BER
BUREAU FOR ECONOMIC RESEARCH

UNIVERSITEIT·STELLENBOSCH·UNIVERS
jou kennisvennoot • your knowledge partn

Bureau for
Economic Research

Department of
Economics

University of Stellenbosch

EARNINGS FUNCTIONS, LABOUR MARKET DISCRIMINATION AND QUALITY OF EDUCATION IN SOUTH AFRICA

Doubell Chamberlain & Servaas van der Berg

Stellenbosch Economic Working Papers : 2 / 2002

Doubell Chamberlain
Genesis Analytics: Department of Economics
University of Stellenbosch
E-Mail: doubellc@genesis-analytics.com

Servaas van der Berg
Department of Economics
University of Stellenbosch
E-Mail: svdb@sun.ac.za

EARNINGS FUNCTIONS, LABOUR MARKET DISCRIMINATION AND QUALITY OF EDUCATION IN SOUTH AFRICA[1]

Doubell Chamberlain[2] & Servaas van der Berg[3]

ABSTRACT

Education is a key determinant of earnings, as several South African studies have confirmed. Years of schooling completed, however, provides an imperfect approximation of the effective level of education achieved, mainly due to variations in the quality of education received. This study addresses this issue by, for the first time in South Africa, incorporating quality of education in the modelling of earnings. Differences in quality of education are viewed as a form of pre-labour market discrimination. By decomposing the wage gap before and after controlling for educational quality, more accurate estimates of the true levels of labour market discrimination are obtained. The main hypothesis tested is that controlling for quality will reduce the component of the wage gap ascribed to labour market discrimination.

The results show a systematic decrease in the labour market discrimination component with increased adjustments for quality of education. Almost half of the previous labour market discrimination can be explained by differences in quality, yet the proportion of racial wage differentials ascribed to labour market discrimination is still found to be significant. The clear implication is that current estimates of labour market discrimination are exaggerated and a more careful analysis of earnings is required to re-assess the levels of discrimination in the South African labour market.

[1] This paper flows from the masters thesis at the University of Stellenbosch by the first author (Chamberlain 2001). Financial assistance from the NRF is gratefully acknowledged. A previous version of this paper was presented to the Econometric Society of South Africa's 2002 conference in the Kruger National Park and to the DPRU/FES 2002 conference on *Labour Markets and Poverty in South Africa* in Johannesburg.

[2] Genesis Analytics & University of Stellenbosch

[3] University of Stellenbosch

1. Introduction

The burgeoning literature on the earnings function in South Africa has paid much attention to evidence of labour market racial discrimination, conventionally measured as that part of earnings differentials between groups not accounted for by productive characteristics. Yet several articles in this literature admit, as Jacob Mincer (1974) also did in his groundbreaking work on human capital and earnings functions, that traditional measures of education (years of schooling completed) are an inaccurate measure of the human capital transferred by education. This measurement deficiency is caused by differences in the content (e.g. subject choice and school curricula) quality of education provided. To accurately determine the contribution of education to an individual's productive characteristics, it is necessary to find better measures of the human capital accumulated through schooling. This study takes a first step towards addressing this issue by, for the first time in South Africa[4], incorporating measures of schooling quality into the analysis of earnings.

In the human capital model developed by Mincer (1974), earnings were explained as a function of acquired human capital which, in turn, was expressed as a function mainly of education (proxied by years of schooling completed) and experience. This model has since been expanded to include other factors that may influence earnings, such as gender and location. Already at the time of specifying the original model, Mincer acknowledged that differences in the quality of education received were a potential weakness and suggested that making provision for this in the model would greatly enhance its explanatory power (Mincer, 1974: 55).

Adjusting the earnings model to incorporate quality of education also affects the measurement of the proportion of the wage gap (i.e. the difference in average earnings of two groups) ascribed to labour market discrimination. For purpose of this study (and based on the work of Oaxaca (1973)), labour market discrimination will be defined as the component of the wage gap left unexplained by differences in measurable productive characteristics of the compared groups. The main hypothesis is that taking account of the quality of education will reduce the component ascribed to labour market discrimination between whites and blacks, the two largest population groups in South Africa. This does not necessarily imply less discrimination, but presupposes that discrimination can affect the individual both in and before entering the labour market. In South Africa, racial differences in the quality of education received can better be considered as pre-labour market discrimination. Underlying the hypothesis is, therefore, the proposition that educational discrimination leads to disparities in the quality of education received which will lead to varying valuations of educational attainment in the labour market.

[4] In a recent conference paper, thus far unpublished, Kingdon & Knight (2002) attempt a similar procedure, but using educational inputs <u>at the magisterial district level</u> as measure of quality.

This suggests that labour market discrimination may currently be overestimated. From a policy perspective, this supports the importance of attention to access to quality education.

2. The Model Set-up

2.1. Specifying the Model

In the simple Mincerian model, earnings are specified as a function of acquired human capital, which in turn is a function of education (measured in years of schooling completed) and experience. In a perfectly functioning labour market the individual's earnings are determined by the value that the market places on the acquired human capital, because of the productivity associated with it.

Thus human capital acquisition takes place pre-labour market and in the labour market (see figure 1 below). The pre-labour market period covers the period up until entering employment and includes primary, secondary and tertiary full-time education. The individual enters the first stage with a human capital stock determined by the unmeasured variable "ability". As it is very difficult to control for the influence of this variable on the accumulation of human capital and sufficient data is not available, it will not be explicitly accounted for in this analysis. This should not detract from its potential influence on the accumulation of human capital. Ability should not play a major role in intergroup differences if it is assumed to be distributed equally across different groups, although it may account for earnings differentials between *individuals* with similar measured human capital. In the first stage, the individual is in full time education and human capital accumulation is determined by years of schooling completed and the quality of such schooling.

The labour market period follows the participation decision. Here the contribution of further education (on the job training and part-time further education) towards the human capital stock gradually diminishes relative to that of experience.

In both these periods, discrimination may arise. Usually in the earnings function literature, the effect of labour market discrimination is considered, but in South Africa a significant component of discrimination occurs in the pre-labour market period, *inter alia* through differences in the quality of education provided. The fact that both pre-labour market and labour market discrimination are strongly structured along racial lines makes it difficult to distinguish the exact effect and its influence on earnings.

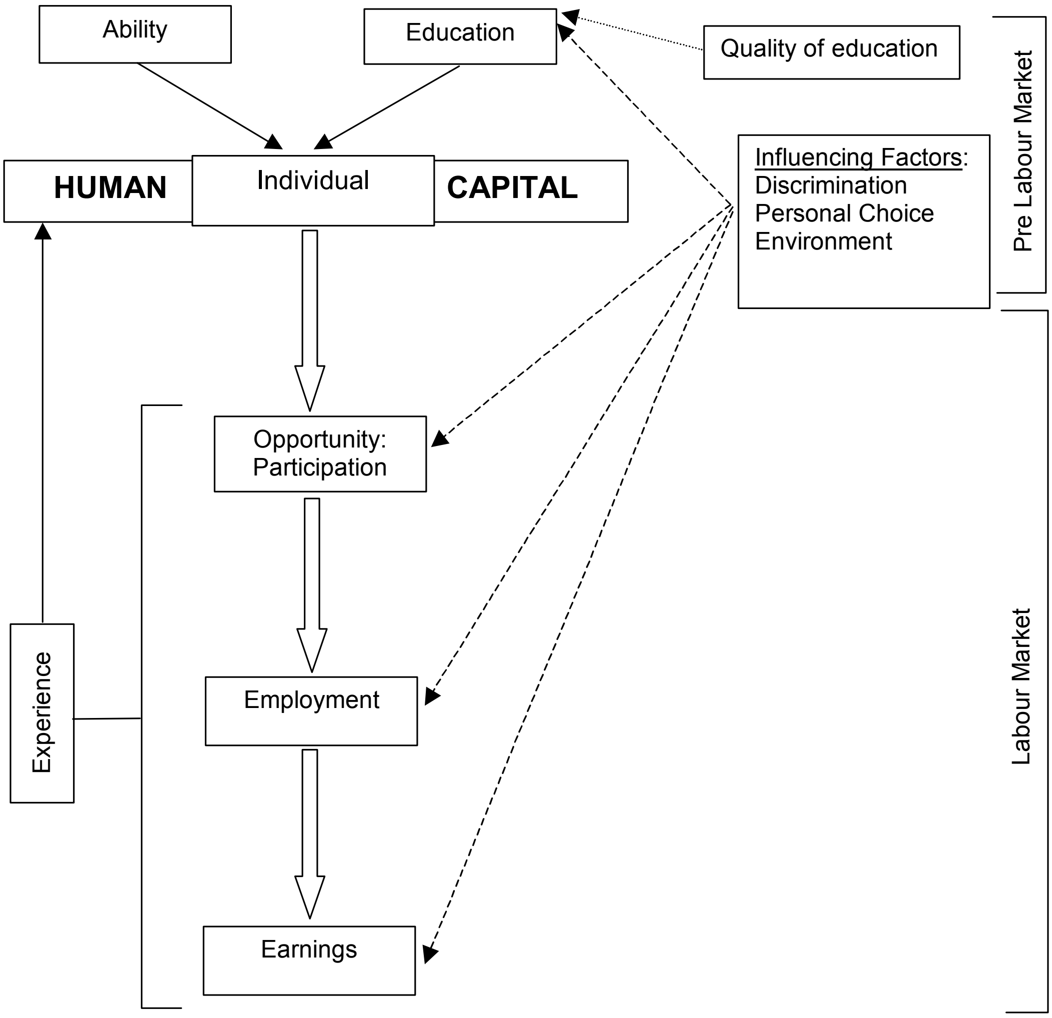

Figure 1: Human Capital Accumulation over the Lifetime of an Individual

Two potential complications to this model in the South African context are that:

- The level of education as measured by years of schooling completed (educational attainment) is not necessarily an accurate measure of the acquired human capital (due to differences in the quality of education received).
- The market does not function perfectly and the market valuation of an individual's human capital stock is often biased, due to imperfect information, discriminatory policies or outright bias on the part of employers (Fallon & Lucas, 1998: 12).

One possible way of dealing with the first complication is to control for the effect of differences in the quality of education received on the individual's level of human capital and therefore earning ability. The measurable variable *years of schooling completed* is adjusted to approximate the unmeasurable variable *effective years of schooling* received. The implicit assumption is that the latent variable *effective schooling* (and therefore human capital) accumulates on a continuous and linear scale parallel to the observable variable *years of schooling*. By adjusting years of schooling, an individual completing, for example, grade 7 may possibly only accumulate, say, 90% of the effective schooling of a standard quality grade 7 year. In this manner, an individual may complete twelve years of schooling but only accumulate, say, eleven years of effective education.

By controlling for the first problem mentioned above in this way, a clearer picture of the extent of the second problem emerges. Differences in earnings between two groups (in this case blacks and whites) can then be decomposed to distinguish between differences in characteristics (e.g. level of education) and how the same characteristics are rewarded between groups. This difference in reward to productive characteristics serves as an estimate of discrimination, as it is unexplained by measurable differences between the groups. By controlling for quality of education (and, therefore, pre-labour market discrimination), the estimate of discrimination arrived at is reduced to represent labour market discrimination only. Comparing this estimate of discrimination before and after controlling for quality provides a measure of the extent of the second problem referred to above.

2.2. Modelling the employment process

The process of selection into employment is often modelled as a binary choice model where individuals are confronted with the 'choice' of being employed or remaining unemployed. The assumption underlying such an approach is that all unemployment is voluntary (Bhorat & Leibbrandt, 2001b: 113), which is clearly not the case in South Africa. Bhorat and Leibbrandt (2001b) show that it is not appropriate to model selection into employment in this manner. Firstly, the individual can only choose to participate in the labour market. This choice will be affected by several factors such as the individual's preferences, location, level of education and general circumstances. Secondly, the individual's choice to participate does not guarantee employment. The selection into employment may be influenced by the same factors as the choice to participate, but also by other factors such as demand for labour and the characteristics of competing participants.

Although only those in employment can earn an income, it would lead to biased results if earnings were simply modelled on the sub-sample of wage earners, who may not be a representative sample of all persons receiving education. Thus a fitted regression earnings function may confound the relationship between education and earnings with parameters of the function determining the probability of selection into the sample (Heckman, 1979: 154). In his seminal article on this topic, Heckman (1979) showed that this problem reduces to the omitted variable problem. To correct for the bias, we include the inverse of the Mill's ratio (λ) as a regressor in the earnings function. Consequently, in order to control for sample selection bias, the modelling of earnings functions takes place in three sequential phases, modelling participation, employment and eventual earnings (of those employed) separately. Modelling the three stages separately also makes it possible to include explanatory factors appropriate to each stage, leading to a better specified model.

2.3. Earnings Functions

In its simplest form, the Mincerian model considers earnings as a function of the individual's measured human capital stock which depends on education, experience and the proportion of time invested in human capital attainment after basic schooling (Mincer, 1974: 9; Fallon & Verry, 1988: 149). To illustrate the model proposed by Mincer the following variables need to be defined:

Y_t: Potential earnings after t years

r : The rate of return (continuous) to investment in human capital

T: Total working lifetime of individual

h_t: The proportion of time invested in human capital attainment during time t

S: Years of schooling (educational attainment)

x: Years of experience

t: Time in years

Consider the level of income of an individual with S years of schooling (Y_S). Assuming that h_t=1 during schooling (for t $\in$ [0,S]), the level of income at the end of S years can be given by:

$$Y_S = Y_0 e^{rs} \qquad (2.3a)$$

In the post-school period, the ratio h is assumed to decline linearly with age and the lifetime relation can be defined as follows:

$$h_t = h_0 - \left(\frac{h_0}{T}\right)t \qquad (2.3b)$$

where h_0 is defined as the proportion of time invested in human capital attainment in the period directly following full time schooling. To simplify the calculations it is assumed that the return to investment in human capital is a constant continuous rate r. If we combine this with the equations (2.3a) and (2.3b) above, the income of an individual x years after leaving school is:

$$Y_x = Y_S e^{r\int_0^x h_t dt} \qquad (2.3c)$$

Substituting (2.3b) in (2.3c):

$$\begin{aligned} Y_x &= Y_S e^{r\int_0^x \left(h_0 - \left(\frac{h_0}{T}\right)t\right)dt} \\ &= Y_S e^{r\left(xh_0 - \frac{h_0}{2T}x^2\right)} \qquad (2.3d) \end{aligned}$$

Taking the natural logarithms:

$$\ln Y_x = \ln Y_S + r\left(xh_0 - \frac{h_0}{2T} x^2 \right) \qquad (2.3e)$$

Substituting (2.3a) in (2.3e)

$$\begin{aligned} \ln Y_x &= \ln Y_0 e^{rS} + rxh_0 - \frac{rh_0}{2T} x^2 \\ &= \ln Y_0 + rS + rh_0 x - \left(\frac{rh_0}{2T} \right) x^2 \end{aligned} \qquad (2.3f)$$

This equation expresses income as a function of schooling (S) and experience (x) and is known as the ‘standard’ Mincerian income function. In a similar fashion, the standard function can be expanded to control for other factors such as gender, location and sector of employment, but in practice the Mincerian function is still often used in its simplest form, controlling only for experience and schooling.[5] This is in part evidence to the intuitive elegance of the standard Mincerian function and its powerful explanatory capability, but also often a result of insufficient data to properly specify and test an elaborated version function.

2.4. Modelling Quality

There may be significant differences in the effective level of schooling received by students with the same educational attainment. One possible explanation is the differences in the quality of education received.

Other than introducing quality measures as dummy variables, two methods have been proposed to account for differences in quality within the Mincerian framework (Behrman & Birdsall, 1983: 931; Heckman, Layne-Farrar & Todd, 1996: 564). The first method adjusts the rate of return to reflect differences in quality:

Taking the natural log of (2.3a) we get the equation:

$$\ln Y_S = \ln Y_0 + rS \qquad (2.4a)$$

Let the rate of return (r) be a function of quality, so that higher quality education is awarded a higher rate of return. As functional relationship between rate of return and quality, Behrman & Birdsall (1983) proposed the quadratic form.

$$\begin{aligned} r &= r(q) \\ &= r_0 + r_1 q + r_2 q^2 \end{aligned}$$

[5] Psacharopoulos & Tzannatos (1992) contains several examples where Mincerian earnings functions were applied in their simplest form.

where q is a measure of quality. This specification allows for the possibility of diminishing return to q (if rS > 0 and r_2 < 0). Substituting this into (2.4a) leads to the final Mincerian equation adjusted for quality:

$$\ln Y_S = \ln Y_0 + \left(r_0 + r_1 q + r_2 q^2\right)S \quad (2.4b)$$

The second method uses a measure of quality to adjust the actual years of schooling completed in order to reflect effective years of education:

Define effective years of schooling (S*) as a function of years of schooling (S) and a proxy for quality of education (q):

$$S^* = S^*(S, q) \quad (2.4c)$$

Substitute (2.4c) into (2.4a):

$$\ln Y_S = \ln Y_0 + rS^*(S, q) \quad (2.4d)$$

One common specification of equation (2.4c) is S* = qS, where the proxy for quality (q) serves as a scale adjustment to years of schooling (Heckman, Layne-Farrar & Todd, 1996: 565). The second method will be followed in this study.

Measurement of educational quality van be approached via inputs into the system (resources) or via the outputs (results). Heckman, Layne-Farrar & Todd (1996) and Behrman & Birdsall (1983) both defined their measures of quality from the input side, using a series of complex relationships between inputs and pricing equations. This approach assumes that the level and quality of inputs play a dominant role in the determination of educational quality. But measurement from the input side does not take into account the efficiency with which inputs are applied. It may provide a good indication of *potential* educational quality, *actual* quality achieved may be quite different.

The alternative approach focuses on educational outputs, measuring what has been achieved without requiring assumptions on the efficacy of inputs (Archer, 1995: 6). But as it does not control for levels of inputs used to achieve the measured outputs, it can also not shed any light on the efficiency with which the inputs were applied and on how levels of quality were achieved. Care should also be taken to select measures of achievement that encompass all aspects of quality education.

Which approach to use will be determined by the specific questions posed and data availability. The output approach provides better measures of quality (although data is seldom available), while the first provides a better picture of how quality was achieved. In this study, the focus is not on understanding quality as such but rather on obtaining a good indicator of quality to adjust the years of schooling regardless of how the levels of quality were attained. The output-side approach, therefore, seems appropriate for approximating quality. Input-side methodology will also be used to transfer the proxy for quality (test scores) from the LSDS 1993 to the October Household Survey 1995 (OHS95) dataset (see below).

To apply this output-based specification of quality requires a quality of education index (q). This is obtained using scores on the literacy and numeracy tests included with the South African Living Standards and Development Survey (LSDS93; see Saldru 1994). As the literacy and numeracy tests were included in the LSDS93 survey, but not in the October Household Survey 1995 (OHS95), it was necessary to obtain predicted test scores for the OHS95 using a model fitted on the LSDS93 data. In essence, the procedure uses an instrumental variable, obtained from another dataset (LSDS93) and then applied to the earnings function dataset (OHS95) to calculate predicted scores.

Level of education is the main predictor of these test scores (Van der Berg, Wood & Le Roux, 2002), but as the intention is to use the quality measure to adjust the level of education in the OHS95, we could not include education as a predictor of test scores. To deal with this, the fitting of the quality function was divided into two stages. In the first stage, a simple regression equation was fitted, regressing education only on test scores in a single regression for both blacks and whites. Using a single regression for both population groups provided a way of standardising the residuals around the same regression function. In the second stage, the residuals from step one (the component of test scores not explained by education) were used to fit a more elaborate regression function (including factors such as provincial and urban location and gender) for blacks and whites separately[6]. By separating the two population groups, it was possible to test for different influencing factors in each. In this way, it was possible to control for education in the LSDS93 function without explicitly including it in the final regression.[7]

[6] As the adjusted education variable will be applied in a regression function modelling levels of earnings, variables pertaining to levels of income also had to be excluded from the total score regression.

[7] Running separate equations for, for example, males and females is justified as they can possibly be considered as two separate employment markets. It is, therefore, not a question of sensoring the sample, but rather of using two different samples from two different populations. This justification cannot be applied to the population groups as all groups operate within the same employment market. To handle this properly a particular racial sub sample should have been considered as a censored sample of the population. Due to the increased complexity of doing this, it was decided not to apply the strict methodology and to run the different racial equations without controlling for censoring. Although this can potentially bias the results, this should not be an enormous problem as some degree of segregation still exists and the population groups, therefore, could be considered as quasi-independent markets.

The regression function obtained in stage two was used to predict test scores for the OHS95. The maximum combined total score obtainable for the literacy and numeracy test is fourteen. Quality of education is not the main contributor towards the human capital stock of an individual, but rather has an enhancing or dampening effect on the ability of the student to absorb the educational content provided. To model this modulating effect of quality on the level of education, the predicted test scores were used to proportionally adjust school attainment. The choice of the proportion to be adjusted depends on the weights given to years of schooling and quality of schooling respectively. This, in turn, is based on assumptions of how debilitating or enhancing the effect of poor or good quality of education can be on the accumulation of human capital stock. As there are no clear guidelines for these weights, alternative weights ranging from 0 to 80% were tested as proportions by which years of schooling should be adjusted for quality. This also provided a simple sensitivity test for the choice of adjustment. Let x be the proportion (expressed as a percentage) to be adjusted, then effective education (E*) can be defined as:

$$E^* = (1-x)\times E + (x)\times\left(\frac{Score}{14}\right)\times E \tag{2.4e}$$

where x is the percentage adjustment allowed and 'Score' is the predicted test score on the literacy and numeracy test. Using the adjusted years of education (effective years of education), new education splines were then calculated.[8]

In the case of a 10% adjustment, the assumption would be that poor quality of education can at worst cause the individual to end up with 10% less than the potential level of education in a standard education. In the case of a 80% adjustment, the proposition is that most of the educational attainment of the student should be disregarded due to the poor quality thereof.

3. Model Results

3.1. Estimating the Quality of Education Function and Quality Proxy

In the first step the level of education was regressed on the total score variable. The residuals represented the variation on test scores not explained by total education. Regressing the remaining explanatory factors on the residuals resulted in a function that accounted for the level of education while not explicitly including it. The results for the separate quality functions fitted for blacks and whites are shown in Table 1 and 2.[9] Several specifications of quality functions for the respective population groups were tested. The general-to-specific approach

[8] The spline categories are kept the same, but the values within the splines are replaced by the adjusted values for years of education received.

[9] Detailed results available on request.

was followed – modelling started with the most elaborate specification, which was then reduced based on statistical inferences. In general, the reduction of the models improved the fit slightly, but the coefficients on the remaining variables were very stable. This was encouraging, as the specified models seemed quite robust.

Table 1: Quality Function for White Population Group

Variable	Coefficient	Standard Error	t	P > |t|
Urban dummy	-2.008035	0.790475	2.540	0.012
Location dummy: Kwazulu Natal	2.060628	0.491822	4.190	0.000
Location dummy: Mpumalanga	-3.670158	1.510216	-2.430	0.016
Persons per room	1.313843	0.468802	2.803	0.005
Age	0.036102	0.009414	3.835	0.000
Number of Observations	297			
F(5,292)	42.81			
Prob > F	0.0000			
R-squared	0.4320			
Adjusted R-squared	0.4131			

Several variables from different categories were tested for inclusion in the test score regression. Amongst these were household environment (including variables such as brick house, electricity, total number of rooms in house, number of people in household, family size and number of persons per room), location (provincial and rural/urban) and personal characteristics (age and gender).

Table 2: Quality Function for Black Population Group

Variable	Coefficient	Standard Error	t	P > |t|
Urban dummy	0.397368	0.182768	2.174	0.030
Location dummy: Northern Cape	-0.641825	0.176625	-3.634	0.000
Location dummy: Free State	-0.923303	0.129228	-7.145	0.000
Location dummy: Kwazulu Natal	-2.123559	0.235941	-9.000	0.000
Location dummy: North-West	-1.720311	0.226019	-7.611	0.000
Location dummy: Gauteng	-2.160045	0.292306	-7.390	0.000
Number of Observations	1597			
F(6,1591)	42.39			
Prob > F	0.0000			
R-squared	0.1378			
Adjusted R-squared	0.1346			

The first step in the next phase was to calculate predicted test scores for the OHS95 dataset based on the two equations estimated above. Once these values had been calculated, it was possible (step 2) to adjust the years of education variable for the measured quality of education received using equation (2.4e) above.

As mentioned before, there are no clear guidelines to determine the magnitude of adjustment to be allowed, i.e. the weights given to years of education and quality of education respectively. Accordingly, calculations were done for adjustments of 10%, 20%, 40% and 80%. The remaining calculations were repeated for each of these levels.

3.2. Estimating the Labour Force Participation Probit

Both the participation decision and the employment phenomenon are binary responses: The individual participate/does not participate or is employed/unemployed. Although it is possible to model this type of variable using OLS regression[10], there are several reasons why OLS is not optimal for modelling dichotomous variables (Aldrich & Nelson, 1984: 11; Long, 1997: 38):

- It violates the assumption of homoscedastic error terms on which OLS regression is based. This means that the estimates of β are inefficient and the standard errors are biased.
- The value of the dependent variable is not restricted to the range [0,1], which makes the interpretation of predicted values problematic.
- Linearity implies that changes in the independent variables have a similar effect on the value of the dependent variable irrespective of the current value. The sigmoid shape of probit/logit models seems to get closer to the true non-linear relationship between the variables.

For the purpose of this study the probit method was chosen to cope with these problems. As the first of three phases in fitting an earnings function (corrected for selectivity), labour force participation probits were estimated for the black and white population groups using the expanded definition of participation. From these functions, Mills' inverse ratios were calculated to be included in the employment probits fitted in phase two. The results for the respective participation probits are shown in Tables 3 and 4.

The benefit of modelling the labour market process in three stages, i.e. the ability to include variables that are specific to each stage, now becomes. Household variables, for example, play an important role in the participation decision but may have little influence on the probability of finding employment. In the labour force participation probit, household variables such as other household income per capita and the presence of other working age males and females, pensioners and small children were included, the expectation being that higher household income and the presence of pensioners and other working age males and females (representing other sources of income for the household) will reduce the probability of participation. The presence of small children was expected to have a negative effect on the

[10] See discussion of the Linear Probability Model in Long (1997: 35) and Aldrich & Nelson (1984: 14)

probability of participation for females. Other variables included were age, education (with adjustments for quality) and an urban dummy. Being in an urban location, being older and having higher levels of education as well as being male, were all expected to increase the individual's probability of participation.

The urban dummy was significant for both the white and black functions and showed a positive coefficient that was consistent across level of adjustment. This was in line with expectation, as individuals often move to urban areas in search of work and it can, therefore, be expected that urban individuals have a higher propensity to participate.

The different sections of the education spline were found to be significant across both population groups. The only exception was the primary education spline for the white population in the case where there was an 80% adjustment for educational quality. For both whites and blacks, the coefficients on all three splines were positive, confirming the expected positive relationship between education and participation. In the case of the black population, the relationship seemed stronger for the higher education levels, whereas for whites the lower education levels seem to have a stronger impact on participation.

All age categories were found to be significant for blacks, whereas only some categories were significant for whites. The results indicated a larger probability of participation (relative to the control category: 15-24) up until the 46-55 category in the case of blacks. For whites, age seems to decrease the probability of participation as negative coefficients were found in most cases.

The female dummy was negative and significant for both races across all levels of adjustment. The presence of children aged less than 7 years in the household had a significantly negative influence on the probability of participation for the white population group, whereas children aged 8-15 significant reduced black participation. Winter (1998) found similar results and suggested that it is because black women stay attached to the labour market even through their childbearing years. Aggregating the genders could also have reduced the observed negative effects of children on labour force participation.

The coefficients on the number of males ages 16-59 and adults older than 60 in the household were negative and significant for both groups across all levels of adjustment, supporting the expectation that other potential earners or pensioners in the household reduced labour force participation. The results for females aged 16-59 were also significant for all categories but had a positive coefficient. Once again, this coefficient may have been more illuminating if separate functions were fitted for males and females. Bhorat & Leibbrandt (2001b) found a strong positive relationship in the case of black females and a positive but smaller effect in the case of black males.

Contrary to expectation, household income variables showed a positive (though very small) influence on participation. It may be more appropriate to look at ‘other household income’ (excluding that of the individual concerned) and not total household income as was done here.

Table 3: White Participation Probit (Marginal Effects)

Variable Name	No Adjustment		10% Adjustment		20% Adjustment		40% Adjustment		80% Adjustment	
	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|
Urban dummy	0.0822265	0.000	0.0857281	0.000	0.0890541	0.000	0.0970701	0.000	0.1025426	0.000
Education Spline: None-Std 5	0.0315526	0.001	0.0253494	0.000	0.0275455	0.001	0.0313876	0.001	0.0142556	0.404
Education Spline: Std 6-10	0.0121757	0.000	0.0109378	0.000	0.0117956	0.000	0.0129661	0.000	-0.0068995	0.000
Education Spline: Tertiary	0.0129840	0.000	0.0122025	0.000	0.0132131	0.000	0.0150282	0.000	0.0126632	0.000
Age: 26-35	0.0063034	0.421	0.0063114	0.421	0.0060122	0.443	0.0053828	0.494	0.0062319	0.437
Age: 36-45	-0.0196405	0.017	-0.0203683	0.013	-0.0210584	0.011	-0.0226612	0.000	-0.0224705	0.000
Age: 46-55	-0.0436836	0.000	-0.0454048	0.000	-0.0467347	0.000	-0.0501452	0.000	-0.0544404	0.000
Age: 56-65	-0.0875182	0.000	-0.0913565	0.000	-0.0937006	0.000	-0.0999814	0.000	-0.1092333	0.000
Female dummy	-0.3322145	0.000	-0.3326177	0.000	-0.3326328	0.000	-0.3327713	0.000	-0.3343684	0.000
No. of Kids < 7	-0.0219725	0.000	-0.0219842	0.000	-0.0220188	0.000	-0.0220734	0.385	-0.0217405	0.000
No. of Kids 8-15	0.0034264	0.357	0.0034062	0.360	0.0033629	0.367	0.0032487	0.000	0.0030128	0.423
No. of Males 16-59	-0.0325843	0.000	-0.0331124	0.000	-0.0333339	0.000	-0.0339688	0.002	-0.0354500	0.000
No. of Females 16-59	0.0150179	0.000	0.0148679	0.001	0.0146958	0.001	0.0142454	0.000	0.0133706	0.005
No. of Adults > 60	-0.0299349	0.000	-0.0302432	0.000	-0.0305512	0.000	-0.0312712	0.000	-0.0315407	0.000
Per Capita Hhld Income	0.0000068	0.000	0.0000069	0.000	0.0000069	0.000	0.0000070	0.000	0.0000076	0.000
Per Capita Hhld Income Squared	-0.00000000002	0.000	-0.00000000002	0.000	-0.00000000002	0.000	-0.00000000002	0.000	0.0000000000	0.000
Pseudo R2		**0.3069**		**0.3060**		**0.3058**		**0.3051**		**0.3007**
Nr Obs		**4,649,077**		**4,649,077**		**4,649,077**		**4,649,077**		**4,649,077**

Table 4: Black Participation Probit (Marginal Effects)

Variable Name	No Adjustment		10% Adjustment		20% Adjustment		40% Adjustment		80% Adjustment	
	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|
Urban dummy	0.1146254	0.000	0.1141309	0.000	0.1136572	0.000	0.1122890	0.000	0.1086900	0.000
Education Spline: None-Std 5	0.0071631	0.000	0.0064314	0.000	0.0072664	0.000	0.0097043	0.000	-0.0438170	0.000
Education Spline: Std 6-10	0.0085710	0.000	0.0080385	0.000	0.0091245	0.000	0.0124186	0.000	-0.0249845	0.000
Education Spline: Tertiary	0.0193643	0.000	0.0189719	0.000	0.0215010	0.000	0.0292359	0.000	0.0561570	0.000
Age: 26-35	0.0894897	0.000	0.0895112	0.000	0.0895424	0.000	0.0896300	0.000	0.0915227	0.000
Age: 36-45	0.0834835	0.000	0.0833450	0.000	0.0834317	0.000	0.0836296	0.000	0.0827892	0.000
Age: 46-55	0.0324997	0.000	0.0324849	0.000	0.0326172	0.000	0.0328956	0.000	0.0280119	0.000
Age: 56-65	-0.0574947	0.000	-0.0569811	0.000	-0.0567919	0.000	-0.0564331	0.000	-0.0661105	0.000
Female dummy	-0.2496103	0.000	-0.2495703	0.000	-0.2495432	0.000	-0.2494668	0.000	-0.2512928	0.000
No. of Kids < 7	-0.0014367	0.561	-0.0014412	0.559	-0.0014278	0.563	-0.0013889	0.574	-0.0010367	0.675
No. of Kids 8-15	-0.0068600	0.002	-0.0068396	0.002	-0.0068403	0.002	-0.0068451	0.002	-0.0070987	0.001
No. of Males 16-59	-0.0216462	0.000	-0.0215997	0.000	-0.0216148	0.000	-0.0216645	0.000	-0.0229485	0.000
No. of Females 16-59	0.0082297	0.000	0.0082919	0.000	0.0082558	0.000	0.0081513	0.000	0.0073721	0.000
No. of Adults > 60	-0.0230556	0.000	-0.0229669	0.000	-0.0229938	0.000	-0.0230647	0.000	-0.0227550	0.000
Per Capita Hhld Income	0.0000454	0.000	0.0000458	0.000	0.0000457	0.000	0.0000454	0.000	0.0000531	0.000
Per Capita Hhld Income Squared	-0.0000000004	0.000	-0.0000000004	0.000	-0.0000000004	0.000	-0.0000000004	0.000	-0.0000000005	0.000
Pseudo R2		**0.1600**		**0.1600**		**0.1601**		**0.1603**		**0.1569**
Nr Obs		**27,986,645**		**27,986,645**		**27,986,645**		**27,986,645**		**27,986,645**

3.3. Estimating the Employment Probit

The next step was to fit the employment probits on the sample of participants using the inverse Mills' ratios (lambdas (λ)) calculated from the participation probit to control for selectivity. The lambdas were significant in all runs, indicating that sample selection bias was a relevant problem. The results are shown in Tables 5 and 6.

In contrast to the probit functions for participation, household variables were not included in the employment probit, as these should not influence the individual's probability of finding employment. Higher age and education levels should increase the probability of being employed, whereas being female will decrease it. In addition to these variables, some provincial dummies capture the effects of systematic differences in employment relative to the Western Cape.

There were a number of surprising findings. The urban dummy was negative and significant for both races across almost all levels of adjustment for educational quality. The results seem to indicate that individuals with similar other characteristics find employment in rural areas more readily than in urban areas when they do participate, but that participation rates in rural areas are significantly lower. A more detailed analysis of unemployment and participation is clearly necessary. The female dummy was negative and mostly insignificant for whites, but strongly positive and significant for blacks.

The coefficients on the education variables were significant for both population groups, except for primary education which was not significant for whites. For whites, the results showed a positive relationship between education and the probability of being employed. Surprisingly, however, only the highest education levels were found to have a positive effect on employment for blacks. Within the sample of participants, it seems that all those with less than tertiary education struggled to find employment. The sign of the coefficients for blacks remained the same but its magnitude increased with increased adjustment for education quality.

The age variables were significant and positive for both race groups. For whites, the location dummies for four provinces were found to be insignificant, whilst in the other four provinces, the probability of employment was higher than for the reference province (Western Cape). For blacks, only the Gauteng dummy was insignificant with all the other provinces (except the Free State) showing lower probabilities of employment than the Western Cape.[11]

[11] The changes in the coefficients and significance of the Kwazulu-Natal and Mpumalanga variables for whites and Northern Cape, Free State, Kwazulu-Natal and North West in the case of blacks stemmed from their inclusion in the education quality function (see section 3.1).

Table 5: White Employment Probit (Marginal Effects)

Variable Name	No Adjustment		10% Adjustment		20% Adjustment		40% Adjustment		80% Adjustment	
	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|
Urban dummy	-0.0146748	0.011	-0.0144854	0.013	-0.0140505	0.017	-0.0130385	0.032	-0.0028978	0.702
Education Spline: None-Std 5	0.0066447	0.302	0.0052510	0.282	0.0056404	0.296	0.0067206	0.326	0.0082770	0.619
Education Spline: Std 6-10	0.0043298	0.000	0.0036938	0.000	0.0039245	0.000	0.0045142	0.001	-0.0041446	0.002
Education Spline: Tertiary	0.0060068	0.000	0.0054993	0.000	0.0059344	0.000	0.0070755	0.000	0.0165015	0.000
Age: 26-35	0.0184453	0.000	0.0185144	0.000	0.0184262	0.000	0.0181977	0.000	0.0159752	0.000
Age: 36-45	0.0253205	0.000	0.0252608	0.000	0.0251436	0.000	0.0248761	0.000	0.0216446	0.000
Age: 46-55	0.0218981	0.000	0.0219361	0.000	0.0218326	0.000	0.0216324	0.000	0.0184455	0.000
Age: 56-65	0.0281678	0.000	0.0283200	0.000	0.0283409	0.000	0.0284305	0.000	0.0279422	0.000
Female dummy	-0.0135178	0.076	-0.0115769	0.128	-0.0109583	0.149	-0.0088578	0.242	-0.0014267	0.852
Eastern Cape dummy	-0.0016198	0.753	-0.0014904	0.773	-0.0014364	0.781	-0.0012757	0.805	-0.0003706	0.942
Northern Cape dummy	0.0228255	0.000	0.0229344	0.000	0.0229603	0.000	0.0230271	0.000	0.0231079	0.000
Free State dummy	0.0117632	0.016	0.0120206	0.014	0.0120799	0.014	0.0122309	0.013	0.0129755	0.007
Kwazulu Natal dummy	0.0042407	0.406	0.0036472	0.480	0.0026946	0.606	0.0002199	0.968	-0.0189869	0.064
North-West dummy	0.0040396	0.510	0.0041331	0.502	0.0041163	0.504	0.0040791	0.509	0.0037398	0.547
Gauteng dummy	0.0189973	0.000	0.0191588	0.000	0.0192022	0.000	0.0192941	0.000	0.0193306	0.000
Mpumalanga dummy	-0.0035256	0.529	-0.0012990	0.812	0.0007620	0.886	0.0055090	0.283	0.0219444	0.006
Northern Province dummy	0.0152967	0.019	0.0149627	0.022	0.0149520	0.023	0.0148836	0.023	0.0146140	0.025
Lambda	-0.0489534	0.000	-0.0523763	0.000	-0.0534639	0.000	-0.0571822	0.000	-0.0703081	0.000
Pseudo R2		**0.1232**		**0.1215**		**0.1211**		**0.1200**		**0.1182**
Nr Obs		**3,809,315**		**3,809,315**		**3,809,315**		**3,809,315**		**3,809,315**

Table 6: Black Employment Probit (Marginal Effects)

Variable Name	No Adjustment		10% Adjustment		20% Adjustment		40% Adjustment		80% Adjustment	
	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|	dF/dx	P > |z|
Urban dummy	-0.1951064	0.000	-0.1962081	0.000	-0.1951322	0.000	-0.1922920	0.000	-0.2125826	0.000
Education Spline: None-Std 5	-0.0159394	0.000	-0.0143817	0.000	-0.0162078	0.000	-0.0215111	0.000	-0.0724271	0.000
Education Spline: Std 6-10	-0.0133302	0.000	-0.0131251	0.000	-0.0147705	0.000	-0.0195844	0.000	-0.0840545	0.000
Education Spline: Tertiary	0.0087901	0.000	0.0074876	0.000	0.0085438	0.000	0.0119270	0.000	0.0431737	0.000
Age: 26-35	0.0418186	0.000	0.0399712	0.000	0.0402187	0.000	0.0407905	0.000	0.0208611	0.006
Age: 36-45	0.1847909	0.000	0.1833139	0.000	0.1835146	0.000	0.1840541	0.000	0.1716868	0.000
Age: 46-55	0.2890619	0.000	0.2882656	0.000	0.2882991	0.000	0.2884945	0.000	0.2912086	0.000
Age: 56-65	0.3729297	0.000	0.3728616	0.000	0.3727851	0.000	0.3726208	0.000	0.3776515	0.000
Female dummy	0.2439315	0.000	0.2482853	0.000	0.2474029	0.000	0.2452434	0.000	0.2882231	0.000
Eastern Cape dummy	-0.1908544	0.000	-0.1907459	0.000	-0.1909287	0.000	-0.1914016	0.000	-0.1835885	0.000
Northern Cape dummy	-0.0647498	0.002	-0.0651128	0.002	-0.0650181	0.002	-0.0646886	0.002	-0.0651096	0.002
Free State dummy	0.0301587	0.036	0.0300224	0.036	0.0299571	0.037	0.0299094	0.037	0.0433886	0.003
Kwazulu Natal dummy	-0.0602708	0.000	-0.0601303	0.000	-0.0604799	0.000	-0.0615457	0.000	-0.0738692	0.000
North-West dummy	-0.0819959	0.000	-0.0820358	0.000	-0.0821676	0.000	-0.0825514	0.000	-0.0832841	0.000
Gauteng dummy	-0.0135812	0.355	-0.0135642	0.356	-0.0137029	0.351	-0.0141418	0.336	-0.0218538	0.137
Mpumalanga dummy	-0.0250736	0.087	-0.0248756	0.090	-0.0248610	0.090	-0.0246943	0.092	-0.0159487	0.283
Northern Province dummy	-0.1659315	0.000	-0.1653655	0.000	-0.1649738	0.000	-0.1635995	0.000	-0.1378546	0.000
Lambda	-1.2407390	0.000	-1.2539410	0.000	-1.2512830	0.000	-1.2447500	0.000	-1.3778870	0.000
Pseudo R2		**0.1561**		**0.1565**		**0.1565**		**0.1565**		**0.1574**
Nr Obs		**21,536,895**		**21,536,895**		**21,536,895**		**21,536,895**		**21,536,895**

3.4. Estimating the Earnings Function

In the third step, earnings functions were fitted on the sample of wage earning individuals using the lambdas from the employment probit (also indirectly incorporating those calculated in the participation probit) to control for selectivity. All the lambdas were, once again, significant. The results are shown in Tables 8 and 9.

The basic earnings function regresses education and experience variables on earnings. Hours worked was controlled for by including a proxy of weekly hours worked. In addition, dummies were included to capture the effect of urban and provincial location as well as gender and union membership. To take account of systematic differences in earnings among different sectors and occupations, a number of sectoral and occupational dummies were included in the regression. Union membership was included as a dummy variable.

The urban dummy was positive and significant[12] for the black population but insignificant for the white population, indicating that there were significant rural/urban earnings differentials for blacks but not for whites when standardising for other characteristics. For blacks, all the education splines were positive and significant with the expected increasing returns to higher levels of education. The estimated returns to primary, secondary and tertiary education were 3.2%, 4.9% and 5.3% respectively before adjusting for education quality (Table 7). With the 40% adjustment, this increased to 3.5%, 6.5% and 7.4% respectively. For the 80% adjustment level, the results showed a negative return for primary schooling as well as a big increase to the returns for tertiary education (the level of significance, however, dropped to 10% for the latter), but 80% may be an excessive adjustment in calculating effective education and should mainly be interpreted as part of the sensitivity tests. For whites, the primary education spline was insignificant. The secondary and tertiary splines were positive and significant and estimated returns were at 5.4% and 5.8% respectively before controlling for quality.

The results in Table 7 show systematic changes in returns across different levels of adjustment, with returns to lower levels of education systematically reduced by quality adjustments and returns to higher levels of education tend to increase. Returns to tertiary education may usually be obscured by the varying impact of quality of education.

---

[12] The Heckman procedure in STATA does not have survey options that take account of frequency weighting in the calculation of the t-statistics (as it does for the probit functions). The artificially large sample size created by weighting, therefore, causes all the t-statistics to be significant. In an attempt to get some idea of the significance of the coefficients, the t-statistics were also calculated for the unweighted samples. These t-statistics were included in the regression tables and will be used to provide some indication of the significance of the results.

Table 7 Estimated Returns to Education

	No Adjustment	10% Adjustment	20% Adjustment	40% Adjustment	80% Adjustment
Whites					
Primary Education	-0.80%	-1.29%	-1.91%	-4.09%	-16.42%
Secondary Education	5.35%**	4.54%**	4.78%**	5.21%**	-1.33%**
Tertiary Education	5.79%**	5.22%**	5.56%**	6.30%**	14.64%**
Blacks					
Primary Education	3.19%**	2.29%**	2.60%**	1.53%**	-4.33%**
Secondary Education	4.90%**	4.28%**	4.83%**	3.54%**	0.50%*
Tertiary Education	5.29%**	4.92%**	5.54%**	6.45%**	18.06%**

** significant at 1% level *significant at 10% level

For blacks, after standardising for other factors, location dummies were insignificant for the Eastern Cape and North West; Northern Cape and Free State had lower levels of earnings than the Western Cape; and Kwazulu-Natal, Gauteng, Mpumalanga and the Northern Province (Limpopo) had higher earnings. The results were consistent for the different levels of adjustment. For whites, dummy variables for the Eastern Cape, Free State and Mpumalanga were insignificant, Northern Cape had lower earnings than the Western Cape, and Kwazulu-Natal, North West, Gauteng and Northern Provinces had higher earnings. Once again, the results were stable across adjustment levels.

Most sectoral dummies showed positive significant coefficients relative to the reference sector (Agriculture). There were no significant changes across levels of adjustment. The situation was the same for occupational dummies with most coefficients both positive and significant relative to the control occupation, unskilled labour.

As expected, the female dummies gave significantly negative coefficients for both whites and blacks, with much larger gender wage differentials for whites. Although, therefore, there were little difference found in the employment of white males and females, there seems to be some gender discrimination in wage setting, a finding supported by Winter (1998). Union membership was also found to be significant, with the positive coefficient for blacks significantly larger than that for whites. The wage differential between union and non-union members was estimated at 23% for blacks, constant across all adjustment levels. This is similar to the results of Bhorat & Leibbrandt (2001b) and Hofmeyr & Lucas (1998).

The results on experience and hours worked were as expected, with experience significant and positive, experience squared significant and negative, and hours worked significant and positive. The coefficient for experience was larger for whites, with the increase in earnings for an additional year of experience at 6% compared to 1.5% for blacks. The coefficient for the log of hours worked was also found to be much larger of whites than for blacks.[13]

[13] The coefficient for blacks seems to be in line with the estimates of Bhorat & Leibbrandt (2001b). The white coefficient, however, seems unreasonably large.

Table 8: White Earnings Regression

Variable Name	No Adjustment			10% Adjustment			20% Adjustment			40% Adjustment			80% Adjustment		
	Coeff.	P > |z| (weighted)	P > |z| (unweighted)	Coeff.	P > |z| (weighted)	P > |z| (unweighted)	Coeff.	P > |z| (weighted)	P > |z| (unweighted)	dF/dx	P > |z| (weighted)	P > |z| (unweighted)	dF/dx	P > |z| (weighted)	P > |z| (unweighted)
Urban dummy	-0.06531270	0.000	0.321	-0.057077700	0.000	0.185	-0.0476491	0.000	0.094	-0.0258204	0.000	0.015	0.1280650	0.000	0.000
Education Spline: None-Std 5	-0.00802600	0.000	0.535	-0.012901900	0.000	0.784	-0.0190995	0.000	0.862	-0.0408845	0.000	0.907	-0.1642754	0.000	0.571
Education Spline: Std 6-10	0.05351880	0.000	0.000	0.045350000	0.000	0.000	0.0477695	0.000	0.000	0.0521424	0.000	0.000	-0.0133242	0.000	0.004
Education Spline: Tertiary	0.05790680	0.000	0.000	0.052167400	0.000	0.000	0.0555780	0.000	0.000	0.0629812	0.000	0.000	0.1463774	0.000	0.000
Location dummy: Eastern Cape	-0.02529490	0.000	0.152	-0.024528900	0.000	0.162	-0.0240051	0.000	0.168	-0.0227226	0.000	0.183	-0.0118634	0.000	0.331
Location dummy: Northern Cape	-0.18664290	0.000	0.000	-0.186882900	0.000	0.000	-0.1871577	0.000	0.000	-0.1879561	0.000	0.000	-0.1919100	0.000	0.000
Location dummy: Free State	0.00553130	0.001	0.539	0.007315600	0.000	0.572	0.0079761	0.000	0.584	0.0095471	0.000	0.613	0.0158643	0.000	0.717
Location dummy: Kwazulu Natal	0.07138040	0.000	0.091	0.063862700	0.000	0.147	0.0542624	0.000	0.257	0.0319864	0.000	0.677	-0.1369757	0.000	0.000
Location dummy: North-West	0.11717090	0.000	0.000	0.117340200	0.000	0.000	0.1174815	0.000	0.000	0.1178442	0.000	0.000	0.1176471	0.000	0.000
Location dummy: Gauteng	0.18574350	0.000	0.000	0.186665500	0.000	0.000	0.1870040	0.000	0.000	0.1877159	0.000	0.000	0.1912345	0.000	0.000
Location dummy: Mpumalanga	0.04946650	0.000	0.506	0.067510100	0.000	0.230	0.0854970	0.000	0.089	0.1271017	0.000	0.006	0.4198330	0.000	0.000
Location dummy: Northern Province	0.14309600	0.000	0.005	0.142014700	0.000	0.005	0.1419313	0.000	0.005	0.1415034	0.000	0.006	0.1447957	0.000	0.005
Sector dummy: Mining	0.29923370	0.000	0.000	0.299740900	0.000	0.000	0.2997116	0.000	0.000	0.2997553	0.000	0.000	0.2924826	0.000	0.000
Sector dummy: Manufacturing	0.18193230	0.000	0.007	0.181570200	0.000	0.007	0.1816127	0.000	0.007	0.1818465	0.000	0.006	0.1747116	0.000	0.010
Sector dummy: Electricity	0.18251940	0.000	0.005	0.184906400	0.000	0.005	0.1857324	0.000	0.005	0.1881159	0.000	0.004	0.1858825	0.000	0.007
Sector dummy: Construction	0.15239020	0.000	0.002	0.151368300	0.000	0.002	0.1513026	0.000	0.002	0.1512424	0.000	0.002	0.1404127	0.000	0.007
Sector dummy: Wholesale/Retail	0.08667660	0.000	0.148	0.085449700	0.000	0.150	0.0853009	0.000	0.151	0.0850122	0.000	0.153	0.0776920	0.000	0.196
Sector dummy: Transport	0.19956920	0.000	0.011	0.196982700	0.000	0.012	0.1969520	0.000	0.012	0.1967383	0.000	0.012	0.1951290	0.000	0.014
Sector dummy: Financial	0.20136190	0.000	0.001	0.202100200	0.000	0.001	0.2024068	0.000	0.001	0.2032632	0.000	0.001	0.1953977	0.000	0.001
Sector dummy: Community Services	0.05031430	0.000	0.590	0.050331800	0.000	0.583	0.0504038	0.000	0.583	0.0507076	0.000	0.583	0.0415387	0.000	0.707
Occupation dummy: Managers	0.82969690	0.000	0.000	0.840444900	0.000	0.000	0.8425119	0.000	0.000	0.8481248	0.000	0.000	0.8363383	0.000	0.000
Occupation dummy: Professional	0.72922040	0.000	0.000	0.741583500	0.000	0.000	0.7446978	0.000	0.000	0.7534838	0.000	0.000	0.7399171	0.000	0.000
Occupation dummy: Technicians	0.47977410	0.000	0.000	0.489552900	0.000	0.000	0.4914149	0.000	0.000	0.4964899	0.000	0.000	0.4829436	0.000	0.000
Occupation dummy: Clerks	0.16788830	0.000	0.000	0.176834500	0.000	0.000	0.1781108	0.000	0.000	0.1816141	0.000	0.000	0.1663220	0.000	0.000
Occupation dummy: Service & Sales	0.05323210	0.000	0.004	0.060688500	0.000	0.002	0.0618871	0.000	0.002	0.0650031	0.000	0.002	0.0509766	0.000	0.004
Occupation dummy: Skilled Agriculture	0.98566960	0.000	0.000	0.995397900	0.000	0.000	0.9973569	0.000	0.000	1.0026350	0.000	0.000	0.9737015	0.000	0.000
Occupation dummy: Craft	0.19952460	0.000	0.000	0.205550200	0.000	0.000	0.2061441	0.000	0.000	0.2076222	0.000	0.000	0.1886226	0.000	0.000
Occupation dummy: Machine Operator	-0.01174800	0.000	0.556	-0.011036300	0.000	0.538	-0.0112608	0.000	0.537	-0.0120063	0.000	0.536	-0.0290372	0.000	0.675
Female dummy	-0.47115770	0.000	0.000	-0.470702600	0.000	0.000	-0.4702457	0.000	0.000	-0.4690589	0.000	0.000	-0.4595118	0.000	0.000
Union Membership dummy	0.03871250	0.000	0.001	0.037231000	0.000	0.001	0.0370573	0.000	0.001	0.0365335	0.000	0.001	0.0414569	0.000	0.000
Experience	0.06024010	0.000	0.000	0.059908000	0.000	0.000	0.0595941	0.000	0.000	0.0588304	0.000	0.000	0.0537397	0.000	0.000
Experience Squared	-0.00111050	0.000	0.000	-0.001111200	0.000	0.000	-0.0011083	0.000	0.000	-0.0011013	0.000	0.000	-0.0010391	0.000	0.000
Log of Hours Worked Previous Week	0.50947450	0.000	0.000	0.510088800	0.000	0.000	0.5100225	0.000	0.000	0.5097780	0.000	0.000	0.5111184	0.000	0.000
Constant	4.83620900	0.000	0.000	4.854620000	0.000	0.000	4.8698560	0.000	0.000	4.9168740	0.000	0.000	4.7521140	0.000	0.000
Lambda	-0.18289270	0.002	0.037	-0.190648900	0.002	0.036	-0.1944817	0.002	0.035	-0.2061448	0.002	0.034	-0.2566143	0.002	0.028
Chi squared		**2,710,000**	**7,436**		**2,710,000**	**7,436**		**2,710,000**	**7,440**		**2,720,000**	**7,449**		**2,740,000**	**7,541**
Level of significance		**1%**	**1%**		**1%**	**1%**		**1%**	**1%**		**1%**	**1%**		**1%**	**1%**
No obs		**3,809,315**	**10,884**		**3,809,315**	**10,884**		**3,809,315**	**10,884**		**3,809,315**	**10,884**		**3,809,315**	**10,884**

Table 9: Black Earnings Regression

	No Adjustment			10% Adjustment			20% Adjustment			40% Adjustment			80% Adjustment		
Variable Name	**Coeff.**	**P > |z| (weighted)**	**P > |z| (unweighted)**	**Coeff.**	**P > |z| (weighted)**	**P > |z| (unweighted)**	**Coeff.**	**P > |z| (weighted)**	**P > |z| (unweighted)**	**dF/dx**	**P > |z| (weighted)**	**P > |z| (unweighted)**	**dF/dx**	**P > |z| (weighted)**	**P > |z| (unweighted)**
Urban dummy	0.1565470	0.000	0.000	0.157474	0.000	0.000	0.1562026	0.000	0.000	0.1525401	0.000	0.000	0.1264351	0.000	0.000
Education Spline: None-Std 5	0.0319263	0.000	0.000	0.022864	0.000	0.000	0.0259693	0.000	0.000	0.0354479	0.000	0.000	-0.0432751	0.000	0.000
Education Spline: Std 6-10	0.0490203	0.000	0.000	0.042830	0.000	0.000	0.0482904	0.000	0.000	0.0645505	0.000	0.000	0.0049611	0.000	0.077
Education Spline: Tertiary	0.0528786	0.000	0.000	0.049235	0.000	0.000	0.0554382	0.000	0.000	0.073716	0.000	0.000	0.1805651	0.000	0.000
Location dummy: Eastern Cape	0.0032247	0.001	0.910	0.008797	0.000	0.740	0.0105243	0.000	0.685	0.0161376	0.000	0.517	0.076785	0.000	0.001
Location dummy: Northern Cape	-0.1024636	0.002	0.001	-0.102313	0.000	0.001	-0.1016897	0.000	0.001	-0.099856	0.000	0.001	-0.0993869	0.000	0.001
Location dummy: Free State	-0.4359893	0.001	0.000	-0.432039	0.000	0.000	-0.4273569	0.000	0.000	-0.41321	0.000	0.000	-0.29856	0.000	0.000
Location dummy: Kwazulu Natal	0.1353722	0.001	0.000	0.136321	0.000	0.000	0.1336077	0.000	0.000	0.1256659	0.000	0.000	0.0681678	0.000	0.000
Location dummy: North-West	0.0278347	0.001	0.448	0.030010	0.000	0.399	0.0299433	0.000	0.400	0.0299373	0.000	0.399	0.0368613	0.000	0.214
Location dummy: Gauteng	0.1714394	0.001	0.000	0.173083	0.000	0.000	0.1727959	0.000	0.000	0.1720994	0.000	0.000	0.1673166	0.000	0.000
Location dummy: Mpumalanga	0.0862275	0.001	0.000	0.091356	0.000	0.000	0.0949755	0.000	0.000	0.1059753	0.000	0.000	0.1951078	0.000	0.000
Location dummy: Northern Province	0.3252911	0.001	0.000	0.333454	0.000	0.000	0.3391574	0.000	0.000	0.3566148	0.000	0.000	0.4953164	0.000	0.000
Sector dummy: Mining	0.4307004	0.001	0.000	0.432920	0.000	0.000	0.4334921	0.000	0.000	0.4353461	0.000	0.000	0.4485429	0.000	0.000
Sector dummy: Manufacturing	0.4285869	0.001	0.000	0.431894	0.000	0.000	0.4323535	0.000	0.000	0.4340542	0.000	0.000	0.454015	0.000	0.000
Sector dummy: Electricity	0.7386162	0.002	0.000	0.744906	0.000	0.000	0.7451535	0.000	0.000	0.746129	0.000	0.000	0.7366639	0.000	0.000
Sector dummy: Construction	0.3434288	0.001	0.000	0.344165	0.000	0.000	0.344606	0.000	0.000	0.3460679	0.000	0.000	0.3510611	0.000	0.000
Sector dummy: Wholesale/Retail	0.3325891	0.001	0.000	0.335235	0.000	0.000	0.3359767	0.000	0.000	0.3384749	0.000	0.000	0.3651559	0.000	0.000
Sector dummy: Transport	0.4886138	0.001	0.000	0.490189	0.000	0.000	0.4907752	0.000	0.000	0.4927246	0.000	0.000	0.5113812	0.000	0.000
Sector dummy: Financial	0.4030259	0.001	0.000	0.408134	0.000	0.000	0.4084468	0.000	0.000	0.4097856	0.000	0.000	0.4272107	0.000	0.000
Sector dummy: Community Services	0.3221877	0.001	0.000	0.325796	0.000	0.000	0.3264631	0.000	0.000	0.3287774	0.000	0.000	0.3476379	0.000	0.000
Occupation dummy: Managers	1.0933800	0.001	0.000	1.100236	0.000	0.000	1.1002040	0.000	0.000	1.1006040	0.000	0.000	1.1159640	0.000	0.000
Occupation dummy: Professional	1.0003180	0.002	0.000	1.014441	0.000	0.000	1.0159120	0.000	0.000	1.0219590	0.000	0.000	1.1489820	0.000	0.000
Occupation dummy: Technicians	0.8847490	0.001	0.000	0.896974	0.000	0.000	0.8976478	0.000	0.000	0.9006622	0.000	0.000	0.9650427	0.000	0.000
Occupation dummy: Clerks	0.5299700	0.001	0.000	0.540326	0.000	0.000	0.5403812	0.000	0.000	0.5409611	0.000	0.000	0.5484065	0.000	0.000
Occupation dummy: Service & Sales	0.3734874	0.001	0.000	0.378894	0.000	0.000	0.3790083	0.000	0.000	0.3796162	0.000	0.000	0.3844129	0.000	0.000
Occupation dummy: Skilled Agriculture	0.8639589	0.002	0.000	0.870109	0.000	0.000	0.8711509	0.000	0.000	0.8746694	0.000	0.000	0.921291	0.000	0.000
Occupation dummy: Craft	0.3213645	0.001	0.000	0.324061	0.000	0.000	0.3242275	0.000	0.000	0.3248684	0.000	0.000	0.3334334	0.000	0.000
Occupation dummy: Machine Operator	0.2567055	0.001	0.000	0.259113	0.000	0.000	0.2590998	0.000	0.000	0.2591947	0.000	0.000	0.263067	0.000	0.000
Female dummy	-0.2720662	0.000	0.000	-0.271796	0.000	0.000	-0.2720363	0.000	0.000	-0.2726017	0.000	0.000	-0.269928	0.000	0.000
Union Membership dummy	0.2360785	0.000	0.000	0.237457	0.000	0.000	0.237708	0.000	0.000	0.2385641	0.000	0.000	0.2507096	0.000	0.000
Experience	0.0158135	0.000	0.000	0.014890	0.000	0.000	0.0148882	0.000	0.000	0.0148204	0.000	0.000	0.0119872	0.000	0.000
Experience Squared	-0.0001932	0.000	0.000	-0.000184	0.000	0.000	-0.0001838	0.000	0.000	-0.0001840	0.000	0.000	-0.0001574	0.000	0.000
Log of Hours Worked Previous Week	0.0708547	0.001	0.000	0.070090	0.000	0.000	0.0698643	0.000	0.000	0.0691369	0.000	0.000	0.0619132	0.000	0.000
Constant	5.5583200	0.002	0.000	5.574335	0.000	0.000	5.574098	0.000	0.000	5.575644	0.000	0.000	5.646351	0.000	0.000
Lambda	-0.3593964	0.001	0.011	-0.364952	0.001	0.011	-0.365	0.001	0.011	-0.367	0.001	0.011	-0.427	0.001	0.009
Chi squared		**12,100,000**	**40,153**		**12,000,000**	**39,935**		**12,000,000**	**39,904**		**12,000,000**	**39,786**		**11,700,000**	**38,426**
Level of significance		**1%**	**1%**		**1%**	**1%**		**1%**	**1%**		**1%**	**1%**		**1%**	**1%**
No obs		**21,500,000**	**66,240**		**21,500,000**	**66,240**		**21,500,000**	**66,240**		**21,500,000**	**66,240**		**21,500,000**	**66,240**

3.5. Estimating Discrimination

In 1973, Oaxaca provided a methodology to decompose the wage gap and to estimate the levels of discrimination in the labour market. Oaxaca's decomposition is often used to estimate levels of gender discrimination[14] and in this study will be applied to estimate the level of discrimination between blacks and whites.

After estimating earnings functions for blacks and whites:

$$\ln Y_w = X_w b_w \qquad \text{(a)} \quad 15$$

$$\ln Y_b = X_b b_b \qquad \text{(b)}$$

where

Y_b and Y_w are black and white levels of earnings respectively,

X_b and X_w are vectors of regressors for black and white earnings functions and

b_b and b_w are vectors of parameters for black and white earnings functions,

the difference between black and white earnings (wage gap) can then be defined as:

$$\ln Y_w - \ln Y_b = X_w b_w - X_b b_b \qquad \text{(c)}$$

Using (c), the decomposition can be done relative to black or white earnings levels. The choice depends on an assumption on which wage level will prevail in the absence of discrimination. There is no theoretical basis on which to decide this – the 'index number problem'.[16] The results will thus rather be shown relative to both white and black earnings. The first expresses, as a proportion of the wage gap, the difference between what blacks are currently earning and what they would have earned *had their characteristics been rewarded similar to whites*. The second expresses, also as a proportion of the wage gap, the difference between what whites are currently earning and what they would have earned *had they received black rewards for their characteristics.*

---

[14] Psacharapoulos & Tzannatos (1992) contains several examples of this type of application.

[15] In order to simplify the discussion we ignore the error terms.

[16] In his initial article, Oaxaca opted to use both these methods and interpreted the two answers as upper and lower bounds for discrimination. In later work (Oaxaca & Ransom, 1994: 8), he indicated that this was an incorrect interpretation and introduced a neutral benchmark (a non-discriminatory wage level) to get rid of the index number problem and distinguish between positive and negative discrimination.

The original method (without the neutral benchmark) will still be used in this study: Firstly, the relative level of discrimination between the black and white population groups is of particular interest. Secondly, the latter method is not without criticism, as it requires quite strong assumptions and adds much complexity to the analysis.

Firstly then, the decomposition relative to black earnings levels:
By adding and subtracting $X_b b_w$ to/from (c) and re-arranging the following equation is derived:

$$\ln Y_w - \ln Y_b = b_w\left(X_w - X_b\right) + X_b\left(b_w - b_b\right) \qquad \text{(d)}$$

The first term can be interpreted as measuring the actual differences in endowment levels between the two groups and the second term as measuring the difference in the market evaluation of the same endowments. The latter term is used as an estimate of discrimination in the market place. The proportion of the wage gap ascribed to discrimination can then be expressed as:

$$\text{\% Discrimination} = \frac{X_b\left(b_w - b_b\right)}{\ln\left(Y_w\right) - \ln\left(Y_b\right)} \times 100 \qquad \text{(e)}$$

A similar decomposition can be done relative to white earnings levels.

Based on the earnings functions estimated above, the decomposition of the wage gap across adjustment levels is as shown in Table 10:

Table 10: Oaxaca's Decomposition of Wage Gap (difference in log of wage)

Level of Adjustment	Based on white means		Based on black means	
	Discrimination	Characteristics	Discrimination	Characteristics
None	42.23%	57.77%	53.61%	46.39%
10%	42.03%	57.97%	53.02%	46.98%
20%	41.23%	58.77%	52.41%	47.59%
40%	38.86%	61.14%	51.23%	48.77%
80%	23.63%	76.37%	36.64%	63.36%

Looking at the calculations based on white means, the discrimination component of the wage gap (i.e. without considering educational quality) is estimated to be 42%, cannot be ascribed to the fact that blacks characteristics are less favourable than those of whites. This is conventionally regarded as labour market discrimination. Adjusting education for quality, the discrimination component of the wage gap is systematically reduced from 42% to 24% with increased levels of adjustment for educational quality. Almost half of the component previously ascribed to labour market discrimination, therefore, could perhaps be due to pre-labour market discrimination. A significant component ascribed to labour market discrimination, however, remains.

These results can also be presented as in Table 11. Based on white mean characteristic as the standard, the wage situation can be expressed thus: Mean black wage per month was R1 689, compared to a mean white wage of R6 989. Without adjusting for education quality, black wages would have been R1 589 + R2 150 = R3 739 if blacks had had white characteristics. This leaves an unexplained shortfall (discrimination) between white and black wages of R3 250, or 46.5% of white levels. Accounting for the quality of schooling reduces this discriminatory element only marginally to 43.8% if a 40% weight is given to educational quality. However, when the weight given to educational quality rises to 80%, the residual is reduced substantially to only 29.5% of the white wage.

Table 11: Size of the residual: Effect of applying white characteristics to the regression for the black population

Level of a educational quality	Black wage	accounted for by wh characteristics	unaccounted for even with white characteristics		residual as % of white wage
0%	R1 589	R2 150	R3 250	**R6 989**	**46.5%**
10	R1 589	R2 161	R3 239	R6 989	**46.3%**
20%	R1 589	R2 206	R3 194	R6 989	**45.7%**
40%	R1 589	R2 341	R3 059	R6 989	**43.8%**
80%	R1 589	R3 336	R2 064	R6 989	**29.5%**

These results strongly support our main hypothesis. Even accepting that the results are based on rather tentative estimates of quality, the indications are strong and systematic enough that differences in quality of education received do in fact explain a significant part of the wage gap previously only attributed to labour market discrimination. But it is important to note that the effect of adjustments for quality of education are not very large, *unless the weight given to quality becomes quite large* (80%) – something which some may find less credible. In addition, the results confirm that even after adjusting for differentials in the quality of education, in 1995 labour market discrimination still accounted for a significant component of wage differentials.

4. Conclusion

The initial premise of this study was that discrimination is a much more complicated phenomenon and operates on more levels of society than is commonly accounted for. Based on this, it was attempted to test whether controlling for quality affects the usual estimates of discrimination and returns to education in a systematic manner.

The main obstacle to an analysis such as this is certainly the absence of clear, tested and standardised measures of quality, as exposed by its conspicuous absence from most earnings functions analyses. At best, it is usually only mentioned in a last paragraph on shortcomings and suggestions for future research. Within the known limitations of the data currently available, this obstacle was addressed in an admittedly simplistic manner that cannot be a perfect rendition of the elusive quality. Yet it certainly provides good first estimates both of quality and of its effect on estimated earnings functions and discrimination in South Africa.

It was shown that it was necessary to control for sample selection bias (coefficients on the inverse Mill's ratio were all significant). The three-phased approach was also vindicated by evidence that showed that the same variables operated differently in each of the three phases, thus providing a more complete picture of the labour market process. For the participation and employment probits, most variables showed coefficients in line with theory and previous studies, and allowing for an increased influence of educational quality changed the coefficients in a systematic and expected manner. Where results differed from previous studies, it could be explained by the more specific approach followed in this study using the three-phased approach. The returns to education in the final earnings regression were also as expected, with higher returns to higher levels of education. In the case of blacks, all splines showed positive returns to education. For whites, the coefficient for primary education was negative but not significant. Although this requires more research, it may be interpreted as support for the hypothesis of Mwabu and Schultz (1996) that the returns to a specific level of education decreases as the proportion so educated increases.

Repeating the exercise before and after controlling for quality showed two things: Firstly, the consistency of the results across different levels of adjustment showed that the models were robust. Secondly, increasing the influence of quality systematically decreased the component of the wage gap ascribed to labour market discrimination. By controlling for quality, almost half the proportion of the wage gap previously ascribed to labour market discrimination could be explained by pre-labour market discrimination. A significant proportion ascribed to labour market discrimination, however, remained.

This study, therefore, contributes to our body of knowledge by virtue of both its strengths and weaknesses. Acknowledging that there is much room for improvement, it provided an interesting view of the labour market process, confirmed the general results found in similar studies and showed some first evidence on the impact of quality of education on labour market earnings in South Africa. Most importantly, it showed that it is necessary to rethink and revisit the estimates of labour market discrimination and to control for quality of education when doing so.

References:

Aldrich, J.H. & Nelson, F.D., 1984. *Linear Probability, Logit, and Probit Models.* Sage University Paper. Sage Publications: London.

Archer, S., 1995. *Quality and Efficiency Indicators in Education: A Survey.* EDUPOL Research Report.

Behrman, J.R. & Birdsall, N., 1983. The Quality of Schooling: Quantity Alone Is Misleading. *American Economic Review.* Vol. 73(5): 928 – 946.

Bhorat, H. & Leibbrandt, M. 2001a. Correlates of Vulnerability in the South African Labour Market. Chapter 2 in: Bhorat, Haroon; Leibbrandt, Murray; Maziya, Muzi; Van der Berg, Servaas; & Woolard, Ingrid (Eds.). 2001. *Fighting poverty: Labour Markets and Inequality in South Africa.* UCT Press: Lansdowne: 171-106.

Bhorat, H. & Leibbrandt, M. 2001b. Modelling Vulnerability and Low Earnings in the South African Labour Market. Chapter 3 in. Bhorat, Haroon; Leibbrandt, Murray; Maziya, Muzi; Van der Berg, Servaas; & Woolard, Ingrid (Eds.). 2001. *Fighting poverty: Labour Markets and Inequality in South Africa.* UCT Press: Lansdowne: 107-129.

Breen, R., 1996. Regression Models: Censored, Sample-Selected, or Truncated Data. Sage Publications: London.

Fallon, P. & Lucas, R., 1998. *South African Labor Markets Adjustment and Inequalities.* Informal Discussion Papers on Aspects of the Economy of South Africa. World Bank: Washington D.C.

Fallon, P. & Verry, D., 1988. *The Economics of Labour Markets.* Peter Allan: New York.

Fuller, B., Pillay, P. & Sirur, N., 1995. *Literacy Trends in South Africa: Expanding Education while Reinforcing Unequal Achievement?* LSDS Working Paper. University of Cape Town: Cape Town.

Heckman, J.J., 1979. Sample Selection Bias as a Specification Error. *Econometrica*. Vol. 47(1): 153 – 161.

Heckman, J., Layne-Farrar, A. & Todd, P., 1996. Human Capital Pricing Equations with an Application to Estimating the Effect of Schooling Quality on Earnings. *Review of Economics and Statistics.* Vol. 78(4): 562 – 610.

Hofmeyr, J.F. & Lucas, R., 1998. *The Rise in Union Wage Premia in South Africa.* IED Discussion Paper Series No. 83. Institute for Economic Development: Boston University.

Johnson, G.E. & Stafford, F.P., 1996. On the Rate of return to Schooling Quality. *Review of Economics and Statistics.* Vol. 78(4): 686 – 691.

Kingdon, Geeta & John Knight. 2002. *Quality of schooling and the race gap in labour market outcomes in South Africa.* Paper to the Conference of the Centre for the Study of African Economies: Oxford. March.

Lee, J. & Barro, R.J.,1997. *Schooling Quality in a Cross Section of Countries.* NBER Working Paper 6198: Cambridge, Mass.

Long, J.S., 1997. *Regression Models for Categorical and Limited Dependent Variables.* Sage Publications: London.

Mincer, J., 1974. *Schooling, Experience and Earnings.* Columbia University Press: New York.

Moffitt, R.A., 1996. Symposium on School Quality and Educational Outcomes: Introduction. *Review of Economics and Statistics.* Vol. 78(4): 559 – 561.

Moll, P.G., 1996. The Collapse of Primary Schooling Returns in South Africa 1960 – 1990. *Oxford Bulletin of Economics and Statistics.* Vol. 58(1): 185 – 209.

Moll, P.G., 1998. Primary Schooling, Cognitive Skills and Wages in South Africa. *Economica.* Vol. 65: 263-284.

Mwabu, G. & Schultz, T.P., 1996. Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa. *American Economic Review.* Vol. 86(2): 330 – 334.

Oaxaca, R.L., 1973. Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review.* Vol. 14(3): 693 – 709.

Oaxaca, R.L. & Ransom, M.R., 1994. On Discrimination and the Decomposition of Wage Differentials. *Journal of Econometrics.* Vol. 61: 5-21.

Psacharopoulos, G. & Tzannatos, Z. (Eds.),1992. *Case Studies on Women's Employment and Pay in Latin America.* World Bank: Washington D.C.

SALDRU, 1994. *South Africans Rich and the Poor: Baseline Household Statistics.* School of Economics, University of Cape Town: Cape Town.

Van der Berg, S., Wood, L. & Le Roux, N., 2002. Differentiation in Black Education. *Development Southern Africa* Vol. 18(3).

Winter, C., 1998. Women Workers in South Africa: Participation, Pay and Prejudice in the Formal Labour Market. Mimeo. World Bank: Washington D.C.