gauteng, i.e. Gauteng acts as a magnet attracting black migrants looking for work.

In the OHS94 there is information available on distance to the nearest telephone. This is used as a proxy for the remoteness of the community. It is likely to capture aspects of the cost of job-search: we expect a positive sign on this variable[7]. Tables 6 and 7 show that this measure of remoteness has a highly significant positive effect on the chances of unemployment. The more remote the community, the higher the cost of job-search and, accordingly, the higher the probability of unemployment. In the African unemployment probit using SALDRU data, shown in Appendix 1, living in a cluster with impassable roads increases the chances of unemployment significantly. This too is consistent with the notion of the cost of job-search being higher in remote clusters[8].

To sum up, the results on the human capital variables, education and experience, are consistent with the market for more skilled workers being tighter than that for the unskilled. Wage setting may clear the skilled labor market but wage floors may prevent the unskilled labor market from clearing. There may also be elements of internal labor markets, in which experienced incumbents are protected against competition from labor market entrants. The importance of residence suggests that workers in remote locations face high search costs; their disadvantage would be exacerbated if they became discouraged from searching. The female disadvantage in unemployment is associated with the inflexibility imposed by conventional gender roles. The importance of race in determining unemployment, *ceteris paribus*, deserves further investigation, to which we now turn.

5. Decomposition of the race gap in unemployment probability

The broad unemployment rate among Africans (41%) and whites (6%) in the OHS94 data indicates that the raw African-white race gap in unemployment rate is 35 percentage points. After standardising for observed characteristics in the pooled unemployment probit of Table 6, however, this race gap is reduced to 21 percentage points. In other words, 14 percentage points out of the 35 percentage point gap is explained by the African-white difference in observed characteristics. Thus 40% of the racial gap in the probability of unemployment is attributable to differences in measured characteristics. The unexplained residual (60%) is due to racial discrimination or to differences in the unobserved traits of blacks and whites, or to a combination of both. However, this method of inferring the extent of the unexplained gap in unemployment probability is unsatisfactory because of its restrictive assumption that the probit of unemployment is identical for blacks and whites in all respects except the intercept.

A more satisfactory method is to allow for the possibility that the coefficients of the variables differ by race and then to apply the familiar Oaxaca (1973) method of decomposing the raw race-gap in unemployment probability into explained and unexplained components[9]. The Oaxaca methodology makes use of the fact that in linear regression, the fitted regression line passes through the mean of the variables, so that the dot product of the vector of coefficients and the vector of mean variable values gives the mean of the dependent variable. One feature of probit analysis is that, unlike OLS, the actual mean of the dependent variable and the predicted mean in a regression equation need not be the same. However, they are usually close together and the Oaxaca method can, therefore, be extended to the case of discrete choice models[10]. This adaptation of the Oaxaca technique for discrete choice models has recently been used to study UK

Black-White unemployment gaps (Blackaby. *al*., 1998, 1999), UK male-female unemployment gaps (Gomulka and Stern, 1990), and Zambian male-female formal sector employment gaps (Nielsen, 1998).

However, as with all applications using the Oaxaca (1973) method, this decomposition suffers from the familiar index number problem, namely that the estimated sizes of explained and residual components will depend on which unemployment structure - African or white probit coefficient vector - is used as the non-discriminatory structure. Oaxaca and Ransom (1994) suggest a way to circumvent the index number problem, whereby the pooled coefficient vector - pooled for the two races being compared - is taken as the non-discriminatory unemployment structure.

Since the white racial group has been the most advantaged in South Africa, with the lowest unemployment rate, we are interested in decomposing the gap in unemployment probability between each non-white race group (African, colored, Indian) and whites. That is, we will make binary comparisons between each race group and whites at any one time. The method can be illustrated by focusing on the African-white comparison.

Let I_i^{*j} be a latent variable for the *i*th individual in the *j*th race group, where

$$I_i^{*j} = \alpha X_i + u_i \tag{1}$$

where j = (African, coloured, Indian, white), X_i is a vector of variables that determine unemployment, α is an associated vector of coefficients, and u_i is an error term distributed N(0,1). Suppose that a binary indicator variable indicating unemployment status is given by

$$\begin{aligned} I_i^{j}=1 &\quad \text{if } I_i^{*j}>=0 \text{ , i.e., the individual is unemployed, and} \\ I_i^{j}=0 &\quad \text{if } I_i^{*j}<0 \text{ , i.e., the individual is employed.} \end{aligned} \tag{2}$$

Denote the probability of observing I_i^j=1 by P(I_i^j =1). This probability of observing unemployment is given by the cumulative normal distribution P(α , X_i) and it can be estimated using a simple binary probit model. Thus, the average of predicted probabilities of unemployment for Africans will be $\overline{P}(\alpha_a, X_a)$ and the average of predicted probabilities of unemployment for whites will be $\overline{P}(\alpha_w, X_w)$, where the subscripts a and w denote African and white respectively.

Define α^* as the unemployment structure that would prevail in the absence of racial differences in the return to unemployment-generating characteristics. Deviations from the race-neutral unemployment structure (represented by α^*) could arise from either discrimination or other unexplained sources of racial differences. On the assumption that a probit estimate for the pooled sample represents the determinants of unemployment in the absence of discrimination or unobserved racial differences, the difference between the average unemployment probability of Africans and what their average unemployment probability would be without discrimination or unobserved influences in returns is:

$$\overline{P}(\alpha_a, X_a) - \overline{P}(\alpha^*, X_a) \tag{3}$$

The comparable expression for whites is:

$$\overline{P}(\alpha^*, X_w) - \overline{P}(\alpha_w, X_w) \tag{4}$$

Thus, the total race gap in average African and white unemployment probability can be written as

$$T = \overline{P}(\alpha_a, X_a) - \overline{P}(\alpha_w, X_w) \tag{5}$$

$$T = \{\overline{P}(\alpha^*, X_a) - \overline{P}(\alpha^*, X_w)\} + \{\overline{P}(\alpha_a, X_a) - \overline{P}(\alpha^*, X_a)\} + \{\overline{P}(\alpha^*, X_w) - \overline{P}(\alpha_w, X_w)\} \tag{6}$$

$$T = \{\text{term 1}\} + \{\text{term 2}\} + \{\text{term 3}\}$$

The first term in (6) uses the race-neutral pooled unemployment structure to predict unemployment probabilities for both Africans and whites but allows the characteristics of Africans and whites to differ. It is the explained part of the total race gap in unemployment probability since it shows the gap in unemployment probability explained by differences in African and white characteristics. The second and third terms together constitute the non-explained part of the total

African-white gap in unemployment probability. The second term shows the difference between returns to African characteristics and those that would exist with a race-neutral unemployment structure. It might be interpreted as the African disadvantage in hiring by employers. The third term shows the difference between returns to white characteristics and those that would exist with a race-neutral unemployment structure. It might be interpreted as the white advantage in hiring by employers.

In what follows, we use the probit equations of unemployment for Africans and whites in Table 7 to obtain α_a and α_w. We also use the pooled African and white unemployment probit (not reported) to obtain the race-neutral unemployment structure α^*. For each individual we produce the predicted probability of unemployment and then calculate the mean of the predicted probabilities summing over observations. Thus, $\overline{P}(\alpha_a, X_a)$ is the average, across the African sample, of the predicted probabilities of unemployment, using African coefficients and African characteristics; $\overline{P}(\alpha_w, X_w)$ is the average of predicted probabilities of unemployment, across the white sample, using white coefficients and white characteristics; $\overline{P}(\alpha^*, X_w)$ is the average of predicted probabilities of unemployment, across the white sample, using the *pooled* coefficients and white characteristics; and so on. Similar computations are made for the comparison between whites and other race groups.

The results of the decomposition exercise are reported in Table 8. Of the total African-white race gap in unemployment probability (33.7 percentage points), 25.4 percentage points is explained and only 8.3 percentage points is not explained by African-white differences in unemployment-generating characteristics. In other words, about 75% of the African-white gap in unemployment is due to superior white characteristics such as education and better location and only 25% of the gap is unexplained. Of the coloured-white unemployment gap, 60% is explained by differences in characteristics and 40% is unexplained. 63% of the Indian-white unemployment gap is explained by differences in characteristics and 37% is unexplained.

The raw unemployment rates by race suggest most discrimination against Africans, followed by coloureds, and least discrimination against Indians. If we attribute the unexplained component to employer discrimination, then Table 8 shows that the probability of unemployment that is due to discrimination is 8.3, 6.5, and 3.2 percentage points for Africans, coloureds and Asians, respectively[11].   The racial hierarchy apparent in raw unemployment data persists.

According to the decomposition results in Table 8, the major part of the reason why Africans have a much higher unemployment rate than whites is their lower levels of employment-enhancing characteristics such as education and their location in areas of high unemployment. While we refer to this as the 'explained' component of the race gap in unemployment, i.e. as the part that cannot be attributed to labor market discrimination, both the lower education and poorer location of Africans are manifestations of pre-labor-market discrimination. The apartheid location policies forcibly confined millions of Africans to the 'homelands' which are very low employment areas. Moreover, there was discrimination in the schooling system, Africans being subjected to poorer access to, and quality of, education.

The non-explained part of the race gap in unemployment probability may arise from employer discrimination but that is not inevitable. It may also (or instead) reflect the lack of control, in our unemployment probit, for expected productivity or productivity-related characteristics such as quality of schooling which employers may observe but which are unmeasured in our (and most other) datasets. Case and Deaton (1999) report that Africans faced very much poorer quality of education in apartheid South Africa than did whites. There is also evidence in South Africa that some firms recruit non-African workers in preference to African workers on the basis of their higher expected productivity. Frijters (1999), in his case study of a large South African clothing firm, found that the firm was significantly more likely to recruit

Indians than Africans, even after controlling for score on a test of applicants' nimbleness. Productivity, as measured by the number of faultless garments produced per unit of time, was lower for the firm's African than for its Indian employees, and it was inferred that Indian workers were favored on the economically rational basis of expected productivity[12]. Other factors which may account for the so-called unexplained residual in Table 8 are traits such as skills, attitudes, trust, and social networks, which employers may gauge at the time of recruitment but which are unmeasured in most datasets.

6. Conclusions

Unemployment is very inequitably distributed in South Africa and certain groups are much more likely to enter it and to stay in it than others. Young uneducated Africans living in homelands and remote areas are most vulnerable to unemployment. There are two particularly striking features of South African unemployment: firstly, the fact that rural unemployment rates are higher than urban rates is atypical among countries and is explained by historical policies restricting mobility. Secondly, the majority (62%) of the unemployed have never held a job before, i.e., they entered unemployment from the time of entering the labor force. The very long duration of unemployment (>1 year) among a high proportion (68%) of the unemployed suggests that the demand-side of the labor market is responsible for a good part of the unemployment.

Our analysis tells us the characteristics of the unfortunate people who are liable to be at the end of the queue for employment. Improving their characteristics may improve their place in the queue, but it will not necessarily reduce unemployment. In the African group - the group that suffers catastrophically high unemployment rates - human capital characteristics such as education and employment experience dramatically reduce the chances of unemployment. A policy prescription that African education and skills should therefore be upgraded may not solve the problem: unless there are more jobs in the economy, upgrading the education of Africans will at best change the composition of employment in their favor. Of course, it is possible that expanding education and skills will reduce overall unemployment. The mechanism might be to increase the supply of skilled labor, for which there is market clearing, and to decrease the supply of unskilled labor, for which the market fails to clear and there is a surplus of workers. However, that is straying beyond the evidence of this paper.

The analysis suggests that racial differences in unemployment incidence cannot simply be dismissed as a problem of the poorer productive characteristics of the African, coloured, and Indian groups relative to the whites in South Africa. While a substantial part of the race gap in the incidence of unemployment in the mid-1990s was explained by inter-group differences in observed characteristics, there remained a residual that could not be explained in this way. The residual may be due to employer discrimination or to racial differences in unmeasured determinants such as the quality of education. Further research incorporating data on the quality of education will be fruitful, and longitudinal data sets will be needed to examine the policy questions concerning unemployment dynamics.

References

Allanson, Paul, Jonathan Atkins, and Timothy Hinks, "Did the End of Apartheid Spell the Beginning of the End for the Racial Wage Hierarchy in South Africa?," *Review of Development Economics* (2002) forthcoming.

Bhorat, Haroon, Murray Leibbrandt, Muzi Maziya, Servaas van der Berg, and Ingrid Woolard, *Fighting Poverty: Labour Markets and Inequality in South Africa*, Cape Town: UCT Press (2001).

Blackaby, David, K. Clark, D. Leslie, Philip Murphy, "Black-white male earnings and employment prospects in the 1970s and 1980s: Evidence for Britain," *Economic Letters* 46 (1994): 273-80.

Blackaby, David, Derek Leslie, Philip Murphy and Nigel O'Leary "The Ethnic Wage Gap and Employment Differentials in the 1990s: Evidence for Britain" *Economics Letters* 58 (1998): 97-103.

Blackaby, David, Derek Leslie, Philip Murphy, and Nigel O'Leary "Unemployment among Britain's Ethnic Minorities," *Manchester School* 67 (1999): 1-20.

Case, Anne and Angus Deaton, "School Inputs and Educational Outcomes in South Africa," *Quarterly Journal of Economics*; 114 (1999): 1047-84.

CSS, "October Household Survey 1994", Statistical Release PO317, Central Statistical Service, Pretoria, March (1995).

Darity, William Jr., *Economics and Discrimination, Volume II*, Aldershot: Edward Elgar (1995).

Erichsen, G. and Jeremy Wakeford, "Racial Wage Discrimination in South Africa: Before and After the First Democratic Election", DPRU Working Paper 01/49, University of Cape Town (2001).

Frijters, Paul, "Hiring on the Basis of Expected Productivity in a South African Clothing Firm," *Oxford Economic Papers* 51 (1999): 345-54.

Gomulka, Joanna and Nicholas Stern, "The Employment of Married Women in the United Kingdom 1970-1983," *Economica* 57 (1990): 171-99.

Hughes, Gordon and Barry McCormick, "Housing Markets, Unemployment and Labour Market Flexibility in the UK," *European Economic Review* 31 (1987): 615-41.

Jensen, Robert, "An Early Assessment of Racial Wage Differentials in Post-Apartheid South Africa", manuscript, Harvard University (1999).

Kingdon, Geeta and John Knight, "Are Searching and Non-searching Unemployment Distinct States When Unemployment is High? The Case of South Africa, WPS/2000.2, Centre for the Study of African Economies, University of Oxford (2000).

Klasen, Stephan and Ingrid Woolard, "Levels, Trends, and Consistency of Employment and Unemployment Figures in South Africa," *Development Southern Africa* 16 (1999): 3-36.

Knight, John B. and Michael D. McGrath, "An Analysis of Racial Wage Discrimination in South Africa," *Oxford Bulletin of Economics and Statistics* 39 (1977): 245-71.

-------- "The Erosion of Apartheid in the South African Labour Market: Measures and Mechanisms", Applied Economics Discussion Paper 35, Institute of Economics and Statistics, University of Oxford (1987).

ILO, *Restructuring the Labour Market: The South African Challenge: An ILO Country Review*, International Labour Organisation, Geneva (1996).

ILO, *Yearbook of Labour Statistics*, International Labour Organisation, Geneva (2001).

Moll, Peter, "Discrimination is Declining in South Africa but Inequality is Not," *Journal of Studies in Economics and Econometrics*, November (2000).

Nickell, Stephen, "A Picture of Male Unemployment in Britain" *Economic Journal*, 90 (1980): 776-94, December.

Oaxaca, Ronald, "Male-Female Differentials in Urban Labor Markets," *International Economic Review*, 3 (1973): 603-709.

Oswald, Andrew, "Thoughts on Nairu: On Homes and the Natural Rate of Unemployment," *Journal of Economic Perspectives*, Correspondence, 11 (1997): 227-28.

Poswell, Laura, "The Post-Apartheid South African Labour Market: A Status Report," Development Policy Research Unit, University of Cape Town, February (2002).

Rospabe, Sandrine, "Making Racial Wage Relations Fair in South Africa: A Focus on the Role of Trade Unions", DPRU Working Paper no 01/48, University of Cape Town, April (2001).

SALDRU, *South Africans Rich and Poor: Baseline Household Statistics*, Project for Statistics on Living Standards and Development, South African Labour and Development Research Unit, Cape Town (1994).

StatsSA, *Unemployment and Employment in South Africa*, Statistics South Africa, Pretoria (1998).

StatsSA, Statistical Release PO317, StatsSA Homepage, http://www.statssa.gov.za/ (2000).

StatsSA, *Labour Force Survey: September 2001*, Statistical Release PO210, Statistics South Africa, Pretoria, March (2002).

Table 1
Unemployment rates in South Africa, 1993-2001

	Source	Broad definition	Narrow definition	Broad-narrow gap
1993	SALDRU	31.2	13.0*	18.2
1994	OHS	31.5	20.0	11.5
1995	OHS	29.2	16.9	12.3
1996	OHS	35.6	21.0	14.6
1997	OHS	37.6	22.9	14.7
1998	OHS	38.6	26.1	12.5
1999	OHS	36.2	23.3	12.9
2000	LFS	35.9	25.8	10.1
2001	LFS	41.5	29.5	12.0

Source: SALDRU data; OHS figures from StatsSA (1998, p3) and StatsSA's webpage (StatsSA, 2000); Labor Force Survey (LFS) figures from Tables B and H of StatsSA (2002).

*The large difference in narrow unemployment rates between SALDRU and both OHS and LFS sources is due to the fact that the SALDRU survey used a reference period (for job-search) of one week whereas the OHS and LFS surveys use one of four weeks.

Table 2

Unemployment rate (%), by age, education, gender, region, and race, OHS94

	Broad definition	Narrow definition	Broad-narrow gap
Age			
16-24	51.4	37.8	13.6
25-35	35.3	23.3	12.0
36-45	25.2	14.3	10.9
46-55	21.3	11.0	10.3
56-64	16.9	8.5	8.4
Education			
None	38.7	20.1	18.6
Primary	42.5	26.8	15.7
Junior	35.3	23.5	11.8
Secondary	28.3	19.5	8.8
Higher	5.7	3.9	1.8
Gender			
Male	26.2	17.3	8.9
Female	40.7	25.3	15.4
Region			
Rural	40.3	23.4	16.9
Urban	27.9	19.1	8.8
Race			
African	41.2	26.2	15.0
Colored	23.3	19.4	3.9
Indian	17.1	14.3	2.8
White	6.3	4.2	2.1

Table 3
Entry into unemployment by age, education, gender, region, and race, OHS94

	All unemployed (N)	Never worked before (%) (a)	Worked before (%) (b)	Of those who worked before, proportion who entered unemployment voluntarily (%)
Age				
16-24	4128	82.8	17.2	25.4
25-35	5245	64.6	35.4	26.4
36-45	2646	52.8	47.2	24.1
46-55	1244	47.4	52.6	22.4
56-64	338	39.8	60.2	13.3
Education				
None	1265	63.3	36.7	23.1
Primary	4507	63.6	36.4	23.5
Junior	4476	60.5	39.5	23.1
Secondary	3056	74.9	25.1	29.1
Higher	297	57.3	42.7	38.7
Gender				
Male	5572	58.9	41.1	15.8
Female	8029	69.9	30.9	34.2
Region				
Rural	5642	72.3	27.7	24.8
Urban	7959	58.4	41.6	24.4
Race				
African	10130	68.4	31.6	22.6
Colored	2236	43.7	56.3	22.9
Indian	609	46.6	53.4	27.9
White	626	30.3	69.7	49.3
Total	13601	61.8	38.2	24.6

The unemployed are divided into those who have never worked before and those who have worked before, i.e., a + b = 100%.

Table 4
Binary probits of entry into unemployment, OHS94

	Probit of entry into unemployment from employment				Probit of voluntary entry into unemployment, among all unemployed who worked before			
	Coefficient	**Marginal effect**	**Robust t-value**		**Coefficient**	**Marginal effect**	**Robust t-value**	
Age 25-35*	0.7087	0.267	16.14	***	0.0517	0.017	0.76	
Age 36-45*	0.9158	0.352	16.61	***	-0.0284	-0.009	-0.4	
Age 46-55*	1.0202	0.389	16.29	***	-0.0962	-0.030	-1.07	
Age 56-64*	1.0870	0.408	12.19	***	-0.5306	-0.139	-3.75	***
Male*	0.3007	0.113	8.31	***	-0.6091	-0.191	-13.52	***
Household head*	0.3278	0.126	7.73	***	0.0174	0.006	0.36	
Married*	0.1685	0.064	5.00	***	0.1189	0.038	2.34	***
Numdep	0.0126	0.005	1.42		0.0069	0.002	0.54	
African*	-0.9260	-0.354	-10.56	***	-0.6668	-0.221	-6.22	***
Colored*	-0.3958	-0.140	-3.85	***	-0.6197	-0.174	-6.35	***
Indian*	-0.5349	-0.177	-4.42	***	-0.4283	-0.118	-2.84	***
Urban homeland*	-0.0362	-0.013	-0.29		0.0577	0.019	0.43	
Rural non-homeland*	0.4760	0.186	3.92	***	0.3777	0.131	2.91	***
Urban non-homeland*	0.4297	0.160	4.45	***	0.0969	0.031	0.8	
Numemp1	0.0237	0.009	1.15		0.0201	0.006	0.73	
Primary*	0.1000	0.038	1.73	*	-0.0229	-0.007	-0.28	
Junior*	0.1089	0.041	1.80	*	-0.1004	-0.032	-1.11	
Secondary*	-0.1334	-0.049	-1.96	**	0.0788	0.026	0.8	
Higher*	0.2565	0.099	1.24		0.2405	0.082	1.15	
Vocational training*	-0.3374	-0.117	-1.88	*	-0.1550	-0.047	-0.69	
Lives in owned home*	-0.0342	-0.013	-0.68		0.0954	0.030	1.46	
Wcape*	0.4043	0.157	3.44	***	-0.2747	-0.081	-1.8	
Ncape*	0.2416	0.093	1.49		0.8251	0.304	5.51	***
Ecape*	0.0841	0.032	0.69		-0.0971	-0.030	-0.7	
Natal*	0.2615	0.100	2.44	**	-0.0098	-0.003	-0.07	
Ofs*	0.0088	0.003	0.06		0.2967	0.102	1.82	*
Etvl*	0.2459	0.095	1.68	*	-0.1198	-0.037	-0.67	
Ntvl*	-0.1102	-0.041	-0.62		0.2427	0.083	1.42	
Nw*	0.5047	0.197	3.72	***	0.4040	0.141	2.37	***
Constant	-0.8769		-5.28	***	-0.0768		-0.36	
Log L		-7555.071				-2683.73		
Restricted Log L		-9044.858				-3059.77		
Pseudo R-square		0.1647				0.1229		
N		13601				5195		
Mean of dependent variable*		0.3820				0.2758		

The starred variables are 0/1 variables. Their mean represents the proportion of ones in the sample. Numdep is the number of dependants (aged <16 or >64); Numemp1 is number of employed members in the household. The base category for age is age16-24, for sex female, marital status unmarried, status non-homeland, home not owned, province Gauteng.

Table 5
Duration of unemployment, by age, education, gender, region, and race, OHS94

	Number of unemployed	Duration of unemployment (months)	% distribution of duration of unemployment					
			<1 month	1-2 months	2-6 months	6-12 months	12-36 months	>36 months
Age								
16-24	4128	21.4	7.2	5.4	9.7	20.1	35.8	21.9
25-35	5245	28.5	4.7	3.2	7.7	13.9	28.9	41.6
36-45	2646	30.4	5.2	3.1	6.4	12.7	24.7	48.0
46-55	1244	30.8	6.5	2.1	6.9	11.6	23.3	49.6
55-64	338	32.3	7.4	1.8	5.0	11.2	20.2	54.5
Education								
None	1265	30.0	5.3	3.8	6.7	11.0	27.1	46.1
Primary	4507	29.8	4.8	3.3	6.5	13.1	26.8	45.5
Junior	4476	27.4	5.4	3.4	8.0	14.6	30.2	38.4
Secondary	3056	23.4	6.3	4.1	10.3	19.8	30.9	28.6
Higher	297	16.5	22.2	5.0	8.6	24.0	22.9	17.3
Gender								
Male	5572	27.3	4.7	3.6	8.3	15.1	30.1	38.2
Female	8029	27.2	6.5	3.6	7.5	14.9	29.0	38.5
Region								
Rural	5642	28.2	5.6	3.4	6.5	13.8	30.4	40.2
Urban	7959	26.4	5.7	3.7	9.1	16.1	28.7	36.6
Race								
African	10130	28.3	4.8	3.3	7.2	14.5	29.5	40.7
Colored	2236	21.3	6.2	5.9	12.2	19.2	33.9	22.6
Indian	609	19.7	7.6	5.5	12.5	21.8	33.6	18.9
White	626	14.6	25.9	5.8	13.8	15.4	26.2	12.9
Type of U								
Searching U	7725	26.3	4.5	4.1	9.3	16.7	29.4	36.0
Non-search U	5876	28.1	7.0	3.1	6.2	13.2	30.4	40.2
TOTAL	13601	27.2	5.7	3.6	7.9	15.0	29.6	38.2

Source: October Household Survey, 1994.
The OHS94 survey truncates the duration question at 12 months, *i.e.* the longest duration information code provided is 'greater than 1 year'. Since 67.8% of all unemployed persons had unemployment duration of greater than 1 year, there is a great loss of information on variation of unemployment duration within this large group. For the purposes of computing column 2 'mean duration in months', the mid-points of the categories <1 month, 1-2 months, 2-6 months, 6-12 months, 1-3 years and >3 years are taken as 0.5, 1.5, 4, 9, 24, and 48 months respectively.

Table 6

Unemployment probits, whole sample, OHS94

	coefficient	robust t-value		marginal effect
Age				
Age 21-25	-0.3318	-9.58	***	-0.091
Age 26-35	-0.7792	-23.41	***	-0.209
Age 36-45	-1.0890	-29.59	***	-0.261
Age 46-55	-1.1810	-28.24	***	-0.243
Age 56-64	-1.2997	-21.16	***	-0.224
Education				
Primary	0.0149	0.39		0.005
Junior	-0.0781	-1.68	*	-0.023
Secondary	-0.3200	-6.03	***	-0.091
Higher	-1.0376	-10.67	***	-0.215
Voc diploma	-0.0098	-0.13		-0.003
Other variables				
Ownship	0.0554	1.09		0.017
Numdep	0.0514	8.80	***	0.016
Urban	0.4690	7.70	***	0.134
Male	-0.3578	-13.41	***	-0.109
Race				
African	0.6957	10.95	***	0.206
Colored	0.4468	5.02	***	0.147
Indian	0.3264	4.44	***	0.108
Location				
Homeland	0.5458	7.01	***	0.180
W. cape	-0.4449	-3.85	***	-0.118
N. cape	-0.0200	-0.17		-0.006
E. cape	-0.1010	-1.00		-0.030
Kwazulu natal	-0.3596	-3.63	***	-0.099
Ofs	-0.3117	-2.24	***	-0.084
Mpumalanga	-0.2168	-2.28	***	-0.061
N. province	-0.2571	-2.11	**	-0.071
N.W. province	-0.3331	-3.12	***	-0.089
Community variables				
Disttel	0.0729	5.99	***	0.022
Downship	0.5349	4.21	***	0.161
Constant	-0.7659	-6.86	***	
Ln L		-22330.51		
Restricted Ln L		-28501.08		
Pseudo R^2		0.2165		
N		47667		
Mean of dependent variable		0.285		

The base or reference categories are as follows: age: persons aged 16-20 years old; education: persons with no education; race: whites; and province: PWV (now called Gauteng). *, **, and *** represent statistical significance at the 10%, 5% and 1% levels respectively. Disttel is distance to nearest telephone (a proxy for remoteness); downship is district home ownership rate, *i.e.* the proportion of households in the district that lived in owned homes. Other definitions and omitted categories are the same as for Table 4.

Table 7
Unemployment probit, by race, OHS94

	African				Colored				Indian				White			
	Coeffi-cient	robust t-value		marginal effect	Coeffi-cient	robust t-value		marginal effect	Coeffi-cient	robust t-value		marginal effect	Coeffi-cient	robust t-value		marginal effect
Age																
Age 21-25	-0.2088	-4.44	***	-0.078	-0.4764	-8.79	***	-0.111	-0.4458	-6.07	***	-0.076	-0.6203	-7.63	***	-0.045
Age 26-35	-0.7196	-16.41	***	-0.259	-0.8859	-14.08	***	-0.203	-0.8558	-11.30	***	-0.140	-1.0714	-11.54	***	-0.083
Age 36-45	-1.1258	-22.86	***	-0.368	-1.1094	-16.14	***	-0.222	-0.9574	-8.26	***	-0.151	-1.0628	-12.36	***	-0.082
Age 46-55	-1.2832	-22.89	***	-0.371	-1.1036	-12.06	***	-0.190	-1.1218	-11.16	***	-0.143	-0.9170	-10.34	***	-0.064
Age 56-64	-1.5460	-20.56	***	-0.373	-1.2739	-11.42	***	-0.182	-1.2370	-8.90	***	-0.125	-0.5852	-5.78	***	-0.041
Education																
Primary	-0.0351	-0.82		-0.013	0.1222	1.32		0.033	--	--		--	--	--		--
Junior	-0.1798	-3.41	***	-0.068	0.0113	0.11		0.003	-0.0821	-1.07		-0.016	--	--		--
Secondary	-0.3291	-5.20	***	-0.120	-0.2854	-2.33	***	-0.070	-0.3312	-3.23	***	-0.067	-0.3884	-6.92	***	-0.040
Higher	-1.7038	-10.91	***	-0.393	-0.9411	-3.19	***	-0.159	-0.9552	-4.30	***	-0.124	-0.5728	-6.88	***	-0.051
Vocational diploma	0.1803	1.19		0.070	0.2373	0.80		0.070	0.2151	0.94		0.048	0.1271	1.30		0.014
Other var																
Ownship	0.1433	2.51	***	0.054	0.0057	0.08		0.002	-0.2623	-3.68	***	-0.056	-0.1503	-3.26	***	-0.017
Numdep	0.0550	7.60	***	0.021	0.0351	2.67	***	0.009	0.0462	2.46	***	0.009	0.0917	4.19	***	0.010
Urban	0.4175	5.92	***	0.158	0.9269	5.38	***	0.210	0.4618	4.57	***	0.077	0.0592	0.66		0.006
Male	-0.3891	-13.96	***	-0.148	-0.2597	-3.82	***	-0.070	-0.3989	-5.70	***	-0.087	-0.4650	-9.84	***	-0.052
Location																
Homeland	0.5675	6.56	***	0.215	1.0266	2.78	***	0.367	1.2648	2.89	***	0.417	1.3408	2.87	***	0.333
W. cape	-0.3824	-3.22	***	-0.136	-0.5850	-6.66	***	-0.153	0.0060	0.05		0.001	0.2662	2.69	***	0.033
N. cape	-0.1996	-1.45		-0.074	0.1956	1.73	*	0.056	0.8811	2.33	**	0.267	0.1271	1.33		0.014
E. cape	-0.1239	-1.05		-0.047	-0.2997	-2.52	***	-0.073	-0.0236	-0.19		-0.005	0.3800	4.70	***	0.051
Kwazulu natal	-0.5882	-5.13	***	-0.207	-0.3010	-1.48		-0.071	0.1384	1.46		0.028	0.2902	3.32	***	0.036
Ofs	-0.4918	-3.27	***	-0.171	0.0560	0.25		0.015	0.1652	0.26		0.037	0.1027	0.46		0.012
Mpumalanga	-0.2981	-2.68	***	-0.108	-0.6582	-4.42	***	-0.126	-0.2390	-1.32		-0.042	0.1202	1.22		0.014
N. province	-0.3473	-2.46	***	-0.125	-0.4608	-6.29	***	-0.098	-0.2175	-0.83		-0.039	-0.1120	-1.05		-0.011
N.W. province	-0.4737	-3.60	***	-0.166	-0.0689	-0.37		-0.018	-0.8479	-2.77	***	-0.103	0.1736	1.79	*	0.021
Community var																
Disttel	0.0736	4.41	***	0.028	0.0438	2.02	**	0.012	0.0635	1.58		0.013	0.0984	4.04	***	0.010
Downship	0.4881	3.14	***	0.186	0.1258	0.62		0.034	-0.0046	-0.03		-0.001	0.0453	0.19		0.005
Constant	-0.0888	0.67			-0.3973	-2.48	***		-0.1895	-1.09			-0.3935	-2.35	***	
Log L		-13875.81				-4344.63				-1489.50				-2011.13		
Restricted Log L		-16839.64				-5239.37				-1701.73				-2276.49		
Pseudo R^2		0.1760				0.1708				0.1247				0.1166		
N		24929				9709				3972				9057		
Dependent variable mean		0.406				0.230				0.153				0.069		

Variable descriptions as in Table 5. The base category for education is ‘no education’ for Africans and coloreds, ‘primary or less’ for Indians and ‘Junior or less’ for white persons. The proportion in the base category for education are: 13.8%, 7.5%, 7.5% and 22.0% respectively for Africans, coloreds, Indians, and whites.

Table 8
Decomposition of the race gap in unemployment probability

	African - white	Colored - white	Indian - white
Total race gap in unemployment probability	0.337	0.161	0.084
Part explained by characteristics	0.254	0.096	0.052
Part not explained by Characteristics, i.e. residual	0.083	0.065	0.032
Of which:			
Part due to non-white disadvantage	0.022	0.031	0.022
Part due to white advantage	0.061	0.034	0.010

The total race gaps in unemployment probability here (row 1) differ somewhat from the race gaps implied in the last rows of Table 2. This is because the unemployment rates reported in Table 2 are weighted averages of unemployment across all individuals of a given race whereas the unemployment rates for each race implied here in Table 8 (from which the gaps here are computed) are the average of predicted probabilities of unemployment in probit models for each race. However, the implied race gaps in Tables 2 and those reported in Table 8 are quite close to each other - only about one percentage point apart for African-white and colored-white comparisons and about two percentage points apart for the Indian-white comparison in the two tables.

Appendix Table 1
Unemployment probit (SALDRU 1993 data)

	Pooled coefficient	Pooled robust t value	Pooled marginal effect	Pooled mean	African coefficient	African robust t value	African marginal effect	African mean
Age								
Age 21-25	-0.2713	-4.60 ***	-0.082	0.170	-0.2045	-2.87 ***	-0.075	0.173
Age 26-35	-0.8106	-13.96 ***	-0.233	0.331	-0.7739	-11.10 ***	-0.272	0.339
Age 36-45	-1.1593	-18.12 ***	-0.292	0.247	-1.1342	-14.71 ***	-0.359	0.241
Age 46-55	-1.2499	-18.78 ***	-0.272	0.139	-1.3106	-17.60 ***	-0.363	0.130
Age 56-64	-1.4605	-15.87 ***	-0.260	0.052	-1.4677	-13.95 ***	-0.355	0.054
Education								
Primary	-0.0044	-0.09	-0.001	0.303	-0.0440	-0.84	-0.017	0.377
Junior	-0.0641	-1.14	-0.020	0.287	-0.1458	-2.49 ***	-0.054	0.287
Secondary	-0.2162	-3.70 ***	-0.066	0.202	-0.2085	-3.24 ***	-0.076	0.157
Higher	-0.8391	-5.10 ***	-0.202	0.090	-1.0136	-3.90 ***	-0.287	0.035
Voc diploma	-0.1803	-1.00	-0.055	0.060	-0.4502	-1.64	-0.153	0.030
Other variables								
Ownship	0.1450	3.33 ***	0.046	0.659	0.2411	5.10 ***	0.089	0.662
Numdep	0.0682	8.76 ***	0.022	2.204	0.0679	8.25 ***	0.026	2.509
Urban	0.2705	3.80 ***	0.086	0.568	0.2554	3.39 ***	0.097	0.427
Male	-0.2113	-6.33 ***	-0.068	0.542	-0.2208	-5.86 ***	-0.083	0.534
Race								
African	0.9512	6.96 ***	0.257	0.722				
Colored	0.5095	3.41 ***	0.182	0.098				
Indian	0.3735	2.46 ***	0.132	0.034				
Location								
Homeland	0.4125	4.45 ***	0.136	0.379	0.3633	3.73 ***	0.136	0.524
W. cape	-0.1355	-1.30	-0.042	0.102	-0.1099	-0.66	-0.041	0.028
N. cape	0.5219	2.57 ***	0.190	0.013	-0.2646	-0.80	-0.094	0.003
E. cape	0.3134	3.18 ***	0.108	0.110	0.3246	2.80 ***	0.126	0.130
Kwazulu natal	-0.0963	-1.06	-0.030	0.203	-0.1394	-1.29	-0.052	0.216
Ofs	-0.1243	-1.29	-0.038	0.077	-0.2281	-2.53 ***	-0.083	0.093
Mpumalanga	-0.2121	-2.06 **	-0.064	0.090	-0.2807	-2.40 ***	-0.101	0.114
N. province	0.1139	0.92	0.038	0.083	0.0896	0.64	0.034	0.106
N.w. province	-0.1266	-1.06	-0.039	0.092	-0.1492	-1.16	-0.055	0.118
Community var								
Impass	0.0007	0.21	0.000	4.451	0.0037	1.88 *	0.001	3.197
Numfaci1	0.1026	2.06 **	0.033	0.374	0.1139	2.13 **	0.043	0.492
Cownship	0.1863	1.70 *	0.060	0.668	0.2125	1.74 *	0.080	0.680
Constant	-0.9924	-5.22 ***			-0.0484	-0.36		
Ln L		-6416.24				-5306.94		
Restricted Ln L		-8165.33				-6349.34		
Pseudo R^2		0.2142				0.1642		
N		13154				9496		
Mean of dependent variable		0.312				0.390		

The base or reference categories are as follows: Age: persons aged 16-20 years old; Education: persons with no education; Race: Whites; and Province: PWV or Gauteng. Footnote 7 contains definitions of the community variables.

[1] For example, an ILO report on the South African labor market (ILO, 1996) claims that it may be wrong to consider as labor force participants jobless persons who report that they want work but who do not search. The report points out that in the OHS94, many such persons were in their 30s and 40s and had never been employed before, implicitly casting doubt on the notion that these were genuine labor force participants. We cannot tell from the SALDRU survey whether the non-searching unemployed had ever held a job before.

[2] The October Household Survey questionnaire first asks about a person's main activity during the past seven days, with the options 'unemployed but looking for work' and 'not working, not looking for work'. From these, it excludes those who did some work (formal or informal) for pay, profit, or family gain during the past year and those who may not have worked in the past week but who had a job or enterprise or an attachment to a job or enterprise such as a business, farm, etc. Further, it excludes those who may not desire to work by asking 'if a suitable job is offered to (name), will (name) accept it?'. It then excludes persons who - though they may desire work - cannot be regarded as genuine work force participants such as housewives, students, and disabled persons. Finally, it asks a 'sweeper' question about how the unemployed labor force participants supported themselves, in order to determine their access to income. The first option in the question was 'did odd jobs during the past week'. Those who answered in the affirmative to this option were excluded from the list of the unemployed. Thus, the October Household surveys ask a detailed set of hurdle questions before admitting an individual as unemployed (Bhorat, 1999).

[3] Some caution is necessary when considering the separation of voluntary quitters from involuntary quitters. For example, a worker who knew her firm was going to fold shortly might quit 'voluntarily' before the event occurred.

[4] The unrestricted log likelihood was obtained from a pooled unemployment probit which included all the variables as well as all variables interacted with the race dummies. The restricted log likelihood was obtained from a pooled unemployment probit which included just the variables and no race interaction terms. Thus, for example, the pooling of the white and African samples was easily rejected: $\chi^2 = 1282.4$. The pooling of other races was rejected as well.

[5] We wished to utilise cognitive skill test scores as proxies for quality of schooling received. However, there are several drawbacks associated with the test score data in the SALDRU survey. Firstly, tests were administered only to one in six of the sample households and within each of these households, it was given to only two members of the household, one of whom was in the age group 13-17 and one over 17. In total, 1330 individuals older than 17 took the test, but less than 500 of these were labor force participants. The test takers over the age of 17 are split 65:35 women to men. It seems that the tests were administered at times when school children were present, but when working adults were likely not to be. As a result, the adult test takers are predominantly women and few report any wage income. This selection is likely to jeopardise any general inferences about the links between test scores and labor market outcomes such as earnings or employment.

[6] The higher incidence of unemployment among the old (>45 years old) might be explained by their waning productivity, especially if it falls relative to their wage. This is likely to result in their greater incidence of involuntary entry into unemployment. Moreover, being less adaptable, they are more likely to have longer duration of unemployment.

[7] The SALDRU dataset has rather better community level information available which enables us to capture aspects of cost of job-search and demand for labor, both potentially important determinants of unemployment. Whether there are any roads that become impassable at certain times of the year (Impass) is a proxy for cost of job-search. Total number of facilities in the community (Numfaci) - such as restaurant, post-office, bank, daily market, etc. - is a measure of the economic development of the community and, as such, at least a crude measure of the local demand for labor. Because community level information is missing on 24 clusters in the SALDRU data, we have assigned the overall mean value of Numfaci to clusters where Numfaci was missing. But, given the discrete (0/1) nature of the Tarroad and Impass variables and given the likelihood that the clusters with missing community schedules are those in remote areas, we have assigned a value of 0 for Tarroad and of 1 for Impass in these 24 clusters. The results are shown in Appendix 1.

[8] The effect of Numfaci is different for Africans and whites (results for whites not shown in Appendix 1 but available from the authors upon request): while for whites Numfaci proxies the local demand for labor - with greater facilities reducing the risk of unemployment - for blacks, the positive effect of Numfaci suggests that African unemployed job-seekers migrate to clusters where there is greater demand for labor.

[9] See a synthesis of research on discrimination using this method in Darity (1995).

[10] In our analysis, the difference in the two means is relatively small. For example, in the sample of African labor force participants, the actual mean of the dependent variable (unemployed=1; employed=0) is 0.406 and the predicted mean is 0.382.

[11] On the conventional method of decomposition (Oaxaca, 1973; Gomulka and Stern, 1990) and using both the white characteristics and the characteristics of the comparator race group (Africans, coloureds, and Indians in turn), the part not explained by characteristics represents 11.7 or 5.7 percent, 9.9 or 6.7 percent, and 4.9 or 5.2 percent respectively, the average of the two measures being 8.7, 8.3, and 5.1 percent respectively. The results are thus not sensitive to the choice of method.

[12] Frijters noted that the lower observed productivity of African employees may have been because they were in a minority in the firm and, thus, may have had higher transaction costs in their communication with the majority Indian employees. In other words, he did not claim that African workers are in general less productive than others.