



Munich Personal RePEc Archive

# **Education, Employment and Earnings of Secondary School and University Leavers in Tanzania: Evidence from a Tracer Study**

Al-Samarrai, Samer and Reilly, Barry

2006

Online at <https://mpra.ub.uni-muenchen.de/129/>

MPRA Paper No. 129, posted 06 Oct 2006 UTC

# **Education, Employment and Earnings of Secondary School and University Leavers in Tanzania: Evidence from a Tracer Study\***

**Samer Al-Samarrai \*\***  
**Institute of Development Studies**  
**University of Sussex**  
**Falmer**  
**Brighton**  
**BN1 9RE**  
**United Kingdom**  
**e-mail: alsamarrai@msbx.net**

**Barry Reilly**  
**Department of Economics**  
**University of Sussex**  
**Falmer**  
**Brighton**  
**BN1 9SN**  
**United Kingdom**

May 2006

Word count: 8545

\* The data used in this survey are drawn from a wider project undertaken in four Sub-Saharan African countries funded by the United Kingdom's Department for International Development (DfID). Faustin Mukyanuzi of HR Consult implemented the survey in Tanzania and overall findings are reported in Mukyanuzi (2003). The views expressed in this paper are entirely those of the authors and do not necessarily represent the policies or views of DfID. The authors are grateful to Lawrence Haddad and two anonymous referees of this journal for constructive comments on earlier drafts but remain entirely responsible for any remaining errors.

\*\* Corresponding author.

# Education, Employment and Earnings of Secondary School and University Leavers in Tanzania: Evidence from a Tracer Study

## ABSTRACT

*The empirical evidence on the earnings of educated groups in Tanzania is limited. This study uses a recently completed tracer survey of secondary school completers to analyse the impact of educational qualifications on labour market earnings. Our findings suggest that the rates of return to the highest educational qualifications for wage employees are not negligible and, at the margin, provide an investment incentive. However, we find little evidence of human capital effects in the earnings determination process for the self-employment sector. Information contained in the tracer survey allowed the introduction of controls for father's educational background and a set of school fixed effects designed to proxy for school quality and potential labour market network effects. Our analysis reveals that the inclusion of these controls in the earnings determination process is important and tends to reduce the estimated rates of return to educational qualifications. A comparison of our results with the available evidence from other countries in the region suggest that despite an extremely small secondary and university education system the private rates of return to education in the Tanzanian wage employment sector are comparatively low.*

# 1. Introduction

The empirical evidence on labour market outcomes in Sub-Saharan Africa is limited. Government and individuals invest heavily in education and training, but little is generally known for this region about the extent to which these investments generate rewards in the labour market. This is particularly so for individuals who fail to gain access to wage employment and are required to rely on self-employment opportunities.

Earnings equations provide a convenient framework within which the average private rate of return to an educational qualification can be computed. The computation of returns to education is important as it provides information on the benefits associated with investments made by both governments and individuals in education. They also provide policy-makers with important information to guide education policy. However, there is a dearth of evidence on the magnitude of these returns for wage employees in Tanzania and almost none for individuals engaged in self-employment activity. Given the increasing number of secondary school-leavers entering self-employment, this represents a substantial lacuna in the literature. The aim of this paper is to partially fill this gap by estimating separate earnings equations for employees and the self-employed using a tracer survey based on cohorts of secondary school students who completed their junior secondary schooling in Tanzania during the 1990s.

Tracer surveys are designed to find a group of individuals who have shared a specific type of training or educational background. They thus provide a basis to explore the impact of a common training or educational experience on labour market outcomes. If carefully designed, they facilitate the collection of richer and more detailed information than generally provided in conventional household or labour force surveys. This allows for some important refinements to the specification of earnings equations. In particular, the independent role of parental background and schooling quality can be explored, as well as the extent to which the returns to general human capital alter with the inclusion of such measures.

It has been shown that a failure to control for school level characteristics in wage equations represents a potential source of bias in the estimation of rates of return to education (see Behrman and Birdsall, 1983; Glewwe, 1996; Jolliffe, 1998; Glewwe, 2002). The data acquired through the school-based tracer survey used in this study allow the inclusion of school-specific fixed effects in earnings equation specifications. This may serve to attenuate the potential bias noted above, though it is acknowledged that the extent to which we can control for a broad range of school specific characteristics is limited.

The data available to us through the tracer survey are rich in a number of respects but could not be interpreted as entirely perfect for the task at hand. The empirical approach to estimating rates of return to education is constrained to some extent by a failure to control for innate ability. It is well established that if ability is correlated with both the level of education and labour market earnings, conventional rate of return estimates based on OLS (or an equivalent procedure) exhibit an upward bias. Ability factors are generally difficult to capture empirically and investigators have resorted to innovative approaches to overcome this potential problem.<sup>1</sup> Though there is a consensus about the direction of the bias, there is less on its magnitude. The evidence cited in Card (1999) suggests that the magnitude of the bias may be modest in nature. Our data contain little information that could be used to proxy for ability in a meaningful sense and the estimated rates of return are thus subject to a caveat in this regard.<sup>2</sup>

The objective of the study is to exploit earnings equations to provide estimates for a variety of human capital and other characteristics for the Tanzanian labour market. A primary theme of the paper is to investigate the magnitude of the rates of return to formal educational qualifications in both wage and self-employment sectors, where the latter could be interpreted as reflecting an informal segment of the Tanzanian labour market. This sector is of growing importance in Tanzania with an increasing proportion of secondary school-leavers engaged in self-employment activity over the 1990s (see Al-Samarrai and Bennell, 2003).

The structure of the paper can now be outlined. In order to place our empirical analysis in context, the next section provides a brief history of the education system in Tanzania and reviews the changes in the labour market over the 1990s. This section also summarises the findings from the small number of studies that have estimated earnings equations for Tanzania. Section 3 outlines the data and describes some of the advantages and drawbacks in estimating earnings equations using tracer survey data. Given the nature of how the earnings data are recorded and potential selection issues, section 4 outlines the econometric methodology and details how the estimated regression models are evaluated. Section 5 reports the empirical results and section 6 offers some conclusions.

## **2. Background**

In the aftermath of Tanzania's independence in 1962, education policy focussed on expanding access to basic education. This was regarded as a key component of the government's economic and social development strategy. Tanzania enjoyed rapid success, with enrolments in primary schools increasing four-fold during the 1970s, achieving primary gross enrolment rates of 100 per cent by the early 1980s (MOEC BEST, 1985). The dramatic increase in primary education enrolment did not provide the impetus for subsequent increases in secondary school enrolments. Access to the latter was restricted both because of resource constraints arising from primary school expansion, and Government policy that limited the establishment of private secondary schools in an attempt to reduce inequality of access. The sharp rise in primary school enrolments and limited secondary school expansion resulted in a decline in the proportion of primary school-leavers progressing to secondary level. The proportion of standard VII leavers continuing to secondary school fell from over one-third in 1961 to under one-fifth by 1967, and was only seven per cent by 1980 (see Knight and Sabot, 1990).

The consequences of war with Uganda and international economic conditions were responsible for the Tanzanian recession experienced during the late 1970s and early 1980s. The economic downturn and the subsequent adjustment policies led to major declines in primary school enrolment rates, which by 1990 declined to approximately

70 per cent. The impact was very different for secondary schooling. Parental pressure and constrained government finances led to an easing of restrictions on private and community schools during the 1980s (see Samoff, 1987). As a consequence, secondary school enrolments doubled between 1984 and 1990 (see MOEC BEST, various years), largely because of increases in the number of private schools, but also through the provision of an increased number of community-built government schools.

Despite large absolute increases in the numbers going through the secondary school system in the 1980s and 1990s, enrolment rates remained amongst the lowest in the world and compared unfavourably with other countries in the region (for example, 6% in Tanzania as compared to 27 % in sub-Saharan Africa in 2000 (UNESCO, 2003)). An even smaller proportion was selected into senior secondary schooling (Forms 5-6). For instance, in 1999 only 27 per cent of junior secondary school graduates (Forms 1-4) progressed to the senior level (MOEC, 2000). The progression from senior secondary school to university is also subject to selection and while tertiary enrolments more than doubled over the 1990s, less than one per cent of the school-aged population received tertiary education. As table 1 reveals, the Tanzanian education system is highly selective with large numbers of completers at each level failing to continue with their education. Only two per cent of the population aged 25 or over had completed the full six years of secondary education. It is also interesting to note that in a nationally representative survey undertaken in 1993 comprising approximately 5,000 households and 30,000 individuals, not one respondent is reported to have had completed a university degree.

[Table 1 here]

The labour force surveys undertaken in 1991 and 2000 provide some insight into the nature of the labour market that increasing numbers of secondary school completers and university graduates entered during that decade. The economically active population in Tanzania almost doubled during the 1990s, although this rise was not matched by a growth in wage employment opportunities. By the end of the 1990s the share of wage employment in the economically active labour force had declined from

eight to six per cent (Planning Commission, 1993, 2000). The majority of the increase in the economically active population was absorbed by the traditional agricultural sector, with some increases in non-traditional agricultural self-employment. Based on this limited evidence, it seems likely that over the 1990s the rising number of secondary school-leavers and university graduates entering the labour force exceeded the growth in wage employment opportunities. It also appears likely that a greater number of the educated became self-employed during this period.

The changes in the demand for educated individuals and the numbers entering the labour force after completing their education clearly impacts on the private rate of return to education secured in the labour market. Table 2 summarises the findings of studies that have estimated private rates of return to educational qualifications in the Tanzanian formal sector over the last 35 years. The returns to secondary education exceed those to primary in Tanzania and this possibly reflects a return to scarcity given the very limited numbers with secondary education. In addition, average rates of return to university qualifications are generally higher again. Therefore the selective nature of the education system creates labour market conditions whereby those individuals who gained access to secondary and then ultimately university education, secure relatively high labour market returns in the formal sector.

[Table 2 here]

This pattern of returns to education in Tanzania appears to differ from those observed elsewhere in sub-Saharan Africa. Private rates of return to primary, secondary and university education in the region tend to be much higher than those reported in table 2. Moreover, a recent review revealed higher returns to primary than to secondary education in sub-Saharan Africa (see Psacharopoulos and Patrinos, 2004).<sup>3</sup> Table 2 also reveals that the average rates of return to university qualifications are generally higher than those obtained in the primary or secondary sectors. Five out of the nine studies reviewed for sub-Saharan Africa (see Psacharopoulos and Patrinos, 2004: Table A1) showed a similar pattern to table 2.



The studies cited in table 2 are broadly comparable in regard to both the methodology employed and the types of controlling factors used in the earnings equations. In general, age is used as a proxy for labour force experience and controls for settlement type and gender are included. The Knight and Sabot study, of manufacturing employees from 1980, exploits somewhat richer data comprising actual levels of employment experience and measures of cognitive skills and reasoning ability.<sup>4</sup> Including these controls reduces the rate of return estimate by more than half and education's impact on wages becomes statistically insignificant. Interestingly, their measure of reasoning ability is not statistically significant in the wage equation and it is the inclusion of measures for cognitive skills that appears to reduce the estimated rate of return. The rates of return reported in table 2 relate to formal sector workers only and provide no indication of the rates in more informal sectors of the labour market. As wage employees constitute a very small proportion of the total labour force, this focus may fail to provide a complete portrait of the earnings distribution in the Tanzanian labour market.

### **3. Data**

The data used in this paper are drawn from a tracer survey conducted in Tanzania during 2001.<sup>5</sup> In exploring the link between education and labour market outcomes, tracer surveys have some advantages over conventional labour force or household surveys. The school that an individual attends may influence the type of work or training undertaken and the income earned after completion. This may be particularly important in a modestly sized education system such as Tanzania. In general, labour force surveys do not collect information on school characteristics or academic achievement. Many of these types of studies focus only on the quantity or type of schooling acquired. By tracing individuals through the schools they originally attended, a tracer survey allows information on school characteristics to be collected. This may prove helpful in analysing the impact of education on labour market outcomes and earnings.

It should be noted, however, that estimating earnings equations using tracer survey data provides limited information on the age-earnings profile since individuals tend to be clustered around certain ages determined by school starting age and individual progression through the school system. Tracer surveys only include information on a cohort of individuals that has achieved a certain level of education and/or training and therefore contain no information on individuals with either a lower level or no formal education. For example, the focus of the tracer survey in this study is secondary education. All individuals in the analysis have attained at least junior secondary education, and it is therefore not possible for our analysis to inform on rates of return to primary education. This represents another potential source for upward bias in the estimation of returns to educational qualifications in our wage equations given that the sample is conditioned on secondary school-leavers whose higher innate ability may have determined their access to secondary schooling in the first place.

There are other potential drawbacks in the use of tracer surveys. The selection of respondents is potentially biased towards individuals that are more easily traceable and, even with the best designed survey, low response rates may be encountered. If some systematic process governs the low response rates, sample estimates based on the survey may be subject to bias. The tracer survey used in this study was carefully designed to avoid this particular type of problem. It aimed to locate a specific group of secondary school-leavers that completed the first four years of secondary schooling in Tanzania. The sample of leavers was selected from ten average performing secondary schools: five located in Dar es Salaam and five in Dodoma. Although Dar es Salaam is the commercial capital of Tanzania, Dodoma is the capital city. It should be noted that the schools are average performing for the two locations but are in the top 30 per cent of schools nationally at the time our leavers attended.<sup>6</sup> In these selected schools, the sample frame consisted of all students who completed their junior secondary schooling in either 1990 or 1995. A random sample of 50 completers from each of the two years was selected from each school to generate a total of 1000 junior secondary school completers.

Once initial lists of sampled students were collated, research teams attempted to trace each individual's current location. Tracing selected individuals began at the schools,

where teachers and school records provided some information. The research teams then interviewed and completed questionnaires on all sampled respondents living in the vicinity of the schools. In addition to interviewing these respondents, the research teams inquired on the whereabouts of other sampled classmates where the school was unable to provide sufficiently accurate information. The research teams also interviewed parents and other household members of sampled students living elsewhere to obtain information on their current location as well as to collect basic information on the individual in question. Asking family members and classmates about the location of sampled students proved to be an extremely productive tracing technique. Once the research teams completed the work in the vicinity of the sampled schools, researchers were sent to interview respondents in other cities and locations with high concentrations of traced completers. In cases where traced respondents lived in remote areas a questionnaire was sent to them for completion.

The tracer survey response rate is reported in table 3 and is extremely high, with 97 per cent of respondents being traced and their information collected. This compares favourably with other tracer surveys conducted in developing countries (see for example, Psacharopoulos and Loxley 1985; Narman, 1992; Bennell and Ncube, 1993, 1994; Mayanja and Nakayiwa, 1997; Kaijage, 2000) and suggests that conventional problems associated with tracer surveys may be attenuated in our case.<sup>7</sup> The main source of the information collected was through the traced respondent, and in the majority of cases this was through ‘face-to-face’ interviews. In 50 cases, however, basic information on the respondent was collected from their parents.

[Table 3 here]

A questionnaire was completed for each respondent and was designed to elicit information on personal background, further education, training and employment history, current activity and labour market income. It should be noted that no information was collected for respondents who were overseas or deceased at the time of the survey, and these respondents are thus excluded from the present study. Of the remaining 940 respondents, usable information was available for 903.<sup>8</sup> Table A1 in

the appendix provides a full description of the variables used in the analysis and table A2 reports summary descriptive statistics.

It is worth noting that only four per cent of those engaged in self-employment were involved in agricultural activities. The majority of secondary school leavers in this sector were engaged in petty trading and other traditional self-employment activities, such as tailoring and hair-dressing. The self-employed in this tracer study do not tend to employ many additional staff, with almost one-half of self-employed respondents working alone. In contrast to those in wage employment who were asked about their earnings in the previous month, those engaged in self-employment activities were asked about earnings from their activities over the last six months in order to account for temporary fluctuations in self-employment earnings. Given the relatively straightforward activities in which this group was engaged, some of the drawbacks of measuring self-employment income, as identified by Vijverberg (1991), are significantly reduced. Nevertheless, the longer recall period and the lack of written records may render the reported self-employment earnings prone to measurement error. The use of responses from an interval-coded earnings question, however, may act to reduce potential bias attributable to this type of measurement error.

## **4. Methodology**

The primary objective of our research is served through estimating earnings equations. In order to elicit income responses that are less prone to measurement error interviewees were asked where their gross monthly earnings were located within a number of mutually exclusive categories. The intervals used for the sampled employees in our case commence at less than 50,000 shillings and rise by amounts of 50,000 for the next five intervals.<sup>9</sup> The penultimate interval is 300,000 to 400,000, and the final interval is open-ended and captures those individuals with earnings greater than 400,000 Tanzanian shillings. There are thus eight coded intervals for the case of employees. A corresponding exercise was conducted for the self-employed but the intervals used were different and relate to earnings in the six months prior to the survey. However, the econometric methodology outlined is identical for both cases

and for convenience the exposition is cast in terms of the earnings categories for employees.

The interval-coded nature of the responses for the dependent variable requires a maximum likelihood procedure. The appropriate likelihood function is a modification of that used in the estimation of the standard ordered probit model and replaces the unknown threshold values by the set of known values that delineate the intervals (see Stewart, 1983).

In order to understand how the model is implemented, responses are coded 1,2,...,8 to capture the eight distinct earnings categories in our application for employees. Let  $y_i$  denote the observable ordinal variable coded in this way and let  $y_i^*$  denote an underlying variable that captures the earnings of the  $i^{\text{th}}$  individual. This can be expressed as a linear function of a vector of explanatory variables ( $\mathbf{x}_i$ ) using the following relationship:

$$y_i^* = \mathbf{x}_i \boldsymbol{\beta} + u_i \quad \text{where } u_i \sim N(0, \sigma^2) \quad [1]$$

It is assumed that  $y_i^*$  is related to the observable ordinal variable  $y_i$  as follows:

$$\begin{array}{lll} y_i = 1 & \text{if} & -\infty < y_i^* < a_1 \\ y_i = 2 & \text{if} & a_1 \leq y_i^* < a_2 \\ y_i = 3 & \text{if} & a_2 \leq y_i^* < a_3 \\ y_i = 4 & \text{if} & a_3 \leq y_i^* < a_4 \\ y_i = 5 & \text{if} & a_4 \leq y_i^* < a_5 \\ y_i = 6 & \text{if} & a_5 \leq y_i^* < a_6 \\ y_i = 7 & \text{if} & a_6 \leq y_i^* < a_7 \\ y_i = 8 & \text{if} & a_7 \leq y_i^* < +\infty \end{array}$$

where the  $a_j$  for  $j=1, \dots, 8$  denote the interval boundaries. Following Stewart (1983), we treat the first and the last intervals as open-ended in this case so for  $j=0$ ,  $\Phi(a_j) = \Phi(-\infty) = 0$  and for  $j=8$ ,  $\Phi(a_j) = \Phi(+\infty) = 1$ , where  $\Phi(\cdot)$  denotes the cumulative distribution function for the standard normal.

The exact knowledge of the thresholds allows the likelihood function to be specified in a fairly straightforward manner. The variable  $y_i^*$  is best interpreted not as a latent measure but one with a quantitative interpretation. In implementing the procedure the standard normal assumption conventionally invoked for the ordered probit model is replaced by the assumption that  $y_i^* | \mathbf{x}_i \sim N(\mathbf{x}_i \boldsymbol{\beta}, \sigma^2)$ . This then allows specification of the log likelihood function as follows:

$$L = \sum_{j=0}^J \sum_{y_i=j} \log_e \left[ \Phi \left( \frac{a_j - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) - \Phi \left( \frac{a_{j-1} - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) \right] \quad [2]$$

where the first summation operator sums across individuals within the given  $j$  category and  $\log_e(\cdot)$  denotes the natural logarithmic operator. The maximum likelihood procedure involves the estimation of the  $\boldsymbol{\beta}$  parameter vector and the ancillary parameter  $\sigma$ , the standard error of the regression model. Given that the introduction of the known thresholds fixes the scale of the dependent variable, the estimated coefficients are amenable to a direct OLS-type interpretation.

Given an interest in estimating an earnings equation within the established Mincerian tradition, the thresholds used in the log-likelihood function [2] are based on the natural logarithms of the reported thresholds. Thus, in the context of the employee specification,  $a_1 = \log_e(50,000)$  and  $a_7 = \log_e(400,000)$ , and so forth.

It is important to evaluate our empirical models in regard to certain key econometric assumptions. The adequacy of the estimated models is assessed using the efficient score tests suggested by Chesher and Irish (1987). These tests require computation of the models' pseudo-residuals. In general terms, the pseudo-residual for the  $i^{\text{th}}$  individual is defined for the interval regression model as:

$$u_i = \frac{\phi \left( \frac{a_{j-1} - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) - \phi \left( \frac{a_j - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right)}{\sigma \left[ \Phi \left( \frac{a_j - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) - \Phi \left( \frac{a_{j-1} - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) \right]} \quad [3]$$

where  $\phi(\cdot)$  denotes probability density function operator for the standard normal.

These pseudo-residuals are then used in computation of the score tests. In order to implement such tests we also need the score contribution for the  $\sigma$  parameter. This is obtained as:

$$\sigma\_score_i = \frac{(a_{j-1} - \mathbf{x}_i' \boldsymbol{\beta}) \phi\left(\frac{a_{j-1} - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma}\right) - (a_j - \mathbf{x}_i' \boldsymbol{\beta}) \phi\left(\frac{a_j - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma}\right)}{\sigma^2 \left[ \Phi\left(\frac{a_j - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma}\right) - \Phi\left(\frac{a_{j-1} - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma}\right) \right]} \quad [4]$$

The score tests take the form  $\mathbf{i}' \mathbf{R} (\mathbf{R}' \mathbf{R})^{-1} \mathbf{R}' \mathbf{i}$ . In this case,  $\mathbf{i}$  is an  $n \times 1$  ‘summer’ vector of ones, and  $\mathbf{R}$  is an  $n \times q$  matrix of score contributions computed for each of the  $k$  parameters from the original specification and the  $k+1, \dots, q$  parameters that capture the form of the alternative hypothesis, assumed zero under the null. A typical row element for  $\mathbf{R}$  is given by:

$$\mathbf{R} = [u_i, u_i x_{1i}, \dots, u_i x_{ki}, \sigma\_score_i, Q_{1i}, \dots, Q_{qi}] \quad [5]$$

The  $Q_{1i}, \dots, Q_{qi}$  variables capture the alternatives under test. If the  $Q_{1i}, \dots, Q_{qi}$  terms were excluded from  $\mathbf{R}$  the score test would be zero by construction. The introduction of the  $Q_{1i}, \dots, Q_{qi}$  terms induce a departure from zero and the test statistic determines whether the difference from zero is statistically significant.<sup>10</sup>

The score tests focus on two key properties of the econometric specifications. These are homoscedasticity and normality. In regard to the former, the set of explanatory variables included in the earnings specifications provide the basis for the alternative heteroscedastic relationship. The  $Q_{1i}$  to  $Q_{ki}$  terms are constructed by interacting the  $\sigma\_score_i$  term with the  $x_{1i}$  to  $x_{ki}$  explanatory variables from the original regression model. The reported degrees of freedom for this particular test are thus equal to the number of explanatory variables in the original specification. The normality test examines departures from skewness and meso-kurtosis and the  $Q_{1i}$  and  $Q_{2i}$  measures are constructed in this case through interacting the pseudo-residuals with expressions

for the third and fourth moment residuals (see Chesher and Irish, 1987 for details). The number of degrees of freedom in this case is two. The resultant test statistics are all distributed as chi-squared with  $p = q - k$  degrees of freedom. The test represents the outer-product gradient (OPG) form of the score (or Lagrange Multiplier) test.<sup>11</sup>

The presence of heteroscedasticity introduces an inconsistency in the regression model estimates.<sup>12</sup> However, in circumstances where the homoscedasticity and/or normality assumptions are violated, the variance-covariance matrix is corrected using the Huber (1967) ‘sandwich’ estimator, which provides an appropriate asymptotic matrix for an estimator that is biased in an unknown direction. This is defined as:

$$\text{Var-Cov}(\hat{\beta}) = [I(\hat{\beta})]^{-1}(\mathbf{x}_i' u_i^2 \mathbf{x}_i) [I(\hat{\beta})]^{-1} \quad [6]$$

where  $I(\hat{\beta})$  is the information matrix for the  $\hat{\beta}$  vector computed at the maximum likelihood estimates. Greene (2000: pp.823-824) offers some cautionary comments about the use of this adjustment procedure. The violation of either of these two assumptions renders the estimator quasi-maximum likelihood.

Finally, there is potential for selection bias in the estimated earnings equations. The samples of employees and the self-employed may not represent a random drawing from the population of school-leavers. As a consequence, the estimated coefficients for the returns to education and other characteristics may be potentially biased. In order to address this problem we use the procedure developed by Lee (1983), which provides a more general approach to the correction of selectivity bias than that originally offered by Heckman (1979). The procedure is two-step and exploits estimates from a multinomial logit model (MNL) rather than a probit to construct the set of selection correction terms (see Lee, 1983 for the technical details).

A set of instruments is required to identify the parameters of the earnings equation, which includes the selection effects. The identifying variables are required to shift the probability of sectoral attachment but not the level of earnings within the sector. The selection measures suggested by Lee (1983) are computed from estimates derived from a four-category multinomial logit sectoral attachment model comprising



employed, the self-employed, the unemployed, and those still in training. The set of instruments used for identification are admittedly *ad hoc* and comprise of the education level of the respondent's spouse, time spent unemployed and self-employed since completing junior secondary school, and the division achieved on the Form IV secondary school examination. A set of variables based on the interaction of the cohort year with an individual's education level, and variables describing the respondent's duration in different activities since leaving formal education (i.e., wage employment, self-employment, training and unemployment) were also used as instruments. These identifying instruments generally exerted a significant impact on the current labour force activity of respondents but not on their labour market earnings. In order to conserve space the maximum likelihood estimates from the sectoral attachment model are not reported in full, though estimates relating to the role of educational qualifications in sectoral attachment are briefly discussed in the next section.<sup>13</sup>

## 5. Empirical Results

Prior to a detailed discussion of the estimates for the earnings equations, we briefly examine the role of educational qualifications in determining sectoral attachment using impact effects computed from the MNL model. Table A3 in the appendix reports the separate impact effects for senior secondary and a university qualification on sectoral attachment for the four categories used. There appears no independent role for education in determining engagement with the self-employed sector but possessing a university qualification relative to junior secondary schooling increases the probability of attachment to wage employment by 25 percentage points, on average and *ceteris paribus*. The comparable point estimate for secondary schooling for this sector is also positive but only attracts an asymptotic t-ratio of 1.5. The possession of a university degree is clearly beneficial in that it enhances the prospect of securing wage employment.

A basic model of earnings determination comprising controls for gender, religion, the individual's completion year of junior secondary schooling, job location and a variety of human capital controls is initially estimated. The latter controls include measures

designed to capture both general (for example, educational qualifications) and job-specific human capital (for example, job tenure). The basic model is then augmented in turn by controls for job branch sector, father's educational background<sup>14</sup> and a set of school controls (designed to proxy for school quality and potential labour market network effects). The primary motivation for adopting this approach is to determine how returns to the general human capital measures alter with the inclusion of variables that account for parental background and potentially schooling quality.<sup>15</sup>

### **5.1 Earnings Equation – The Employees**

The interval regression model estimates for the employees' sample are reported in table 4. A brief review of the diagnostics provides a prelude to a more detailed discussion of the earnings equation estimates themselves. The result for the normality assumption is somewhat mixed and on the borderline for our preferred specification (4) in table 4. We believe, however, it is innocuous to infer that normality provides a reasonable approximation for the unobservables governing the determination of the earnings process for employees. The null hypothesis of homoscedasticity is decisively rejected for the four estimated models and the Huber (1967) adjustment is used to construct a robust variance-covariance matrix for the coefficient estimates. The goodness-of-fit measure for our preferred specification is satisfactory by the standards of cross-sectional earnings equations, though it does not possess the interpretation associated with the conventional OLS procedure.<sup>16</sup>

[Table 4 Here]

The estimated coefficient for the selection correction variable is poorly determined in all four specifications and suggests that the sample of employees comprise a random drawing from the sample of school-leavers available to us. In addition, it is worth noting that the estimated selection effect is not materially affected by the inclusion of additional regressors in the earnings equation.

The discussion of the estimates will primarily focus on our preferred specification (4) in table 4 but cross-reference will be made to other specifications where necessary.

The cohort effect achieves statistical significance within the five per cent level for our preferred model. This variable is designed to proxy for age effects and suggest that the average *ceteris paribus* earnings for the 1990 cohort are almost 26 per cent higher than for the more recent 1995 cohort of leavers.<sup>17</sup> There is also some evidence of gender and religious advantage in employee monthly earnings with the male premium estimated at 24 per cent and Muslims securing about 16 per cent more than non-Muslims, on average and *ceteris paribus*. The latter wage premium may in part reflect the influence of social networks.

There is a sizeable return of nearly 34 per cent for those with jobs located in Dar-es-Salaam compared to the relatively heterogeneous base comprising all other Tanzanian areas other than the administrative capital of Dodoma. This may capture cost of living differences. Approximately one-third of the employees in the sample work in the health and education sector and earn, on average and *ceteris paribus*, 16.4 per cent less in monthly terms than workers in ‘other’ sectors. Given the recent expansion in educational access in Tanzania more secondary school completers are likely to be absorbed into the education sector as teachers, but the lower wages are unlikely to attract the best qualified candidates. The monthly earnings of public sector workers, however, are not statistically different from their private sector counterparts.

The estimated coefficient for the training variable is robust across all the reported specifications and the estimates are generally invariant to the inclusion of the additional controls for parental background and school effects.<sup>18</sup> In interpreting the estimate it is useful to translate the effect into an annualised return. Thus, an additional year of general training raises monthly earnings by 8.4 per cent, on average and *ceteris paribus*. The estimated effect for on-the-job experience, as measured by job tenure, is a more modest 3.6 per cent per annum but borderline in terms of statistical significance.<sup>19</sup> Given such further training and education generally involves a substantial investment on the part of these individuals and their parents, it is salutary to note that such investments appear, on the average, to have some pay-off in the formal sector.

The estimated returns for the educational qualifications are mildly sensitive to the gradual inclusion of additional controls across the different specifications. The introduction of parental and school controls attenuates the estimated effect for the senior secondary qualification but it retains statistical significance at a conventional level in our preferred model. A similar pattern is evident for the estimated effect for university graduates but the attenuation is most pronounced with the introduction of the father's educational background variables. The estimated coefficient for a university graduate falls by 0.11 log points (or over one-tenth in value) with the inclusion of father's education. This could be taken to highlight the importance of controlling for parental background when attempting to obtain informative estimates for the returns to educational qualifications. A more detailed discussion on the returns to the educational qualifications is provided below in section 5.4.

An examination of the estimated effect of a father's educational background on earnings, however, is important in its own right. The estimates in our preferred specification suggests that individuals whose fathers have higher education earn 44 per cent more than those individuals whose fathers have less than secondary education, on average and *ceteris paribus*. One possible interpretation for father's education is that it is potentially correlated with the innate ability of the wage earner. The inclusion of this regressor could thus also act to reduce any bias in our rate of return estimates that are channelled through unobserved ability effects. There are a number of other possible interpretations for this particular result. It may reflect the fact that educated parents are more effective at inculcating in their children the skills that are ultimately well rewarded in the Tanzanian labour market. Alternatively, it may be the case that individuals whose fathers are highly educated can exploit informal networks in securing better-paid jobs for their off-spring. Both interpretations could be viewed as providing mechanisms that reinforce an inter-generational transfer of inequality. It is also worth noting that the inclusion of the school controls reduces the father's higher educational effect by almost one-third. This may suggest that the school system, as captured by our school controls, also plays some role in ossifying such inter-generational inequality.

## 5.2 Earnings Equation – The Self-employed

The estimates for the sample of self-employed are reported in table 5. In contrast to the regression models for employees, the performance in regard to the diagnostics is considerably poorer. In terms of our preferred specification, the model registers a more decisive rejection for normality and the assumption of homoscedasticity is again violated. In addition, the fits of the regression models are poor by any standard and few of the estimated coefficients are well determined. There are a couple of exceptions to this general statement and one is reserved for the estimated selection effect. The negative coefficient is consistent with the presence of positive selection given the construction and derivation of the selection term in this application (see Lee, 1983). The point estimate for the selection effect suggests that those who select into self-employment earn about 17 per cent more in earnings than those drawn at random from our traced sample of school-leavers.<sup>20</sup>

[Table 5 Here]

An interesting feature of the regression estimates for the self-employed draws on a comparison with table 4. In particular, the estimate for the standard error of the regression is considerably higher for the self-employed than for employees. Part of this may be attributable to the use of more earnings intervals in the self-employed application but this cannot explain all of the difference. This could be taken to confirm to some extent that unobservable prices and quantities are considerably more important in earnings determination for self-employed than for employees.

Ignoring school effects and the selection effect already discussed, only three of the estimated coefficients for the self-employed attain statistical significance at a conventional level in our preferred specification: current job tenure, wage employment experience and holding a university qualification. The amount of experience acquired in wage employment enhances the earnings of the self-employed with an additional month raising self-employed income by two per cent, on average and *ceteris paribus*. This could be taken to tentatively suggest that time spent in the formal labour market enhances general labour market skills and thus earnings in the

self-employed sector. The annualised effect for current job tenure is 8.4 per cent. On the face of it, the possession of a university degree yields sizeable benefits for the self-employed. However, it is worth cautioning that only a small number of the self-employed hold such a qualification (8 out of 211) and these estimates may reflect the characteristics of this small group rather than informing more generally on the returns to this measure of human capital in the self-employed sector.

It is important to complete this sub-section by noting those factors that appear to play no independent role in earnings determination for the self-employed. These include gender and religious affiliation and this could be taken to suggest little evidence of consumer-based discrimination in our Tanzanian sample given most of the self-employed are engaged in activities that directly interact with consumers (see Al-Samarrai and Bennell, 2003). Earnings are reasonably flat in terms of the training measure used and the age of the individual, which is proxied by the school-leaver cohort variable. There appears no role for parental education in earnings determination in the self-employed sector and no evidence of regional variation in self-employed earnings either. It also appears to matter little whether the individual works on ‘own-account’ or employs others. We also estimated an earnings model for the self-employed which allowed for a complete set of interactions between the variable EMP, capturing whether an individual employs others, and all the independent regressors in specification (4) of table 5. There was no evidence of any statistical differences in the estimated coefficients either individually or jointly.<sup>21</sup>

It could be argued that our educational estimates for the self-employed equation are subject to bias because of a failure, through the limitations of the data, to control for the contribution of capital. This ultimately affects the marginal product of labour and, on the assumption of a competitive labour market, earnings. We can only offer some conjectures on the implication of this omission for our educational estimates. If we assume that capital is positively correlated with educational level, because educated individuals are less likely to be constrained in their access to capital markets, and positively associated with earnings through the assumption of a competitive labour market, then the omission of this factor is likely to lead to an upward biased estimate in the returns to educational qualifications. Given the educational estimates obtained

for the self-employed are poorly determined, we take the view that the absence of this measure does not materially alter our inferences on the rates of return to educational qualifications for the self-employed.

### 5.3 The Estimated School Effects

One motivation for the inclusion of the school effects is to control, *inter alia*, for schooling quality in terms of its potential effect on education returns and on earnings more generally. Their inclusion is also motivated by a perception that these controls might also capture labour market network effects. The inclusion of the school controls tends to reduce both the return to senior secondary schooling and to a university qualification but induces a sharper reduction in the former relative to the latter.<sup>22</sup> This is a relatively intuitive result and highlights the importance of controlling for schooling quality/networks in computing rates of return to educational qualifications. It is apparent that parental educational background is more important to the estimated university returns than secondary schooling quality.

The school effects can be interrogated more thoroughly by normalizing the ten estimated school effects as a deviation from an overall school weighted average. This transformation has appeal in that the estimated earnings differences are then expressed relative to an overall average rather than an arbitrary base group and are thus more easily interpretable (see Krueger and Summers, 1988). If we define the school effect for the  $k^{\text{th}}$  school as  $\gamma_k$  the deviation is expressed as:

$$\gamma_k - \sum_{j=1}^{10} \pi_j \gamma_j \quad [7]$$

where  $\pi_j$  is the proportion of the sample in the  $j^{\text{th}}$  school. The corresponding sampling variances are computed as per Zanchi (1998).<sup>23</sup> Table 6 reports the estimated deviations for the ten schools for both wage employees and the self-employed.

[Table 6 Here]

The *ceteris paribus* earnings for those in eight of the ten schools are not statistically different from the average for either the employee or the self-employed groups. However, there are two schools for which the earnings of their *alumni* (both employees and the self-employed) are different from the overall average. One school is well above and the other well below the average. In terms of the employee group the graduates of one school earn *ceteris paribus* about 59 per cent more than the average, while graduates from the less favoured school earn *ceteris paribus* about 38 per cent less than the average. The contrast in earnings performance between the graduates from the two schools appears starker in terms of the point estimates for the self-employed but there is no statistical difference in these effects across the two sectors – either individually or jointly. In addition, the estimated Spearman rank order correlation coefficient for school rankings across the two sectors on the basis of the earnings of their *alumni* is 0.71, which is statistically different from zero at the five percent level.

It is evident that the school an individual attends matters in a number of cases but why it matters may be too complex to unravel satisfactorily given the data available to us. The more favourable school in terms of above average labour market earnings reported in table 6 is a private co-educational non-boarding school based in an urban area of Tanzania. The less favourable school is public but identical in respect of the other characteristics noted. In addition, the national ranking of the public school in the 1990s actually appears superior to that of the private school. However, those that attended the latter tended to be disproportionately drawn from backgrounds where the father had either secondary or higher education. Thus, it could legitimately be argued that this particular school effect is partly reflecting a socio-economic background factor not explicitly captured in our estimated specifications. It might capture labour market network effects in the sense that graduates from this particular school are better able to exploit a network of former graduates who have perhaps attained influential positions in the Tanzanian labour market.

#### **5.4 Estimated Private Rates of Return to Educational Qualifications**



We now turn to an examination of the estimated private rates of return for the two educational qualifications used here – senior secondary and university. These are computed using the differences in the estimated coefficients between adjacent qualifications divided by the difference in years. If we define the interval regression earnings model parameter for a university qualification as  $\gamma_U$  and  $\gamma_S$  for university and senior secondary respectively, then the rate of return for a university qualification is computed as:

$$\text{University Qualification} = \frac{\gamma_U - \gamma_S}{U_{\text{Years}} - S_{\text{Years}}} \quad [8]$$

where  $U_{\text{Years}}$  is total years in education to acquire a university qualification and  $S_{\text{Years}}$  is the corresponding number of years taken to acquire a senior secondary qualification. The rate of return to senior secondary schooling is computed as:

$$\text{Senior Secondary Qualification} = \frac{\gamma_S}{S_{\text{Years}} - J_{\text{Years}}} \quad [9]$$

where  $J_{\text{Years}}$  is total years in education to acquire a junior certification qualification. The sampling variances for the estimated rates of return are easily computed using the conventional formula.

The estimated rates return reported in table 7 are the exponentiated versions of [8] and [9] expressed in percentages. The corresponding sampling variances are then constructed using the delta method. In respect of employees the point estimate for the senior secondary qualification is just under nine per cent and is well determined. The point estimate for the self-employed, though comparable, is imprecisely estimated.

[Table 7 Here]

The rate of return for a university qualification for the employees is well determined. The point estimate suggests an annualised rate of return of 17.1 per cent. The point estimate for the self-employed is slightly larger but again more imprecisely estimated.

In order to test whether there is any variation in the rates of return across characteristics, we introduced a number of interaction terms into the employee specification (4) of table 4.<sup>24</sup> We exclude the public sector from this analysis as this issue is examined in more detail separately below. A summary of the test results are reported in table A4 of the appendix. There is no evidence of variation in the returns to senior secondary or university qualifications across the school leaving cohort year, religion, gender, job location, or parental background characteristics. However, a significant result was detected in regard to the health and education branch sector. This was driven by variation in the return to the senior secondary qualification for those working in this sector. In particular, the estimated return to a secondary qualification was estimated to be statistically lower in this sector than in all other sectors. However, the estimated rate of return to a university qualification for those working within this sector is not statistically different from that reported in table 7.

Given that over one-third of wage employees are in the public sector, we also explored the nature of the differences in estimated effects between the public and private sectors by separating the sample of workers by these two sectors. The wage equation estimates for this exercise are reported in table A5 of the appendix.<sup>25</sup> Overall, the data fail to support complete sectoral separation, though there are statistical differences in some coefficients worthy of independent comment. On the basis of the estimated standard errors for both regression models, there is evidence of a greater wage compression in the public sector, which is not surprising. There are no statistical differences detected for the estimated educational effects across the two employment sectors. Thus, it is difficult to argue that the government compresses the Tanzanian wage structure through constraining rates of return to educational qualifications. In addition, it could be argued that the estimated rates of return to educational qualifications reported in table 7 are relatively constant across these two sectors. However, the public sector in Tanzania, compared to its private sector counterpart, does appear to act in a much more egalitarian fashion in regard to wage rewards for both gender and religious groups. In addition, the returns to job tenure are well determined in public jobs reflecting the existence of a public sector wage

reward structure that is more closely linked to tenure than that which prevails in the private sector.

As noted in the introduction to this paper, the evidence on the estimated rates of return to education in Tanzania is relatively sparse. Nevertheless, it would be useful to situate our findings on returns within the broader context of the literature. Researchers have computed both marginal returns to an additional year in education and returns to qualifications for Tanzania. In terms of the former, Psacharopoulos (1994) reported an estimate of 12 per cent, while in more recent work Söderbom *et al.* (2003), using data for the Tanzanian manufacturing sector, provided estimates in a range from 12 per cent for young workers to 17 per cent for older workers. This study also reported some evidence of an increase in the estimated educational returns in Tanzania over the last decade.

The literature on estimates for returns to qualifications, reported earlier in table 2, is even less extensive for Tanzania. The estimates reported by a number of authors for the returns to secondary schooling are broadly in comport with our findings. The point estimate for the university qualification obtained using the tracer study data is almost identical to that obtained using the 2001 LFS data. In addition, the LFS-based point estimate from 1991 comfortably lies within even the most modest confidence interval generated for the estimate for our wage employees. This could be taken to tentatively suggest that the rate of return to a university qualification has remained relatively stable in the formal labour market in Tanzania over the period covering the last decade of the 20<sup>th</sup> century.

## **6. Conclusions**

A key finding of our analysis is that the point estimates for the private returns to a university education are higher compared to senior secondary and this finding is compatible with the notion of a convex relationship between earnings and educational level as recently posited for Tanzania by Söderbom *et al.* (2003). On the basis of comparison with previous studies, there is no indication that the estimated rates of

return to senior secondary and university qualifications have actually increased over the last decade. Indeed, there is evidence of some degree of stability in these returns over time. We acknowledge that such a comparison can only provide a tentative rather than definitive insight into the temporal movement in the returns to these levels of education.

Further education and training undertaken by secondary school-leavers, once formal education is completed, appears to increase wage employment earnings. This is not surprising given the heavy investments secondary school-leavers make in further education and training. The average respondent in our survey, who had completed junior secondary education in 1990 (1995), had spent approximately three (two) years between 1990 (1995) and 2001 in further education and training (see Al-Samarrai and Bennell, 2003). Unfortunately it did not prove possible to disaggregate the months of training into their different types. A deeper understanding of which types of training have exerted the biggest impact on earnings would be an interesting avenue for future research but one not pursued here given data limitations.

Information contained in the tracer surveys allowed us to include controls for father's educational background and a set of school controls designed to proxy for school quality and potential labour market network effects. The inclusion of school fixed effects reduced the estimated rates of return to education. Our analysis revealed that the school an individual attends is important and this in part may be due to labour market network effects with leavers from particular schools having preferential access to better paying jobs through school or other contacts.

The estimated self-employment earnings equation was found to be poorly determined. The poorer performance of the self-employed earnings equation may be linked to the heterogeneous nature of the sector and the fact that the earnings measure contains a great deal of noise and a greater potential for measurement error. In particular, we found little evidence of well determined human capital effects in the earnings determination process in the self-employment sector. University education appeared to exert a statistically significant impact on self-employment earnings relative to a

base comprising junior secondary, but the small cell size for the university-educated group merits interpretational caution here.

The findings in regard to the self-employment sector are predictable given the type of activities in which secondary school-leavers are traditionally engaged. Typically, secondary school leavers were working as small-scale vendors (i.e., buying and selling goods) with relatively few additional staff (see Al-Samarrai and Bennell, 2003). It is unlikely that post-primary education will be of much benefit in undertaking these activities. Furthermore, secondary school-leavers in Tanzania tend to use self-employment, rather than unemployment, as a queuing strategy for waged employment opportunities. Given the increasing numbers of secondary school-leavers entering self-employment these findings are of potential concern from a policy perspective. Secondary school-leavers interviewed for the survey indicated that they were at least partially aware of the lack of relevance of their secondary education for self-employment as only a minority agreed that the secondary school curriculum was relevant (see Al-Samarrai and Bennell, 2003).

The estimates we obtained for the private rates of return to both secondary and university qualifications for wage employees were comparable in magnitude to those obtained using the most recent round of the Tanzanian Labour Force Survey. Our empirical evidence suggests that the rates of return to the highest educational qualifications are not small and at the margin provide an investment incentive. We can assert, however, that despite an extremely small secondary and university education system, the private rates of return to education in the Tanzanian wage employment sector appear low in comparison to other countries in the region. This is puzzling given the highly restrictive nature of post-primary education opportunities in Tanzania. One putative explanation for this particular finding is that the quality of secondary and university education provision in Tanzania is inferior when compared to provision in other developing countries. This issue is clearly worthy of independent investigation but goes well beyond the scope of the current study. Tanzania has not been represented in any cross-national study that would enable a systematic test of this particular hypothesis, and this important issue must thus be consigned to an agenda for future research.

Finally, it is worth pointing out that the tracer survey and the analysis reported here have generated a substantial amount of detailed information on individual characteristics, labour market outcomes and earnings for a sample of secondary school-leavers in Tanzania. This information is potentially important for both education and labour market policy. The cost of the tracer survey used in this study was approximately £25,000 (in 2001 prices) and represents a highly cost-effective way of generating such detailed information and analysis for policy-makers in developing countries where government financial resources are tightly constrained.

## ENDNOTES

- <sup>1</sup> For example, Bonjour *et al.* (2003) use the wage difference between sets of identical twins to control for innate ability. See also Card (1999: 1843-1844) for a survey of the empirical evidence from using twins in this context.
- <sup>2</sup> The tracer survey also elicited information on measures of educational achievement in secondary schools for the respondents. However, the measures were found to lack variation and did not influence earnings. A good measure of innate ability should be independent of academic achievement in schools, so this would be a misleading measure to use for the purpose here.
- <sup>3</sup> In periodic reviews of rates of return studies globally Psacharopoulos has provided regionally aggregated returns to different levels of education (Psacharopoulos, 1985, 1994; Psacharopoulos and Patrinos, 2004). These tend to reveal that primary education has the highest private returns followed by university and then secondary education. However, there has been substantial debate around the quality of data used in these regional aggregates, as well as the time period covered. For instance, Bennell (1996) shows that the aggregated rates of return to secondary exceed primary for sub-Saharan Africa, provided studies judged by the author of poor quality are actually excluded.
- <sup>4</sup> Reasoning ability was measured using Raven's Progressive Matrices and cognitive skills on reading and mathematics achievement based on questions from the Primary School Leaving Examination and the Form IV national examination.
- <sup>5</sup> The Tanzania secondary school tracer survey analysed in this paper formed part of a larger project conducted for secondary school leavers and university graduates in Malawi, Tanzania, Uganda and Zimbabwe. For details of the wider research findings see Al-Samarrai and Bennell (2003).
- <sup>6</sup> It is acknowledged that this may imply that the rates of return to education estimated with information from these ten schools overestimates the average return to education based on data drawn from all secondary schools in Tanzania.
- <sup>7</sup> Typically response rates in tracer surveys are around 50 per cent. The Psacharopoulos and Loxley (1985) study in Tanzania achieved a response rate of 60 per cent for 1981 Form IV leavers traced only one year after leaving school.
- <sup>8</sup> Of these, 400 were in wage employment and 211 were self-employed. The remainder were either unemployed or undertaking further training.
- <sup>9</sup> At the time of the survey there were 876 Shillings to the US dollar.
- <sup>10</sup> The likelihood function estimation and the computation of the diagnostic tests reported in this paper were all undertaken using the LIMDEP 7.0 software package (see Greene, 1998).
- <sup>11</sup> Orme (1990) provides some criticisms of OPG-based tests.
- <sup>12</sup> This is acknowledged as a difficult issue in these types of models. We attempted to address the issue of heteroscedasticity when encountered through expressing the  $\sigma$  term in the log-likelihood function of expression [2] as a function of a set of explanatory variables and their interactions. Attempts to optimise the log-likelihood function using the optimisation routines in LIMDEP 7.0 proved difficult given a lack of convergence. The lack of convergence may be explained by the small sample size available to us.
- <sup>13</sup> The chi-squared statistic (with 16 degrees of freedom) for the joint significance of the set of identifying variables in the wage employment (self-employment) sectoral equation was 69.23 (123.24) suggesting that the null hypothesis of zero coefficients is rejected at the one per cent level in both cases. The corresponding chi-squared statistics (with 16 degrees of freedom) were 17.02 and 16.93 for this same set of variables in the wage and self-employment earnings equations respectively implying that the null hypothesis of zero coefficients is upheld in both cases.
- <sup>14</sup> The use of variables capturing the educational background of the mother was also experimented with in the earnings equations but on all occasions proved to be poorly determined and was dropped from the analysis.
- <sup>15</sup> We estimate separate earnings equations for the employed and self-employed sub-samples. The fact that the earnings measures are subject to different coding intervals across the two employment types and relate to different measurement periods vitiates any sensible pooling exercise.
- <sup>16</sup> In other words, we cannot use these measures to infer that a certain proportion of the variation in monthly earnings is explained by variation in the explanatory variables. It can only be taken to provide the proportional improvement in the log-likelihood function's value consequent on the introduction of the relevant explanatory variables.

- 
- <sup>17</sup> The dependant variables in Tables 3 and 4 are logged and therefore, in the text, the coefficient estimates on dummy variables are transformed to ease interpretation. The percentage change in the dependent variable when the dummy variable is equal to one is computed as  $(\exp(\beta)-1)\times 100$  where  $\beta$  is the estimated coefficient corresponding to the dummy variable.
- <sup>18</sup> It should be noted that training does not include formal education pursued after respondents left junior secondary school (i.e. senior secondary and first degree). Most commonly, respondents undertook training courses in computing, management and accounting. For a more in-depth analysis of training undertaken by the respondents see Al-Samarrai and Bennell (2003).
- <sup>19</sup> Controls designed to capture previous labour force experience in months in waged-employment, self-employment or unemployment yielded poorly determined effects in the reported specifications and were consequently omitted.
- <sup>20</sup> This is computed as the product of the negative value of the selection coefficient (0.458) multiplied by the average value of the selection term (0.37). This yields the average selection effect which equals 0.169 in this case.
- <sup>21</sup> The Wald value for the test of joint statistical significance is estimated at 25.4 with 23 degrees of freedom. The corresponding prob-value is 0.329.
- <sup>22</sup> In table 4 the estimated coefficient for the senior secondary qualification reduced by over one-quarter with the inclusion of school effects, while the comparable estimate for the university qualification reduced by over one-tenth. This is perhaps unsurprising given that the school controls are not controlling for attendance at different universities and thus, by definition, do not capture university quality or network effects.
- <sup>23</sup> See also Haisken-De New and Schmidt (1997) for a more detailed discussion on deriving the sampling variances in this context.
- <sup>24</sup> Given the poorly determined nature of many of the reported coefficients, this analysis was not undertaken for the self-employed sector.
- <sup>25</sup> The separation of the sample in this way ignores the issue of endogenous selection into public or private sector jobs and this should be borne in mind in interpreting the results.



## References

Al-Samarrai, S. and Bennell, P. (2003) Where has all the education gone in Africa? Brighton: Institute of Development Studies and Knowledge and Skills for Development.

Bennell, P. (1996) Rates of return to education: Does the conventional pattern prevail in Sub-Saharan Africa. *World Development*, 24(1), pp. 183-199.

Bennell, P. (1994) Jobs for the boys? The employment experiences of secondary school leavers in Zimbabwe. *Journal of Southern African Studies*, Vol. 20(2) pp. 301–15

Bennell, P. and Ncube, M. (1993) Education and training outcomes: University graduates in Zimbabwe during the 1980s. *Zimbabwe Journal of Educational Research*, 5(2), pp. 107–123.

Behrman, J.R. and Birdsall, N. (1983) The quality of schooling: Quantity alone is misleading. *The American Economic Review*, 73(5), pp. 928-946.

Bonjour, D., Cherkas, L., Haskel, J., Hawkes, D and T. Spector. (2003). Returns to Education: Evidence from U.K. Twins, *American Economic Review*, 93(5), pp. 1799-1812.

Card, D. (1999) The causal effect of education on earnings, in: O. Ashenfelter and D. Card (eds) *Handbook of Labor Economics*, Vol.3.

Chesher, A. and Irish, M. (1987) Residual analysis in the grouped and censored normal regression model. *Journal of Econometrics*, 34, pp. 33–61.

Haisken-DeNew, J. P. and C. M. Schmidt (1997) Inter-industry and inter-region differentials: Mechanics and interpretation. *Review of Economics and Statistics*, 79, pp. 516-521.

Heckman, J. (1979) Sample selection bias as a specification error. *Econometrica*, 47, pp. 153-161.

Huber, P.J. (1967) The behaviour of maximum likelihood estimates under non-standard conditions, in: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA: University of California Press, pp. 221 – 223.

Glewwe, P. (2002) Schools and skills in Developing Countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature*, 150(2), pp. 436-482.

Glewwe, P. (1996) The relevance of standard estimates of rates of return to schooling for education policy: A critical assessment. *Journal of Development Economics*, 51, pp. 267-290.

Greene, W.H. (1998), *LIMDEP Version 7.0*, User's Manual Revised Edition, Econometric Software, Inc. New York.

Greene, W.H. (2000) *Econometric Analysis*, 4<sup>th</sup> edition. (New York, Macmillan).

Jolliffe, D. (1998) Skills, schooling, and household income in Ghana, *The World Bank Economic Review*, 12(1), pp. 81-104.

Kaijage, E.S. (2000) Faculty of Commerce and Management graduates and their employers: A tracer study. University of Dar es Salaam, Dar es Salaam.

Knight, J.B. and Sabot, R.H. (1990) *Education, productivity and inequality*. (New York: Oxford University Press).

Knight, J.B. and Sabot, R.H. (1981) The Returns to education: Increasing with experience of decreasing with expansion? *Oxford Bulletin of Economic Statistics*, 43, pp. 51-71.

Krueger, A.B. and Summers, L.H. (1988) Efficiency wages and the inter-industry wage structure. *Econometrica*, 56(2), pp. 259-293.

Lee L. (1983) Generalized models with selectivity. *Econometrica*, 51, pp. 507 – 512.

Mason, A.D. and Khandker, S.R. (1997) Household schooling decisions in Tanzania. World Bank, Gender and Public Sector Management Department: Washington D.C.

Mayanja, M.K. and Nakayiwa, F. (1997) Employment opportunities for Makerere University graduates: Tracer study. Makerere University, School of Post Graduates and Research: Kampala.

MOEC (various years) Basic education statistics in Tanzania (BEST). Ministry of Education and Culture: Dar es Salaam.

Mukyanuzi, F. (2003) Where has all the education gone in Tanzania? Brighton: Institute of Development Studies and Knowledge and Skills for Development.

Narman, A., (1992) Trainees at Moshi National Vocational Training Centre: Internal achievements and labour market adoption: Final report on a tracer study project in Tanzania. Swedish International Development Agency: Stockholm.

Orme, C. (1990) The small sample performance of the information matrix test. *Journal of Econometrics*, 46, pp.309-331.

Planning Commission (2000) Labour force survey, draft report. Government Printers: Dar es Salaam.

Planning Commission (1993) Labour force survey. Government Printers: Dar es Salaam.

Psacharopoulos, G. (1994) Returns to education: A further international update and implications. *Journal of Human Resources*, 20(4), pp. 583-604.

Psacharopoulos, G. (1985) Returns to investment in education: A global update. *World Development*, 22(9), pp. 1325-43.

Psacharopoulos, G. and Loxley, W. (1985) *Diversified secondary education and development: Evidence from Colombia and Tanzania*. (Baltimore: John Hopkins Press).

Psacharopoulos, G. and Patrinos, H. (2004) Returns to investment in education: A further update. Policy Research Working Paper No. 2881. The World Bank.

Samoff, J. (1987) School expansion in Tanzania: Private initiatives and public policy. *Comparative Education Review*, 31, pp. 333-360.

Söderbom, M., Teal, F., Wambugu, A. and G. Kahyara (2003) The dynamics of returns to education in Kenyan and Tanzanian Manufacturing. Centre for the Study of African Economies working paper WPS/2003-17, Oxford University.

Stewart, M. (1983) On least squares estimation when the dependent variable is grouped, *Review of Economic Studies*, 50, pp.737-753.

UNESCO (2003) Global monitoring report: Gender and education for all, the leap to equality. UNESCO: Paris.

Vijverberg, W.P. (1991) Measuring income from family enterprises with household surveys. Living Standards Measurement Study Working Paper 84, World Bank.

World Bank (2005) Secondary education development programme. Report No. 27631. World Bank: Washington.

Zanchi, L. (1998) Interindustry wage differentials in dummy variable models, *Economics Letters*, 60, pp. 297-301.

Table 1: Highest Level of Education Achieved in Tanzania (population Aged 25+), 1993 (%)

Education level	Male	Female	Total
No education	15.1	35.7	25.7
Incomplete primary	23.1	18.2	20.6
Primary	46.9	38.6	42.6
Junior secondary	8.8	4.1	6.4
Senior secondary	3.4	0.8	2.0
University degree	0.0	0.0	0.0
Other	2.8	2.6	2.7
<b>Total</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>

Notes to table 1:

(a) Authors' calculations from Human Resource Development Survey 1993.

(b) The proportions given in the table are weighted proportions from the survey that inflate the individual level data to nationally representative figures.

Table 2: Private Rates of Return – Summary of Estimates (%)

Author(s)	Year and source of Data	Sub-sample	Sample size	Primary	Junior Secondary	Senior Secondary	Secondary	University
Mason & Khandker (1997)	1991 – Labour Force Survey	male	2,353	7.0	12.0	2.0	8.5	12.0
		female	1,882	14.0	22.0	3.0	13.9	2.0
Knight and Sabot (1981 and 1990)	1971 and 1980 - Survey of manufacturing sector employees		1980 – 179 1971 – 777	2.1	7.3 (2.8)	§	§	§
World Bank (2005)	2000/01 – Labour Force Survey		3,017	11.0	25.0	9.0	9.4	17.0

Notes to table 2:

(a) The rates of return are author's calculations based on the Mincerian regression models reported in the studies.

(b) § denotes not applicable or not reported.

(c) all rates of return are for wage employment in the formal sector and are calculated using Mincerian regression models.

(d) For comparability purposes the results reported in this table do not include adjustments to the returns for ability etc.

(e) The primary rate of return for Knight and Sabot is from Knight and Sabot (1981) and for junior secondary from Knight and Sabot (1990). The rate of return for junior secondary in parentheses is the rate of return when controls for ability and cognitive skills are included.

Table 3: Tracer Survey Response Rates

	Male	Female	Total
<b>Total Traced</b>	<b>510</b>	<b>455</b>	<b>965</b>
Interviewed	466	408	874
Postal	7	9	16
Parents	25	25	50
Abroad	2	2	4
Deceased	10	11	21
<b>Not Traced</b>	<b>16</b>	<b>19</b>	<b>35</b>
<b>Total Sample</b>	<b>526</b>	<b>474</b>	<b>1000</b>
<b>Response rate (%)</b>	<b>97</b>	<b>96</b>	<b>97</b>

Table 4: Interval Regression Estimates for Earnings Equation – Employees

<b>Variables</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>
<b>Constant</b>	10.591*** (0.165)	10.704*** (0.166)	10.506*** (0.165)	10.588*** (0.178)
Cohort_90	0.106 (0.121)	0.150 (0.122)	0.202* (0.114)	0.231** (0.110)
Male	0.089 (0.077)	0.067 (0.079)	0.134* (0.079)	0.218** (0.104)
Muslim	0.133 (0.089)	0.163** (0.087)	0.169** (0.083)	0.146* (0.079)
<b>Human Capital:</b>				
Senior Secondary	0.276** (0.096)	0.278** (0.093)	0.235*** (0.091)	0.168** (0.086)
University	0.840*** (0.164)	0.839*** (0.161)	0.728*** (0.152)	0.643*** (0.148)
Training (months)	0.009*** (0.003)	0.010*** (0.002)	0.008*** (0.002)	0.007*** (0.002)
Current Job Tenure (months)	0.004** (0.002)	0.004** (0.002)	0.005** (0.002)	0.003* (0.002)
<b>Job Location:</b>				
Dar-es-Salaam	0.358*** (0.115)	0.307*** (0.115)	0.300*** (0.111)	0.291** (0.123)
Dodoma	-0.389*** (0.117)	-0.353*** (0.113)	-0.323*** (0.113)	-0.262** (0.115)
<b>Employment Sector:</b>				
Public	§	-0.140* (0.079)	-0.107 (0.079)	-0.094 (0.078)
Health & Education	§	-0.231*** (0.083)	-0.193*** (0.081)	-0.179** (0.082)
<b>Parental Background:</b>				
Father – Secondary Education	§	§	0.170** (0.086)	0.062 (0.082)
Father – Higher Education	§	§	0.525*** (0.110)	0.365*** (0.107)
<b>School Effects:</b>	<i>No</i>	<i>No</i>	<i>No</i>	<i>Yes</i>

Table 4 (Cont' d)

<b>Selection Correction</b>	-0.044 (0.121)	-0.067 (0.111)	-0.075 (0.111)	-0.134 (0.114)
$\hat{\sigma}$	0.701	0.686	0.661	0.630
Pseudo-R <sup>2</sup>	0.106	0.115	0.133	0.156
Pseudo Log-Likelihood	-637.6	-631.5	-618.5	-601.9
Homoscedasticity ~ $\chi_k^2$	39.189*** (0.000)	45.555*** (0.000)	35.977*** (0.000)	60.676*** (0.000)
Normality ~ $\chi_2^2$	2.204 (0.332)	6.088** (0.048)	3.303 (0.192)	4.789* (0.091)
Observations	400	400	400	400

Notes to table 4:

- (a) The earnings measure relates to current monthly income.
- (b) The lower and upper bounds used in estimation expressed in Tanzanian shillings (000s) are  $[-\infty, 50]$ ,  $[50, 100]$ ,  $[100, 150]$ ,  $[150, 200]$ ,  $[200, 250]$ ,  $[250, 300]$ ,  $[300, 400]$ ,  $[400, +\infty]$ . The natural logarithms of the finite interval boundaries were used in specifying the model's likelihood function.
- (c) The reported estimates are based on the maximum likelihood procedure. The estimated asymptotic standard errors reported in parentheses are based on the Huber (1967) adjustment.
- (d) \*\*\*, \*\* and \* denote statistical significance at the 1 per cent, 5 per cent and 10 per cent level respectively.
- (e) § denotes not applicable in estimation for given model.
- (f) For the heteroscedasticity test the degrees of freedom are k and are determined by the number of explanatory variables in each specification.
- (g) The values in parentheses beneath the diagnostic test values are the significance levels of the tests.
- (h) The Pseudo-R<sup>2</sup> is based on the McFadden R<sup>2</sup> formula.

Table 5: Interval Regression Estimates for Earnings Equation – Self-employed

<b>Variables</b>	<b>(1)</b>	<b>(2)</b>	<b>(3)</b>	<b>(4)</b>
<b>Constant</b>	11.717*** (0.332)	11.537*** (0.352)	11.515*** (0.353)	11.668*** (0.426)
Cohort_90	-0.251 (0.220)	-0.279 (0.223)	-0.270 (0.225)	-0.225 (0.216)
Male	0.184 (0.194)	0.219 (0.195)	0.226 (0.194)	0.195 (0.232)
Muslim	-0.063 (0.177)	-0.062 (0.176)	-0.079 (0.177)	-0.156 (0.173)
<b>Human Capital:</b>				
Senior Secondary	0.076 (0.199)	0.157 (0.205)	0.118 (0.207)	0.157 (0.202)
University	0.832** (0.347)	0.787* (0.344)	0.715* (0.377)	0.648*** (0.366)
Training (months)	0.007 (0.005)	0.005 (0.006)	0.004 (0.006)	0.003 (0.005)
Wage Employment (months)	0.021*** (0.006)	0.022*** (0.006)	0.022*** (0.006)	0.020*** (0.006)
Current Job Tenure (months)	0.009** (0.003)	0.009*** (0.003)	0.009*** (0.003)	0.007** (0.003)
<b>Job Location:</b>				
Dar-es-Salaam	0.085 (0.211)	0.120 (0.211)	0.114 (0.210)	-0.029 (0.260)
Dodoma	-0.159 (0.254)	-0.150 (0.253)	-0.188 (0.256)	-0.361 (0.292)
<b>Self-Employment Type:</b>				
Employees	§	0.306* (0.163)	0.270* (0.162)	0.215 (0.162)
<b>Parental Background:</b>				
Father – Secondary Education	§	§	0.098 (0.163)	0.007 (0.168)
Father – Higher Education	§	§	0.401 (0.319)	0.297 (0.316)
<b>School Effects:</b>	<i>No</i>	<i>No</i>	<i>No</i>	<i>Yes</i>

Table 5 (cont'd)

Selection Correction	-0.433** (0.192)	-0.448** (0.197)	-0.459** (0.203)	-0.458** (0.192)
$\hat{\sigma}$	1.053	1.043	1.038	0.995
Pseudo Log-Likelihood	-453.3	-451.5	-450.6	-442.5
Pseudo-R <sup>2</sup>	0.028	0.032	0.033	0.051
Homoscedasticity ~ $\chi_k^2$	21.432** (0.029)	22.395** (0.033)	24.104* (0.045)	42.947** (0.007)
Normality ~ $\chi_2^2$	5.000* (0.082)	6.895** (0.032)	6.458* (0.040)	14.558*** (0.001)
Observations	211	211	211	211

Notes to table 5:

- (a) The earnings measure relates to self-employed earnings received in the last six months.
- (b) The lower and upper bounds used in estimation expressed in Tanzanian shillings (000s) are  $[-\infty, 100]$ ,  $[100, 150]$ ,  $[150, 200]$ ,  $[200, 250]$ ,  $[250, 300]$ ,  $[300, 350]$ ,  $[350, 400]$ ,  $[400, 500]$ ,  $[500, 800]$ ,  $[800, +\infty]$ . The natural logarithms of the finite interval boundaries were used in specifying the model's likelihood function.
- (c) The reported estimates are based on the maximum likelihood procedure. The estimated asymptotic standard errors reported in parentheses are based on the Huber (1967) adjustment.
- (d) \*\*\*, \*\* and \* denote statistical significance at the 1 per cent, 5 per cent and 10 per cent level respectively.
- (e) § denotes not applicable in estimation for given model.
- (f) For the heteroscedasticity test the degrees of freedom are k and are determined by the number of explanatory variables in each specification.
- (g) The values in parentheses beneath the diagnostic test values are the significance levels of the tests.
- (h) The Pseudo-R<sup>2</sup> is based on the Mc-Fadden R<sup>2</sup> formula.



Table 6: Deviations from Average – School Effects

<b>School</b>	<b>Wage Employees</b>	<b>Self-Employed</b>
<b>School_1</b>	-0.072 (0.157)	-0.134 (0.193)
<b>School_2</b>	0.467*** (0.113)	0.469*** (0.238)
<b>School_3</b>	0.139 (0.114)	0.063 (0.419)
<b>School_4</b>	0.125 (0.105)	0.642 (0.519)
<b>School_5</b>	-0.476*** (0.123)	-0.742*** (0.227)
<b>School_6</b>	-0.165 (0.131)	-0.001 (0.284)
<b>School_7</b>	-0.086 (0.145)	-0.209 (0.202)
<b>School_8</b>	-0.036 (0.142)	0.236 (0.261)
<b>School_9</b>	-0.028 (0.093)	0.124 (0.178)
<b>School_10</b>	-0.019 (0.099)	-0.152 (0.328)

Notes to table 6:

(a) See expression [7] in the text for a description of the calculations.

(b) \*\*\*,\*\* and \* denote statistical significance at the 1 per cent, 5 per cent and 10 per cent level respectively using two-tailed tests.

Table 7: Rates of Return to Educational Qualifications (%)

<b>Qualification</b>	<b>Wage Employees</b>	<b>Self-Employed</b>
Senior	8.8**	8.2
Secondary	(4.7)	(10.9)
University	17.1***	17.8
	(5.2)	(14.5)

Notes to table 7:

(a) The estimates used for this table are based on specification (4) for both models as reported in tables 4 and 5.

(b) See expressions [8] and [9] in the text for the computational details. The estimated effects are exponentiated and the corresponding standard errors are computed using the delta method.

(c) \*\*\*,\*\* and \* denote statistical significance at the 1 per cent, 5 per cent and 10 per cent level respectively using two-tailed tests.

Table A1: Variable Description

Variable Names	Description of Variables
Cohort_90	= 1 if individual left school in 1990; = 0 otherwise.
Male	= 1 if individual is male; = 0 otherwise.
Muslim	= 1 if individual is Muslim; = 0 otherwise.
Junior	= 1 if individual has a junior secondary qualification; = 0 otherwise.
Senior Secondary	= 1 if individual has a senior secondary qualification; = 0 otherwise.
University	= 1 if individual has a university qualification; = 0 otherwise.
Training (months)	The number of months of full-time further education and training an individual has undertaken since completion of junior secondary school.
Employees	=1 if the self-employed individual employs at least one employee; =0 otherwise.
Current Job Tenure (months)	The number of months an individual has been working in their current job
Dar-es-Salaam	= 1 if individual's job is located in Dar-es-Salaam; = 0 otherwise.
Dodoma	= 1 if individual's job is located in Dodoma; = 0 otherwise.
Public	= 1 if the individual works in the public sector; = 0 otherwise.
Health & Education	= 1 if the individual works in the health and education sector; = 0 otherwise.
Other Employment Sector	= 1 if the individual works in any other employment sector; = 0 otherwise.
Father – Primary Education or Less	= 1 if the individual's father had primary education or less; = 0 otherwise
Father – Secondary Education	= 1 if the individual's father had secondary education; = 0 otherwise.
Father – Higher Education	= 1 if the individual's father had higher education; = 0 otherwise.
School_1 to School_10	= 1 if the individual is a graduate of one of the ten secondary schools; = 0 otherwise.
Selection Term	Constructed using estimates from a reduced from multinomial logit model.

Table A2: Summary Statistics

Variable	Wage Employed				Self-employed			
	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
Log(earnings)	11.39	0.87	10.13	13.30	12.23	1.01	10.82	13.71
Cohort_90	0.65		0	1	0.54		0	1
Male	0.48		0	1	0.69		0	1
Muslim	0.25		0	1	0.29		0	1
Junior Secondary	0.58		0	1	0.71		0	1
Senior Secondary	0.29		0	1	0.25		0	1
University	0.13		0	1	0.04		0	1
Training (months)	20.52	18.97	0	96	10.15	16.15	0	74
Current Job Tenure (months)	44.61	28.89	2	126	54.41	31.55	3	132
Period in Waged Employment (months)	<i>f</i>				6.43	16.42	0	122
EMP	<i>f</i>				0.57		0	1
Dar-es-Salaam	0.59		0	1	0.57		0	1
Other Region	0.17		0	1	0.17		0	1
Dodoma	0.24		0	1	0.26		0	1
Public	0.36		0	1	<i>f</i>			
Health & Education	0.31		0	1	<i>f</i>			
Other Employment Sector	0.33		0	1	<i>f</i>			
Father – Primary or less	0.50		0	1	0.62		0	1
Father – Secondary Education	0.31		0	1	0.29		0	1
Father – Higher Education	0.19		0	1	0.09		0	1
School_1	0.08		0	1	0.17		0	1
School_2	0.09		0	1	0.12		0	1
School_3	0.12		0	1	0.07		0	1
School_4	0.12		0	1	0.03		0	1
School_5	0.08		0	1	0.07		0	1
School_6	0.09		0	1	0.09		0	1
School_7	0.07		0	1	0.15		0	1
School_8	0.12		0	1	0.10		0	1
School_9	0.09		0	1	0.13		0	1
School_10	0.14		0	1	0.07			
Selection term	0.26	0.38	0	3.63	0.37	0.58	0.00	3.48
sample size	400				211			

Notes to table A2:

(a) log earnings are the mean of the mid-point of the intervals described in the methodology section and are not actually used in the analysis.

(b) *f* not applicable in estimation.

Table A3: Estimated Impact Effects for Educational Qualification on Sectoral Attachment

	Wage employment	Self employment	Training	Unemployment
Senior secondary	14.7 (9.6)	-8.0 (8.7)	-11.2** (4.5)	4.5 (3.9)
University	25.0* (13.9)	-12.9 (12.8)	-7.6** (3.2)	-4.5 (4.9)

Notes to table A3:

(a) The reported estimates are the change in the probability of sectoral attachment associated with each education level based on the four-category multinomial logit model. These estimates are calculated at the means of all other variables and expressed as percentage points.

(b) Estimated asymptotic standard errors are reported in parentheses.

(c) \*\*\*, \*\* and \* denote statistical significance at the 1 per cent, 5 per cent and 10 per cent level respectively.

(d) In addition to the variables included in the earnings equations the MNL model included controls for education level of the respondent's spouse, time spent unemployed and self-employed since completing junior secondary school, the division achieved on the Form IV secondary school examination, a set of variables based on the interaction of the cohort year with an individual's education level, and variables describing the respondent's duration in different activities since leaving formal education (i.e., wage employment, self-employment, training and unemployment)..

Table A4: Test Results of Educational Qualification Interactions

<b>Senior Secondary and University Variables Interacted with:</b>	$\chi^2_k$
Cohort_90 ~ $\chi^2_2$	3.676 (0.159)
Male ~ $\chi^2_2$	0.029 (0.986)
Muslim ~ $\chi^2_2$	0.479 (0.787)
Job Location ~ $\chi^2_4$	2.713 (0.607)
Parental Background ~ $\chi^2_4$	0.129 (0.998)
Health and Education ~ $\chi^2_2$	11.784*** (0.000)

Notes to table A4:

(a) The interaction terms are constructed by separately interacting the qualification variables with the listed variables above. The interactions are introduced into our preferred specification (4) of table 4.

(b) Given the violation of homoscedasticity, the chi-squared values are Wald tests based on the variance-covariance matrix reported in expression [6].

(c) The prob-values for the tests are reported in parentheses.

(d) \*\*\*, \*\* and \* denote statistical significance at the 1 per cent, 5 per cent and 10 per cent level respectively.

Table A5: Interval Regression Estimates for Earnings Equation – Employees by Sector

Variables	Public Sector Employees	Private Sector Employees	Prob-Value for Test of Coefficient Differences	Public Sector Means	Private Sector Means
<b>Constant</b>	10.368*** (0.283)	10.628*** (0.234)	0.338	1.0	1.0
Cohort_90	-0.109 (0.188)	0.338** (0.141)	0.071	0.762	0.595
Male	-0.110 (0.164)	0.360*** (0.124)	0.033	0.545	0.447
Muslim	-0.075 (0.115)	0.278*** (0.106)	0.031	0.279	0.233
<b>Human Capital:</b>					
Senior Secondary	0.178 (0.124)	0.198* (0.115)	0.918	0.266	0.303
University	0.818*** (0.227)	0.595*** (0.201)	0.415	0.112	0.140
Training (months)	0.009*** (0.003)	0.007** (0.003)	0.464	22.23	19.58
Current Job Tenure (months)	0.009*** (0.003)	0.000 (0.003)	0.025	51.80	40.61
<b>Job Location:</b>					
Dar-es-Salaam	0.389** (0.166)	0.290*** (0.174)	0.627	0.482	0.654
Dodoma	-0.263 (0.166)	-0.228 (0.158)	0.871	0.308	0.202
<b>Employment Sector:</b>					
Health & Education	-0.091 (0.118)	-0.249** (0.113)	0.320	0.434	0.237
<b>Parental Background:</b>					
Father–Secondary Education	0.065 (0.141)	0.071 (0.103)	0.964	0.273	0.331
Father – Higher Education	0.222 (0.165)	0.396*** (0.138)	0.482	0.126	0.229
<b>Selection Correction</b>	0.078 (0.264)	-0.222* (0.129)	0.341	0.173	0.303
$\hat{\sigma}$	0.525	0.658	§	§	§
Pseudo-R <sup>2</sup>	0.171	0.167	§	§	§
Pseudo Log-Likelihood	-190.315	-398.101	§	§	§
Homoscedasticity ~ $\chi^2_{23}$	40.369*** (0.000)	53.784*** (0.000)	§	§	§
Normality ~ $\chi^2_2$	6.231** (0.044)	3.263 (0.196)	§	§	§
Overall Test of Coefficient Differences ~ $\chi^2_{23}$	§	§	0.298	§	§
Observations	143	257	400	143	257

Notes to table A5:

(a) See table 4. The reported regression models are based on specification (4) from table 4.

(b) The prob-values for the tests of coefficient differences are based on individual t-tests from the estimation of an interval regression model using the full sample of employees with a complete set of public sector interactions with all the regressors. The Huber (1967) adjustment is used to compute the variance-covariance matrix and the overall test reported in this table is based on a Wald test using this variance-covariance matrix.