CSAE WPS/2010-02

LEARNING & EARNING IN AFRICA: WHERE ARE THE RETURNS TO EDUCATION HIGH?

Neil Rankin†, Justin Sandefur‡ and Francis Teal‡.

January 2010

Abstract

This paper investigates the role of learning - through formal schooling and time spent in the labor market - in explaining labor market outcomes of urban workers in Ghana and Tanzania. We investigate these issues using a new data set measuring incomes of both formal sector wage workers and the self-employed in the informal sector. In both countries we find significant, convex returns to education and large earnings differentials between sectors when we pool the data and do not control for selection. In Ghana there is a particularly steep age-earnings profile. We investigate how far a Harris-Todaro model of market segmentation or a Roy model of selection can explain the patterns observed in the data. We find highly significant differences across occupations and important effects from selection in both countries. The data is consistent with a pattern by which higher ability individuals queue for the high wage formal sector jobs such that the age earnings profile is convex for the self-employed in Ghana once we control for selection. The returns to education are far higher in the large firm sector than in others and in this sector they are linear not convex. In both countries there is clear evidence of convexity in the returns to education for the self-employed and here the average returns are low.

The data used in this paper were collected by the Centre for the Study of African Economies, Oxford, in collaboration with the Ghana Statistical Office (GSO) and the Tanzania National Bureau of Statistics (NBS). The research, and the surveys on which it is based, has been funded by the Department for International Development (DfID) and the Economic and Social Research Council (ESRC) of the UK and by the IDRC in Canada. We are greatly indebted to numerous collaborators for enabling this data to be collected, particularly Emilian Karugendo and Trudy Owens in Tanzania, and Moses Awoonor-Williams, Geeta Kingdon and Andrew Zeitlin in Ghana. Andrew Kerr provided valuable assistance with the coding. An earlier version of this paper benefited from the input of seminar participants in IZA/Berlin and Cornell. We have discussed the points made in this paper extensively with Måns Söderbom who has offered many valuable suggestions. All errors are ours.

†School of Economic & Business Sciences, University of the Witwatersrand, South Africa
‡Centre for the Study of African Economies, Oxford University

1 Introduction

How incomes can be increased in Africa is clearly the central issue that has to be addressed in understanding how poverty there can be reduced. In this paper we seek to build on three broad elements in understanding links between labor market outcomes and incomes. The first, and probably most intensively researched, is the role of education. The second is the tradition of dual economy models of labor markets which date from early work on surplus labor, summarised and clarified in Sen (1966), and the model of rural-urban migration set out by Harris and Todaro (1970). The third is a multi-market Roy (1951) model of earnings and selection, applications of which have been mainly confined to developed countries, Dahl (2002) . Mortensen (2005) provides a review of recent empirical work, mainly for developed economies, seeking to understand sources of wage dispersion of a similar kind to which we will observe in our two African countries.

That increases in education is the key to raising incomes and reducing poverty is widely believed. Early research reviewed by Psacharopoulos (1994) showed that the significant, robust, positive relationship between education and earnings is nearly universal across middle and low income countries. More recent work by Schultz (2005) draws on data for nine African countries demonstrating that the return to schooling is highly convex, with an extremely modest private economic return at the primary level and rising to over 20% at the secondary level. Both Psacharopoulos and Schultz interpret their results as an endorsement of universal primary education (UPE) as an instrument for poverty alleviation.

In contrast to the focus on human capital in the modern earnings literature, an earlier generation of labor market analysis in developing countries completely ignored the role of skills in income determination. This approach, borne out of the African experience, focused instead on wage-setting institutions and the earnings differentials between formal and informal or urban and rural sectors. Institutionally imposed wage floors in the urban, formal sector were the starting point for Harris and Todaro's (1970) influential model and the large number of papers that followed in its wake. This formal sector wage floor came to be seen as the fundamental market distortion in African economies by many analysts in the 1970s and `80s (Jamal and Weeks (1993)) and played a central role in Bates' (1981) path-breaking analysis of African political economy.

The third element on which we propose to draw is work primarily in developed countries identifying the importance of both unobservable firm and individual characteristics in determining earnings and the processes of selection which may occur as a result of such unobservables. Two such variables which have played a major part in the human capital

explanation of earning differences are the role of unobserved skill and ability. Some of these unobserved skills can, it has been argued, be inferred indirectly from evidence as to how work experience affects earning. Following Becker (1964) the increasing, concave relationship between age and earnings is interpreted as the diminishing marginal return to general skills acquired through work experience. Returns to job tenure reflect job-specific skills that pay a premium within a given job and have strong implications for job movement. Taken literally, the age-earnings profiles which we document below for Ghana suggest that poverty is, on average, a transitory phenomenon from an individual perspective. In Ghana the average earnings of a young person aged 16 to 20 are about US$24 per month, compared to US$49 for those aged over 30, a doubling of income in little more than a decade. How misleading are these cross-section estimates of the implied return on work experience?

In contrast to these partially observed dimensions of skills is the possible role played by unobserved ability. High ability types may both we able to acquire more observable human capital and earn more. The central focus of this paper is the implications for both the age-earning profile and for the interpretation of occupational earnings differentials of the existence of different types in a segmented labor market. We use the basic Harris-Todaro framework but allow for both skills and unobserved high ability types. We ask how large are the earnings differentials between occupational groups with, and without, controls for observable and unobservable skills. The skills we consider may also be specific to the occupation in which the workers is employed as in the Roy (1951) model.

To test the model - i.e., to understand the interaction of human capital and wage-setting institutions in producing income differences - requires income data on individuals across skill classes in both the formal and informal sectors. We know of no existing work using such data for Africa, and very little work elsewhere. Rather, current understanding of earnings determination in low-income economies is based on a highly unrepresentative sample of the labor force: urban wage employees. In his review of the returns to health and education in Africa, Schultz (1999) acknowledges this is a problem, noting that his estimates based on wage workers from Ghana are informative for approximately 10% of the total labor force.

This paper presents new survey data specifically designed to address this gap. Like many existing data sets, the survey collects a wide range of socioeconomic data from workers in urban areas of two African countries: Ghana and Tanzania. The survey is novel in collecting income data on the self-employed in the urban informal sector. This data is intended to be comparable to wage data for urban employees - a hypothesis which is testable up to a point. In both Ghana and Tanzania, the self-employed outnumber wage employees by nearly 2-to-1 in the urban labor

force, and constitute the fastest growing segment of the labor force across rural and urban areas. Understanding earnings determination in this sector, and how workers sort between sectors is thus central to any understanding of labor markets and income distribution in these countries.

The following section presents our modeling framework and section 3 our data. The empirical strategy for addressing selection issues is set out in section 4 and the earnings functions allowing for both the endgeneity of education and selection are presented in Section 5. A final section concludes.

2 Modeling African Labor Markets

In 1970 John Harris and Michael Todaro published a simple two sector, static model of rural-to-urban migration. The model set out to explain how throughout the 1960s rural-to-urban migration continued unabated despite high levels of open unemployment in the capital city. Furthermore, attempts by the Kenyan government to calm political pressure by launching public employment initiatives appeared to exacerbate urban unemployment by drawing in new job-seekers from the countryside. This phenomenon became known as the Harris-Todaro Paradox.

Harris and Todaro's (1970) model depicted potential migrants maximizing expected earnings by weighing a guaranteed income in the rural sector against a lottery for high wage, formal sector jobs in the city. In the years that followed, numerous generalizations of the basic Harris-Todaro framework appeared in the literature. Bardhan and Udry (1999) outline a few of these variants to the model. The version published by Fields (1975) allowing for simultaneous coexistence of open unemployment in the city and an urban informal sector provides the basic setup we use below. While we abandon the focus on migration, our analysis builds on, and proposes to test, several of the institutional assumptions that define the Harris-Todaro framework.

The focus in the Harris and Todaro (1970) and Fields (1975) models is on the role of unemployment (or informality) in establishing the equilibrium between the formal and the informal sector. In the simplest version of these models labour is homogeneous. However in taking the model to the data the nature of the differences across labour in terms of their skills is clearly a crucial element in establishing how far markets, for given skill levels, are segmented. One observable element of these differences is the level of formal education and labour market experience.

The traditional approach to estimating the earnings function has been through seeking instruments for education, an approach we intend to follow. We model occupation sorting using a

latent variable approach. Here the latent variable is the propensity to sort into a given sector. Only the occupation outcome and not the underlying propensity are actually observed:

[1] $$y_j^* = z\gamma_j + \eta_j \quad \text{where } j = (f, i, n)$$

In the empirical analysis of the next section we identify five sectors, the self-employment, wage work in small and large firms, public sector employment and not having a job. Theoretical models usually make a distinction between the formal and informal sector assuming the informal sector is open access and thus does not need to be distinguished from unemployment. However it is an empirical issue as to what constitutes the informal sector. Is there free entry to the informal sector and self-employment, if so does this market clear? By allowing these to be sectors separately identified in the data we can address that question.

While both education and age are observable there are key characteristics of both the individual and the job which, in the Roy (1951) model, can be observed by the individual but not by us. These include the unobserved attributes of the formal and informal sector jobs. There may be some forms of attributes which are well rewarded in the informal sector but not the formal, for example initiative, which those working in the formal sector may have attributes that pay little or nothing in the informal sector. We assume these factors will appear in the residuals of equation [1] - η_j .

The outcome variable y_s is observed if and only if category *s* is chosen which happens when

[2] $$y_s^* > \max_{j \neq s}(y_j^*)$$

Given this sorting we will observe earnings in sector *s*, by definition no earnings accrue to the unemployed. If we define

[3] $$\varepsilon_s = \max_{j \neq s}(y_j^* - y_s^*) = \max_{j \neq s}(z\gamma_j + \eta_j - z\lambda_s - \eta_s)$$

Then under this definition condition [2] is equivalent to $\varepsilon_s < 0$

[4] $$y_s = x\beta_s + u_s \text{ where } s \text{ is the sector which the individual chooses.}$$

[5] $$E(y_s|x, s_i = 1) = \beta x_i + E(u_i|x_i, s_i = 1)$$

[6] $$E(y_s|x, s_i = 1) = \beta x_i + E(u_i|x_i, \varepsilon_s < 0)$$

The modeling issue that arises is how u_i and ε_s are assumed to be correlated. Following McFadden (1973) if it is assumed that the $(\eta_j)s$ are independent and identically Gumbel distributed (the so-called IIA hypothesis) then their cumulative and density functions are

respectively: $G(\eta) = \exp(-e^{-\eta})$ and $g(\eta) = \exp(-\eta - e^{-\eta})$. This specifications leads to the multinomial logit model with:

[7] $$P(\varepsilon_s < 0 \mid z) = \frac{\exp(z\gamma_s)}{\sum_j \exp(z\gamma_j)})$$

On the basis of this expression, consistent maximum likelihood estimates of the $(\gamma_j)s$ can be easily obtained. The problem posed by this selection process is to obtain unbiased estimates of the parameters of the income equation [4]. In general the disturbance term in this equation may not be independent of all the $(\eta_j)s$.

Define Γ as follows: $\Gamma = (z\gamma_1, z\gamma_2, ..., z\lambda_M)$

Following Bourguignon, Fournier and Gurgan (2007) we use the fact that given that the relation between the *M* components of Γ and the *M* corresponding probabilities is invertible then we can obtain consistent estimates of the β_s from either of two regressions:

[8a] $$y_s = x\beta_s + \mu(P_{1,}.....,P_M) + w_s$$

[8b] $$y_s = x\beta_s + \lambda(\Gamma) + w_s$$

where w_s is a residual that is mean-independent of the regressors.

The most commonly used approach to dealing with selection in a model of this kind is that of Lee (1983) who first proposed a generalisation of the two-step selection bias correction method introduced by Heckman (1979).

Call $F_{\varepsilon_s}(\cdot \mid \Gamma)$ the cumulative distribution of ε_s .The cumulative $J_{\varepsilon_s}(\cdot \mid \Gamma)$ defined by the following transform: $J_{\varepsilon_s}(\cdot \mid \Gamma) = \Phi^{-1}(F_{\varepsilon_s}(\cdot \mid \Gamma)$ where Φ is the standard normal cumulative, has a standard normal distribution. Lee's model can be written as implying that:

[9] $$E(u_s \mid \varepsilon_s, \Gamma) = \sigma\rho_s . J_{\varepsilon_s}(\varepsilon_s \mid \Gamma)$$

and yields an estimating equation of the form:

[10] $$y_s = x_s\beta_s - \sigma\rho_s \frac{\phi(J_{\varepsilon_s}(0 \mid \Gamma)}{F_{\varepsilon_s}(0 \mid \Gamma)} + w_s$$

It is this equation which we will use in reporting the results which allow for selection by Lee's method.

As discussed by Bourguignon, Fournier and Gurgan (2007) the Lee approach makes restrictive assumptions as to how the errors across the selection and outcome equations are related. Lee's assumptions imply that the covariance of the errors in the earnings equation for sector *s* and the errors in the occupational sectors are all of the same sign (see Schmertmann

(1994)). For example, let us hypothesise that ability *A* makes it less likely that the individual will be unemployed and, conditional on having a job in the formal sector, that *A* will boost their earnings. This sets up a negative correlation between the unobservable earnings in the formal sector and the unobservable propensity to have no job. Lee's assumptions for addressing the selection issue imply that all the correlations with the other sectors be negative too. Indeed in the form of the model set out in equation [10] they are all assumed to be equal as there is only one parameter ρ_s.

This is clearly restrictive and in this paper we compare the Lee method with that proposed by Dubin and McFadden (1984) which allows these correlations to differ. As we will show this relaxation of the Lee's assumption is important for our empirical results.

The basis for correcting sample selection proposed by Dubin and McFadden (1984) is an equation of the form:

[11] $$E(u_s \mid \eta_1 .....\eta_M) = \sigma \frac{\sqrt{6}}{\pi} \sum_{j=1...M} r_j(\eta_j - E(\eta_j))$$

where r_j is the correlation coefficient between u_s and η_j. Bourguignon, Fournier and Gurgan (2007, p. 178) summarise how this leads to an equation which can be estimated.

[12] $$y_s = x_s\beta_s + \sigma \frac{\sqrt{6}}{\pi} \sum_{j \neq s} \left[r_j \left(\frac{P_j ln(P_j)}{1-P_j} \right) - r_1 ln\,(P_1) \right] + w_s$$

Dubin and McFadden (1984) in addition to [12] assume that $\sum_{j=1...M} r_j = 0$. We do not. Equation [12] will be used in the sample correction equations to relax the Lee assumptions.

The x_s vector in [12] contains the observable elements of human capital. The two of which we propose to focus are *Education* and *Age*. There has been a very large literature concerned with addressing the problem posed by the endogeneity of education.

If there is sorting along the lines hypothesised by a Roy model then in principle all the results of the OLS earnings function will be biased. In particular any finding of sectoral differences in earnings between the formal and informal sector need not imply that a Harris-Todaro model is relevant for understanding these differences, they may arise from a Roy sorting model. If there is extensive "waiting" and those who wait longest are those who have high ability then such a model sets up a selection process by which the cross-section age-earning profile is simply an artifact of the queuing for formal sector jobs.

In the original context for the Harris-Todaro model the formal sector was the public sector where both wages and employment were set exogenously. We have shown in another paper

based on this data, Sandefur, Serneels and Teal (2006), that size is an important correlate of earnings among both the wage and the self-employed. The data suggest not a clear line between formality and informality but a spectrum from low formality for the own self-employed and wage employees in small firms to increasing levels of formality as enterprise size rises with the highest earnings being in the public sector. We set up the data so this potential dimension of formality can be investigated.

3 The Data

Our data is taken from a longitudinal labor market survey conducted by the Centre for the Study of African Economies (CSAE) at Oxford University, under the direction of the authors and in collaboration with the Ghana Statistical Office (GSO) and the Tanzania National Bureau of Statistics (NBS). We refer to the data set as the Ghana and Tanzania Urban Panel Survey (UPS). The survey collects information on incomes, education and labor market experience, household characteristics and various other modules for labor force participants (ages 15 to 60) in urban areas. For Ghana these areas span the four largest urban centers in the country: Accra (and neighboring Tema), Kumasi, Takoradi and Cape Coast. In Tanzania, the sample covers several of the largest urban areas including Arusha, Dar es Salaam, Iringa, Morogoro, Mwanza, and Tanga.

The samples were based on a stratified random sample of urban households from the 2000 census in Ghana and the 2000 Household Budget Survey (HBS) in Tanzania.[1] While the initial sample was household based, interviews were conducted on an individual basis, and the unit of analysis in most of what follows will be at the individual level. A total of 830 and 543 individuals were interviewed in the first round of the survey in Ghana and Tanzania respectively, which was conducted between October 2003 and June 2004.

A features of the UPS data which is important for answering the questions posed in this paper is that it provides comparable information, including income data, on both wage employees and the self employed. All labor force participants in the selected households were to be interviewed. Thus the sample of workers spans the formal and informal sectors, public and private employees, the self-employed, unemployed and so on.

[1] For the narrow group of readers familiar with the CSAE's survey work in Ghana, we should note that the analysis in this paper does *not* incorporate data from the Ghana Manufacturing Enterprise Survey (GMES). The UPS and the GMES are conducted in parallel with a common survey instrument. However, we restrict ourselves in this paper to the population based sample of the UPS, excluding the firm-based sample of manufacturing employees interviewed through the GMES.

Collecting income data on the self-employed in low-income countries is a controversial endeavor. Field guides for the World Bank's Living Standards Measurement Surveys (LSMS), which serve as the international standard for household surveys in development economics, recommend survey managers not collect this information. The stated rationale is that self-employed business people in the informal sector rarely keep written accounts and their self-reported income data may be too noisy to be of use. For household based enterprizes, the distinction between business and personal expenditures may be completely alien to respondents. We acknowledge the validity of these concerns.

However, because the non-agricultural self-employed constitute a majority of the urban working population in both Ghana and Tanzania, we feel the incomes in this sector are too important to ignore. Our income measure for the self-employed is based on self-reported profits. Profits are net of routine operating expenses and gross of fixed capital expenditure, if any. The concepts of ``revenue", ``business costs", and ``profits" are explained to respondents by enumerators with experience in conducting firm and household surveys. As the surveys are entered directly onto handheld computers, a simple mechanical check forces enumerators to go over the numbers again if revenue, cost and profit figures are inconsistent. While enumerators have reported few conceptual difficulties with this portion of the questionnaire, we feel a better test of the validity of self-reported income data is our ability to explain its variation with personal, household and business characteristics, as we attempt to do below.

During the course of July-August 2005 the initial UPS sample was resurveyed and questions were asked which enabled us to link their activities and earnings in 2005 with the same variables in 2003/04, creating a panel of individual workers. The data that will be used in this paper is for both the 2004 and 2005 rounds of the survey.

Table 1 sets out how large are the differences in earnings across the formal and informal sectors within these economies and how closely these are related to education levels. In both Ghana and Tanzania median earnings are higher in self-employment than in the small firm sector (a small firm is one employing less than 10 workers) and the public sector pays substantially more than even the large firm sector. Earning differentials across what might be termed the informal (ie small firm) sector and the formal (public sector workers) appear large in both countries, a differential of nearly 3 times. Table 1 also shows that the public sector worker has a substantially higher level of education than the other sectors, particularly in Tanzania where the median levels for all other sectors is 7 years of education while for the public sector it is 12. How large are these differences across occupation once we control for observable human capital in the form of education and age (as our proxy for work experience) and the endogeneity of education?

In Table 2 we report summary statistics of the variables we are going to use as instruments for education. These are the distances to the nearest primary and secondary schools when the individual being interviewed were at school and the education and occupations of their parents.

We use these instruments to predict education in the first two columns of Table 3. We then ask how much of the occupational differences we observed remain after we control for human capital and allow for the endogeneity of education by means of the residuals from the education equation (these residuals are the control variable reported in Table 3.) The results with the control variable are reported for Ghana in Column (4) and for Tanzania in Column (6). The results show that while the differential is reduced it remains substantial. In Ghana the differential between being in the public sector and working in a small firm is 1.8 times, in Tanzania the differential 2.4 times. In both Ghana and Tanzania the data appear to show a clear pattern by which those in small firms earn the least and those in the public sector the most.

While Table 3 controls for the endogeneity of education it does not allow for Roy selection. To address that issue our empirical strategy has two steps. First, we estimate the determinants of the job search decision, distinguishing between formal employment (in the public or private sector), informal employment (either self-employed or in a small or micro enterprize) and unemployment. We model occupational choice using a multinomial logit with plausible determinants of expected earnings and the reservation wage as independent variables. Second, we estimate the determinants of earnings in the cross section. We report the results of using both the Lee (1983) and Dubin and McFadden (1984) correction terms for selection.

4 The Empirical Strategy for Selection

What determines occupational choices? In the model presented in section 2 occupation choices depend on the probability of acquiring a good job and the relative earnings in each sector, including the reservation wage in unemployment. Following a long tradition in estimating labor supply decisions, we use estimate occupational choices using variables thought to predict earnings in each occupation. In theory, these variables might differ for each sector (public sector employment, formal sector private employment, etc.). In practice, we divide the explanatory variables into just two categories. The first category is a vector of personal characteristics thought to predict earnings conditional on employment in any sector, including gender, age and education. The second category is a vector of personal and family characteristics that will arguably determine the reservation wage, including marital status, number of children, status as household

head, mother and father's education level, and occupation codes for the respondent's mother and father.

We present the results of this multinomial logit for education and age in Table 4 (the actual equation is reported in the Appendix). For both countries the probability of working in the public sector increase with the level of education but this effect is much greater in Tanzania than in Ghana. In Ghana the probability of working in a large firm also increases with education but this is not true in Tanzania. In both countries the probability fo working as either self-employed or in a small firm decreases with education. However self-employment dominates all forms of employment below the level of university.

By far the most important aspect of age in Ghana is the rise in the probability of being in self-employment and the fall in the probability of not having a job. For Tanzania it is the rise with age in the probability of being in the public sector and the fall in the probability of being in the small firm sector. Clearly both these patterns are consistent with a pattern of "waiting" by which there is a preference for a public or large firm job and as the probability of obtaining one declines the worker switches to the informal sector. If so then selection into the sector will be an important determinant of earnings there.

5 The Empirical Strategy for Earnings

In this section we set out earnings functions for Ghana and Tanzania for both wage and self-employment. We begin in Table 5 by asking how the earnings function differs across occupation with no controls either for the endogeneity of education or for occupational selection. It is clear that there are radical changes from the pooled regression in both the age earnings profile and the returns to education. In Ghana the concave age earnings profile is confined to the private sector while the convex education earnings function is confined to the self-employed and those working in small firms. In Tanzania the convex earnings function is confined to the self-employed and those in large firms. In large firms the return on age is convex, not concave as in the pooled regression. Clearly the results we reported in Table 3 are misleading as to how education impacts on earnings within sectors.

The results in Table 5 are consistent with the returns to human capital being sector specific but they do not control for either the endogeneity of education nor the possibility of Roy occupational sorting. That we do next in Tables 6 for Ghana and in Table 7 for Tanzania. In both Tables we report results across the four occupational sectors we have identified and for each

sector we report the results of using the controls for selection due to Lee (1883) and Durbin and McFadden (1984) (DMF).

The results in Table 6 for Ghana suggest that Roy selection is a very important aspect of the determinants of income in the economy. For the self-employed the concave age earnings profile shown in Table 5 disappears with the controls for selection in Table 6. These results hold whether we model selection by the Lee term as in Table 6 Column (1) or those due to DMF Column (2). In fact selection matters in a way that is informative as to the nature of choices available in the economy. Both Lee and DMF imply a negative correlation between the unobservables that induced movement into the sector and earnings when engaged there. How if self-employment is the preferred sector can this correlation be negative? Clearly it is possible if self-employment becomes the preferred sector over time as there is a period of waiting for a high paid job in the formal sector defined here as either the large firm or public sector. Further the DMF selection terms by being more general than those of LEE identify a positive correlation between earnings in self-employment and the unobservables determining the choice of work in the public sector. The reason that the public sector is not "chosen" is that employment there is rationed and, the results imply, these abilities are productive in the self-employment sector.

Not only selection but the endogeneity of education matter in self-employment in Ghana. The returns to education remain convex but the degree of convexity becomes more pronounced relative to the results with no controls as reported in Table 5. The implications of such convexity is that the returns to education for those with low levels is lower than that to those with more education. However in Table 6 we see how misleading were the results of the pooled regressions in Table 3. Using the DMF selection correction terms we see that the returns to education in the large firm sector are linear and large. The average return to education there is 34 per cent (Table 6, Column (6) which compares with 4.7 per cent for the self-employed (column (2) and 6 per cent for those in small firms (Column (4). There ae no significant returns to human capital in the public sector.

The results in Table 7 for Tanzania differ in several respects to those of Ghana. The pooled regression for Tanzania reported in Table 3 did not identify a concave age earnings function while the disaggregation reported in Table 5 showed a significant convex one for large firms. Selection does not alter this result. Selection is however a significant factor in assessing the returns to education in Tanzania, convexity is confined to the self-employed sector. The evidence here too is consistent with "waiting" for formal sector jobs being an important aspect of the selection process. In Table 7 Column (2) we see that here is a highly significant DMF seelction term on the no job occupational outcome which implies a positive correlation between the

unobservables in self-employment earnings and the outcome for that sector. This is consistent with those with higher reservation wages, and more market ability, waiting longer before they enter the self-employed sector. Returns to education in the public sector are actually negative, consistent with both compressed differentials within that sector and high rents which induce searching for those jobs.

6 Conclusions

Where then are the returns to education high? In Ghana we find that controls for the endogeneity of education and for selection imply that the convexity of the returns to education are confined to the self-employed and small firm sector, although within these sectors the degree of convexity increases with the controls. For the large firm sector the returns to education are very large (on average in excess of 30 per cent) and appear to be linear. These conclusions depend on using the Durbin and McFadden (1984) controls. Within the public sector we find no return to education, a result that does not depend on how we adjust for selection. Focusing on the results using the DMF controls we see for Ghana that the concave age earning profile only holds in the small firm sector. In fact the profile is convex in the other sectors.

For Tanzania we find that the convex return to education is confined to the self-employed. Our sample sizes for small and large firms are small but we find for large firms a strongly concave return to education, a result which does not depend on how selection is modeled but does depend on selection. Also, as in Ghana, we find no return to education in the public sector. We noted that the concave age earning profile in Tanzania was not nearly as sizable or well identified as that in Ghana. Once we control for selection we find it is negative and convex for the large firm and public sectors. Selection appears a very important aspects of these labour markets and to operate in a way that is consistent with the kind of market segmentation hypothesized by Harris and Todaro. High rent sector motivate search which influences the time individuals enter specific sectors rendering both the age earnings profile and returns to education on pooled regressions very misleading.

References

Bardhan, P. and C. Udry (1999) *Development Microeconomics*, Oxford University Press.

Bates, R. (1981) *Markets and States in Tropical Africa*, Berkeley: University of California Press.

Becker, G. S. (1964): *Human capital: A theoretical and empirical analysis, with special reference to education*, First edition. Columbia University Press, New York.

Borjas, G. J. (1987): "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77(4), 531–53.

Bourguignon, F., M. Fournier and M. Gurgand (2007) "Selection bias corrections based on the multinomial logit model: Monte Carlo simulations", *Journal of Economic Surveys*, Vol. 21, No.1, 174-205.

Durbin, J.A. and D. L. McFadden (1984) "An econometric analysis of residential electric appliance holdings and consumption", *Econometrica*, 52: 345-362.

Fields, G. (1975) "Rural-urban migration, urban unemployment and underemployment and job search activity in LDCs" *Journal of Development Economics*, 2, 165-188.

Harris, J. and M. Todaro (1970) "Migration, unemployment and development: A two sector analysis", *American Economic Review*, 60 (1).

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

Heckman, J. J., and B. E. Honore (1990): "The Empirical Content of the Roy Model," *Econometrica*, 58(5), 1121–49.

Jamal, V. and J. Weeks (1993) *Africa Misunderstood, or, Whatever happened to the rural-urban gap?* Macmillan Press, London.

Jolliffe, D. (1998): "Skills, Schooling, and Household Income in Ghana," *World Bank Economic Review*, 12(1), 81–104.

Kingdon, G., J. Sandefur, and F. Teal (2006): "Labour market flexibility, Wages and Incomes in Sub-Saharan Africa in the 1990s," *African Development Review*, Vol. 18, No. 3, December 2006.

Lee, L.F. (1983) "Generalized econometric models with selectivity", *Econometrica*, 51: 507-512.

McFadden, D. L. (1973) "Conditional logit analysis of qualitative choice behavior", In P. Zarembka (ed.), Frontiers In Econometrics, New York, Academic Press.

Mortensen, D.T. (2005) *Wage Dispersion: Why are similar workers paid differently?*, The MIT Press, Cambridge, Massachusetts.

Psacharopoulos, G. (1994). 'Returns to investment in education: a global update,' *World Development*, Vol. 22, pp. 1325-1343.

Psacharopoulos, G. and Patrinos, H. A. (2002). 'Returns to investment in education: a further update,' World Bank Research Working Paper 2881, The World Bank.

Roy, A. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3, 135 – 146.

Sandefur, J. P. Serneels and F. Teal (2006) "African poverty through the lens of labor economics: Earnings & mobility in three countries" GPRG-WPS-060, CSAE Oxford.

Schmertmann, C.P. (1994) "Selectivity bias correction methods in polychotomous sample selecton models, *Journal of Econometrics*, 60: 101-132.

Schultz, T.P. (1999) "Health and schooling investment in Africa", *Journal of Economic Perspectives*, 13 (3), 67-88.

Sen, A.K. (1966) "Peasants and dualism with or without surplus labour", *Journal of Political Economy*, 74.

Söderbom, M., Teal, F. and A. Wambugu (2005) "Unobserved heterogeneity and the relation between earnings and firm size: Evidence from two developing countries" *Economic Letter,* 87, 153-159.

Table 1 Earnings and Education in Urban Ghana and Tanzania: 2004-2005

	Ghana			Tanzania		
	Ln (Monthly Earnings in 1995 prices)	Monthly Earnings in US$	Education in Years	Ln (Monthly Earnings in 1995 prices)	Monthly Earnings in US$	Education in Years
	Mean	Median	Median	Mean	Median	Median
Self Employment	10.92	45 (754)	9	5.43	39 (618)	7
Small Firm	10.67	34 (157)	9	5.12	33 (72)	7
Large Firm	11.47	72 (123)	11	5.80	57 (69)	7
Public	11.77	98 (97)	14	6.20	98 (151)	12
Not Employed	NA	NA	10 (509)	NA	NA	7 (61)
Total	11.02 (1131)	45 (1131)	10 (1640)	5.57 (910)	47 (910)	7 (971)

Figures in ( ) are the number of observations.

Table 2 Determinants of Education in Ghana and Tanzania: 2004-2005

	Kdistprim	Kdistsec	Momeduc	Dadeduc	Momformal	Dadformal
	Distance to Primary school when 6	Distance to Secondary School when 16	Mother's education	Father's education	Dummy = 1 if Mother had a formal sector job	Dummy =1 if father had a formal sector job
Ghana						
Median	2	4.5	3	12	0	0
Mean	1.92	5.58	5.17	8.50	0.06	0.37
N = 1640						
Tanzania						
Median	2	4.5	3	6	0	0
Mean	2.42	5.41	3.96	6.73	0.08	0.29
N = 971						

Table 3 Determinants of Education and Pooled Earnings in Ghana and Tanzania

	Ghana	Tanzania	Ghana		Tanzania	
	(1)	(2)	(3)	(4)	(5)	(6)
	Education		Ln (Real Earnings in 1995 prices)			
Male	2.22***	1.03***	0.32***	0.28***	0.32***	0.27***
	(0.20)	(0.20)	(0.05)	(0.06)	(0.07)	(0.07)
Age	0.12	0.22***	0.09***	0.09***	0.04	0.03
	(0.09)	(0.08)	(0.02)	(0.02)	(0.03)	(0.03)
Age^2	-0.09	-0.26***	-0.11***	-0.11***	-0.05	-0.03
	(0.12)	(0.10)	(0.03)	(0.03)	(0.04)	(0.04)
Kdistprim	-0.22***	-0.18***				
	(0.07)	(0.04)				
Kdistsec	-0.06*	0.12***				
	(0.03)	(0.02)				
Momeduc	0.07***	0.11***				
	(0.02)	(0.04)				
Dadeduc	0.31***	0.11***				
	(0.02)	(0.03)				
Momformal	1.14***	1.22***				
	(0.41)	(0.36)				
Dadformal	0.15	0.94***				
	(0.22)	(0.23)				
Lhours			0.32***	0.33***	0.37***	0.36***
			(0.09)	(0.09)	(0.12)	(0.12)
Education			-0.05***	-0.03	-0.01	0.06
			(0.02)	(0.02)	(0.03)	(0.04)
$Education^2$			0.48***	0.47***	0.63***	0.58***
			(0.10)	(0.10)	(0.21)	(0.21)
Small firm			-0.20***	-0.20***	-0.38	-0.37
			(0.06)	(0.06)	(0.23)	(0.23)
Large firm			0.37***	0.36***	0.16**	0.15**
			(0.07)	(0.07)	(0.07)	(0.07)
Public Sector			0.58***	0.58***	0.50***	0.48***
			(0.09)	(0.09)	(0.12)	(0.12)
Control for Education				-0.03**		-0.07***
				(0.01)		(0.03)
Constant	545.65	317.92	102.36	88.73	-254.16*	-260.11*
	(405.31)	(391.29)	(92.87)	(93.15)	(141.94)	(141.29)
Observations	1131	910	1131	1131	910	910
R-squared	0.340	0.239	0.234	0.237	0.206	0.212
Marginal ROR (evaluated at mean)			0.035	0.054	0.100	0.161

Robust standard errors in parentheses*** p<0.01, ** p<0.05, * p<0.1

Table 4 Probabilities of Being in Occupations as a Function of Education and Age

	Self-employment	Small Firm	Large Firm	Public Sector	No employment
Ghana Education					
None	0.62	0.09	0.01	0.00	0.27
Junior Secondary	0.52	0.09	0.05	0.02	0.32
Senior Secondary	0.42	0.08	0.08	0.06	0.36
University	0.29	0.06	0.11	0.17	0.37
Tanzania Education					
None	0.79	0.05	0.05	0.01	0.10
Junior Secondary	0.72	0.07	0.05	0.10	0.05
Senior Secondary	0.62	0.06	0.05	0.24	0.04
University	0.41	0.03	0.03	0.50	0.03
Ghana Age					
Age 20	0.24	0.09	0.06	0.01	0.60
Age 30	0.54	0.10	0.05	0.01	0.30
Age 40	0.66	0.07	0.04	0.01	.21
Age 50	0.61	0.05	0.04	0.03	0.27
Age 60	0.31	0.02	0.03	0.13	0.51
Tanzania Age					
Age 20	0.66	0.27	0.01	0.01	0.04
Age 30	0.77	0.09	0.05	0.05	0.05
Age 40	0.72	0.04	0.07	0.11	0.05
Age 50	0.64	0.04	0.05	0.21	0.05
Age 60	0.52	0.07	0.02	0.34	0.05

Notes; These probabilities are calculated from the multinomial logit reported in the Appendix. All other regressors are held at their mean values.

Table 5 Earnings Function for Ghana and Tanzania: by Occupation and No controls
Dependent Variable Ln (Real Earnings in 1995 prices)

	(1) Self-employment	(2) Small firm	(3) Large firm	(4) Public Sector	(5) Self-employment	(6) Small firm	(7) Large firm	(8) Public Sector
	Ghana				Tanzania			
Lhours	0.34***	0.45*	0.13	0.01	0.42***	2.362	0.189	0.107
	(0.11)	(0.26)	(0.25)	(0.42)	(0.122)	(1.504)	(0.259)	(0.465)
Male	0.39***	0.31***	0.04	0.07	0.36***	1.27***	0.132	-0.103
	(0.07)	(0.11)	(0.14)	(0.13)	(0.075)	(0.413)	(0.112)	(0.195)
Age	0.08***	0.24***	0.13***	0.05	0.010	0.208	-0.13***	-0.016
	(0.02)	(0.05)	(0.04)	(0.04)	(0.029)	(0.129)	(0.035)	(0.079)
Age^2	-0.09***	-0.33***	-0.15**	-0.05	-0.005	-0.308	0.16***	0.025
	(0.03)	(0.08)	(0.06)	(0.06)	(0.035)	(0.200)	(0.045)	(0.087)
Education	-0.05**	-0.05*	0.10	0.02	0.006	-0.354	-0.010	0.118
	(0.02)	(0.03)	(0.07)	(0.12)	(0.033)	(0.421)	(0.085)	(0.093)
$Education^2$	0.37**	0.62***	-0.05	0.18	0.410*	3.129	0.807*	0.016
	(0.17)	(0.20)	(0.40)	(0.45)	(0.212)	(2.736)	(0.481)	(0.496)
Trend	-0.03	-0.17	0.03	0.01	0.066	0.358	0.221**	0.195
	(0.06)	(0.11)	(0.11)	(0.13)	(0.070)	(0.504)	(0.103)	(0.191)
Constant	59.01	346.34	-61.13	-14.79	-128.751	-725.606	-436.4**	-385.233
	(122.21)	(219.36)	(224.73)	(252.13)	(140.0)	(1,014)	(207.6)	(382.6)
Observations	754	157	123	97	618	72	69	151
R-squared	0.094	0.375	0.366	0.196	0.174	0.295	0.621	0.268
Rates of Return (ROR)								
Average (a)	-0.015	0.001	0.074	0.037	0.036	-0.236	0.043	0.091
Marginal (a)	0.018	0.056	0.087	0.069	0.063	0.084	0.102	0.122
Median levels of Education	9	9	11	14	7	7	7	12

(a) The average and marginal rates of return (ROR) are evaluated at the median levels of education for each occupation.

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 6 Earnings Functions for Ghana
Dependent Variable Ln (Real Earnings in 1995 prices)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Self-Employment		Small firm		Large firm		Public Sector	
	LEE	DMF	LEE	DMF	LEE	DMF	LEE	DMF
Lhours	0.35***	0.36***	0.46*	0.45*	0.12	0.00	-0.01	0.01
	(0.11)	(0.11)	(0.25)	(0.26)	(0.25)	(0.28)	(0.43)	(0.42)
Male	0.47***	0.43***	0.18	0.14	-0.05	0.21	0.13	0.33
	(0.08)	(0.12)	(0.12)	(0.18)	(0.25)	(0.29)	(0.16)	(0.22)
Age	-0.07	-0.11**	0.22***	0.20*	0.11**	-0.22	0.07*	-0.04
	(0.05)	(0.05)	(0.05)	(0.11)	(0.05)	(0.15)	(0.04)	(0.18)
Age^2	0.08	0.14**	-0.28***	-0.26*	-0.13*	0.30	-0.08	0.05
	(0.06)	(0.07)	(0.07)	(0.14)	(0.07)	(0.19)	(0.06)	(0.24)
Education	-0.04	-0.04	-0.01	0.01	0.06	0.35**	-0.07	0.15
	(0.03)	(0.03)	(0.04)	(0.04)	(0.08)	(0.16)	(0.17)	(0.25)
$Education^2$	0.76***	1.02***	0.71***	0.59**	0.00	-0.18	0.21	-0.18
	(0.21)	(0.25)	(0.21)	(0.29)	(0.41)	(0.67)	(0.50)	(0.62)
Control	-0.04**	-0.05***	-0.07*	-0.08*	0.02	-0.04	0.07*	0.03
	(0.02)	(0.02)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.06)
Dmf_self		0.44**		0.96		0.51		-1.21
		(0.22)		(1.38)		(1.52)		(1.88)
Dmf_small		0.03		0.03		1.07		0.89
		(0.42)		(0.14)		(1.00)		(2.06)
Dmf_big		0.15		0.36		-0.32		1.13
		(0.62)		(1.02)		(0.22)		(1.32)
Dmf_pub		1.30**		0.38		3.71**		-0.03
		(0.60)		(1.34)		(1.48)		(0.23)
Dmf_none		1.10**		1.16		4.63**		0.83
		(0.55)		(1.41)		(1.88)		(1.07)
Lee	-0.96***		-0.16		-0.20		-0.21	
	(0.26)		(0.30)		(0.33)		(0.36)	
Constant	149.07	227.93*	303.45	339.82	-46.15	252.08	-7.33	7.02
	(122.84)	(132.50)	(223.47)	(232.63)	(225.39)	(239.64)	(234.99)	(285.85)
Obs	754	754	157	157	123	123	97	97
R-squared	0.116	0.119	0.398	0.404	0.369	0.429	0.223	0.254
Rates of Return (ROR)								
Average	0.032	0.047	0.049	0.060	0.065	0.34	-0.045	0.12
Marginal	0.018	0.056	0.087	0.069	0.063	0.084	0.102	0.122
Education	9	9	9	9	11	11	14	14
Test of education	0.00	0.00	0.00	0.098	0.19	0.001	0.91	0.81
Education Coefficient		0.039 (0.021)*		0.053 (0.043)		0.33 (0.09)***		0.09 (0.15)

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Table 7 Earnings Function for Tanzania
Dependent Variable Ln (Real Earnings in 1995 prices)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Self-Employment		Small firm		Large firm		Public Sector	
	LEE	DMF	LEE	DMF	LEE	DMF	LEE	DMF
Lhours	0.43***	0.37***	2.43	1.88	0.12	0.11	-0.20	-0.31
	(0.12)	(0.12)	(1.49)	(1.18)	(0.23)	(0.23)	(0.50)	(0.49)
Male	0.33***	0.19	0.58	0.30	0.48***	0.30	0.54**	-0.03
	(0.08)	(0.13)	(0.39)	(0.93)	(0.17)	(0.19)	(0.24)	(0.24)
Age	-0.01	0.03	0.42	0.50	-0.12**	-0.15**	-0.12	-0.27*
	(0.03)	(0.04)	(0.31)	(0.33)	(0.05)	(0.06)	(0.10)	(0.14)
Age^2	0.02	-0.02	-0.53	-0.56	0.14**	0.16**	0.07	0.24*
	(0.04)	(0.05)	(0.39)	(0.38)	(0.06)	(0.07)	(0.09)	(0.14)
Education	0.05	0.01	-0.06	-0.20	0.21*	0.20*	-0.13	-0.45*
	(0.04)	(0.05)	(0.39)	(0.46)	(0.11)	(0.11)	(0.17)	(0.27)
$Education^2$	0.44**	0.40*	3.37	4.92	0.32	-0.15	-0.08	0.31
	(0.22)	(0.23)	(3.07)	(3.78)	(0.50)	(0.46)	(0.63)	(0.93)
Control	-0.05*	-0.02	-0.33	-0.20	-0.15***	-0.11***	-0.01	0.15*
	(0.03)	(0.03)	(0.24)	(0.26)	(0.04)	(0.04)	(0.06)	(0.09)
Dmf_self		-0.04		3.07		-0.79		1.46
		(0.40)		(5.16)		(1.18)		(0.99)
Dmf_small		-0.59		0.24		-0.48		1.72
		(0.59)		(0.41)		(0.68)		(1.69)
Dmf_big		0.36		4.56		-0.07		-5.75**
		(0.71)		(4.95)		(0.17)		(2.29)
Dmf_pub		0.22		8.46		-2.11***		1.15**
		(0.85)		(7.21)		(0.68)		(0.52)
Dmf_none		2.13***		7.78		1.18		3.59
		(0.77)		(5.73)		(1.11)		(2.18)
Lee	-0.28		-2.03		0.64**		-1.79**	
	(0.27)		(1.32)		(0.26)		(0.88)	
Constant	-143.27	-428.8**	-1,169	-1,258	-500.1**	-795.0***	-757.3*	-1,124.5**
	(140.06)	(201.56)	(1,040)	(1,092)	(204.50)	(207.32)	(389)	(437.76)
Obs	618	618	72	72	69	69	151	151
R-squared	0.179	0.196	0.341	0.375	0.692	0.726	0.343	0.404
Rates of Return (ROR)								
Average	0.085	0.034	0.174	0.144	0.235	0.187	-0.138	-0.408
Marginal	0.117	0.062	0.409	0.488	0.257	0.177	-0.147	-0.371
Education	7	7	7	7	7	7	12	12
Test of Education	0.00	0.07	0.11	0.22	0.00	0.009	0.62	0.17
Education Coefficient		0.059 (0.041)		0.418 (0.269)		0.17 (3.10)***		-0.38 (0.21)*

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Appendix Multinomial Logit for Ghana

```
. mlogit occup $z_occup $g_cities if country=="Ghana", base(5)

Multinomial logistic regression                 Number of obs   =       1640
                                                LR chi2(72)     =     734.92
                                                Prob > chi2     =     0.0000
Log likelihood = -1775.2231                     Pseudo R2       =     0.1715

------------------------------------------------------------------------------
       occup |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
self         |
        male |  -.2231523   .1455154    -1.53   0.125    -.5083574    .0620527
         age |   .3898483   .0502781     7.75   0.000      .291305    .4883915
       agesq |  -.4743283   .0685858    -6.92   0.000    -.6087539   -.3399026
        educ |   .0143789   .0506899     0.28   0.777    -.0849715    .1137292
      educsq |  -.5732546   .3288607    -1.74   0.081     -1.21781    .0713005
   kdistprim |  -.0479074   .0443084    -1.08   0.280    -.1347502    .0389354
    kdistsec |  -.0586812   .0172041    -3.41   0.001    -.0924007   -.0249617
     momeduc |   .0003828   .0162105     0.02   0.981    -.0313891    .0321548
     dadeduc |  -.0273207   .0163047    -1.68   0.094    -.0592774     .004636
   momformal |  -.2844869   .2912727    -0.98   0.329    -.8553709     .286397
   dadformal |   .1531491   .1456307     1.05   0.293    -.1322819    .4385801
     married |   .5243778   .1587083     3.30   0.001     .2133153    .8354404
    children |  -.0274878   .0484397    -0.57   0.570    -.1224278    .0674522
      hhhead |   .4518191   .1582449     2.86   0.004     .1416648    .7619733
       trend |   .2019249   .1284651     1.57   0.116     -.049862    .4537118
      cityG2 |  -.4884512   .3470838    -1.41   0.159    -1.168723    .1918205
      cityG3 |   .0827995    .208191     0.40   0.691    -.3252474    .4908464
      cityG4 |    .406809   .2099275     1.94   0.053    -.0046413    .8182592
       _cons |  -411.0172   257.4919    -1.60   0.110    -915.6921    93.65763
-------------+----------------------------------------------------------------
small        |
        male |   .0298103   .2046155     0.15   0.884    -.3712286    .4308493
         age |   .2870855   .0852462     3.37   0.001      .120006     .454165
       agesq |  -.4043467   .1274457    -3.17   0.002    -.6541356   -.1545578
        educ |   .0290035    .080715     0.36   0.719    -.1291949    .1872019
      educsq |  -.5392167    .505641    -1.07   0.286    -1.530255    .4518214
   kdistprim |   .0620833   .0582399     1.07   0.286    -.0520648    .1762314
    kdistsec |  -.0700276    .027779    -2.52   0.012    -.1244734   -.0155818
     momeduc |  -.0049403   .0231858    -0.21   0.831    -.0503836     .040503
     dadeduc |  -.0188061   .0241072    -0.78   0.435    -.0660554    .0284433
   momformal |    .063231   .4022858     0.16   0.875    -.7252347    .8516967
   dadformal |  -.0844924    .213005    -0.40   0.692    -.5019746    .3329898
     married |   .1644421   .2522728     0.65   0.515    -.3300035    .6588876
    children |  -.3911359   .1029677    -3.80   0.000    -.5929488   -.189323
      hhhead |   .7756996   .2285116     3.39   0.001      .327825    1.223574
       trend |   .1072778   .1911539     0.56   0.575     -.267377    .4819326
      cityG2 |  -.3444049   .5758611    -0.60   0.550    -1.473072    .7842621
      cityG3 |   .2917272   .3281626     0.89   0.374    -.3514597     .934914
      cityG4 |   .5023237   .3350372     1.50   0.134     -.154337    1.158985
       _cons |  -220.3292   383.0921    -0.58   0.565     -971.176    530.5175
-------------+----------------------------------------------------------------
big          |
        male |   1.246777   .2684864     4.64   0.000     .7205528       1.773
         age |   .2013411   .0917237     2.20   0.028     .0215659    .3811162
       agesq |  -.2642877   .1251682    -2.11   0.035     -.509613   -.0189625
        educ |   .1619721   .1221033     1.33   0.185    -.0773458    .4012901
      educsq |  -.3175635    .580052    -0.55   0.584    -1.454445    .8193176
   kdistprim |  -.0590876   .0780058    -0.76   0.449    -.2119762    .0938009
    kdistsec |  -.0169393   .0309789    -0.55   0.585    -.0776569    .0437782
     momeduc |   .0215358   .0265817     0.81   0.418    -.0305634     .073635
     dadeduc |  -.0013565   .0279254    -0.05   0.961    -.0560893    .0533762
   momformal |    .650188   .3862434     1.68   0.092    -.1068351    1.407211
   dadformal |   -.193796   .2437946    -0.79   0.427    -.6716247    .2840327
     married |   .6535238   .2828014     2.31   0.021     .0992431    1.207804
    children |  -.1196238    .091568    -1.31   0.191    -.2990939    .0598462
      hhhead |   1.252162   .2809087     4.46   0.000     .7015911    1.802733
       trend |   .1466963   .2166152     0.68   0.498    -.2778618    .5712543
```

```
      cityG2 |  -1.160681    .808412    -1.44   0.151     -2.74514     .423777
      cityG3 |   .3137569    .348639     0.90   0.368     -.369563    .9970768
      cityG4 |   .4022089   .3612849     1.11   0.266    -.3058966    1.110314
       _cons |  -301.8194   434.1446    -0.70   0.487    -1152.727    549.0885
-------------+----------------------------------------------------------------
public       |
        male |   .3231275   .2893333     1.12   0.264    -.2439553    .8902103
         age |  -.0056742   .1035075    -0.05   0.956    -.2085453    .1971968
       agesq |   .0811773   .1357383     0.60   0.550    -.1848649    .3472195
        educ |   .6576127   .3145067     2.09   0.037     .0411909    1.274034
      educsq |  -1.254633   1.194184    -1.05   0.293     -3.59519    1.085924
   kdistprim |  -.0152569   .0989022    -0.15   0.877    -.2091016    .1785878
    kdistsec |  -.1644958   .0536393    -3.07   0.002    -.2696268   -.0593647
     momeduc |  -.0020955   .0305053    -0.07   0.945    -.0618847    .0576938
     dadeduc |  -.0271856   .0345529    -0.79   0.431     -.094908    .0405369
   momformal |   .4730179   .4411918     1.07   0.284    -.3917022    1.337738
   dadformal |   .4336503   .2843857     1.52   0.127    -.1237354    .9910359
     married |   .8046039   .3137128     2.56   0.010      .189738     1.41947
    children |  -.1538673   .1037152    -1.48   0.138    -.3571453    .0494107
      hhhead |   1.392831   .3348588     4.16   0.000     .7365203    2.049142
       trend |  -.1548345   .2585342    -0.60   0.549    -.6615522    .3518833
      cityG2 |   .7146756   .5281249     1.35   0.176    -.3204302    1.749781
      cityG3 |   .1453116   .4067444     0.36   0.721    -.6518928     .942516
      cityG4 |  -.1445446   .4716051    -0.31   0.759    -1.068874    .7797844
       _cons |   301.8829   518.1798     0.58   0.560    -713.7309    1317.497
```

Appendix Multinomial Logit for Tanzania

```
mlogit occup $z_occup $t_cities if country=="Tanzania", base(5)


Multinomial logistic regression                 Number of obs   =        971
                                                LR chi2(80)     =     340.83
                                                Prob > chi2     =     0.0000
Log likelihood = -928.42373                     Pseudo R2       =     0.1551

------------------------------------------------------------------------------
       occup |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
self         |
        male |   .2468858   .3678285     0.67   0.502    -.4740449    .9678164
         age |   .0039053   .1209237     0.03   0.974    -.2331008    .2409114
       agesq |  -.0184457   .1437428    -0.13   0.898    -.3001764     .263285
        educ |   .0937882    .141824     0.66   0.508    -.1841818    .3717582
      educsq |   -.337557   .9618648    -0.35   0.726    -2.222777    1.547663
   kdistprim |   .0885963   .0747763     1.18   0.236    -.0579626    .2351553
    kdistsec |   .0059949   .0337275     0.18   0.859    -.0601097    .0720995
     momeduc |   .0875089    .056269     1.56   0.120    -.0227763    .1977942
     dadeduc |   .0303874    .043235     0.70   0.482    -.0543516    .1151264
   momformal |  -.2007229   .6990147    -0.29   0.774    -1.570767    1.169321
   dadformal |    .152262   .3556808     0.43   0.669    -.5448595    .8493835
     married |   .4064242   .3334476     1.22   0.223     -.247121    1.059969
    children |   -.044904   .0746582    -0.60   0.548    -.1912315    .1014234
      hhhead |   .4235197    .363808     1.16   0.244     -.289531     1.13657
       trend |   -.591782   .2810034    -2.11   0.035    -1.142539   -.0410254
      cityT2 |  -.6601001   .4762334    -1.39   0.166      -1.5935    .2733001
      cityT3 |   .5322591   .6855646     0.78   0.438    -.8114229    1.875941
      cityT4 |  -1.028887   .5805314    -1.77   0.076    -2.166708    .1089337
      cityT5 |  -.0943441   .6060365    -0.16   0.876    -1.282154    1.093465
      cityT6 |   .1469401   .6444253     0.23   0.820     -1.11611     1.40999
       _cons |   1187.421   563.0607     2.11   0.035     83.84202    2290.999
-------------+----------------------------------------------------------------
small        |
        male |    .568852   .4646153     1.22   0.221    -.3417773    1.479481
         age |   -.289319   .1534352    -1.89   0.059    -.5900466    .0114085
       agesq |    .312551   .1849663     1.69   0.091    -.0499763    .6750784
        educ |    .197128   .2176478     0.91   0.365    -.2294539    .6237099
```

```
      educsq |  -.8850147   1.382572    -0.64   0.522    -3.594806    1.824776
   kdistprim |  -.0422209   .1098428    -0.38   0.701    -.2575089    .1730671
    kdistsec |   -.044297   .0468402    -0.95   0.344     -.136102     .047508
     momeduc |    .114716   .0690537     1.66   0.097    -.0206268    .2500587
     dadeduc |   .0220869   .0558898     0.40   0.693    -.0874552    .1316289
   momformal |   .1199272   .8603306     0.14   0.889     -1.56629    1.806144
   dadformal |  -.5212942   .4735209    -1.10   0.271    -1.449378    .4067897
     married |   .8463326   .4568676     1.85   0.064    -.0491115    1.741777
    children |  -.1193257   .1147238    -1.04   0.298    -.3441802    .1055289
      hhhead |   .5981817   .4808451     1.24   0.213    -.3442573    1.540621
       trend |  -.5994925   .3677569    -1.63   0.103    -1.320283    .1212978
      cityT2 |  -.8047743   .5916912    -1.36   0.174    -1.964468    .3549191
      cityT3 |   .2957965   .8515092     0.35   0.728    -1.373131    1.964724
      cityT4 |  -1.239524   .7874104    -1.57   0.115     -2.78282     .303772
      cityT5 |   .1266175   .7354171     0.17   0.863    -1.314773    1.568008
      cityT6 |   .0225123   .7812385     0.03   0.977    -1.508687    1.553712
       _cons |    1206.58   736.8451     1.64   0.102    -237.6102     2650.77
-------------+----------------------------------------------------------------
big          |
        male |   1.843614   .4895411     3.77   0.000     .8841315    2.803097
         age |   .2789818   .1728603     1.61   0.107    -.0598181    .6177817
       agesq |  -.3447405   .2089946    -1.65   0.099    -.7543624    .0648814
        educ |   .1091392    .226016     0.48   0.629    -.3338441    .5521224
      educsq |   -.369972   1.420124    -0.26   0.794    -3.153365    2.413421
   kdistprim |   .0032273    .102969     0.03   0.975    -.1985882    .2050429
    kdistsec |  -.0066518    .046442    -0.14   0.886    -.0976765    .0843729
     momeduc |   .0757266    .069709     1.09   0.277    -.0609006    .2123538
     dadeduc |   .0538672   .0562091     0.96   0.338    -.0563007     .164035
   momformal |   .4062419   .8379858     0.48   0.628     -1.23618    2.048664
   dadformal |   -.177371   .4715687    -0.38   0.707    -1.101629    .7468866
     married |  -.4113186   .4616361    -0.89   0.373    -1.316109    .4934716
    children |  -.1268992    .117069    -1.08   0.278    -.3563502    .1025517
      hhhead |   .2712471   .4870992     0.56   0.578    -.6834497    1.225944
       trend |  -.5090753   .3745428    -1.36   0.174    -1.243166    .2250152
      cityT2 |  -1.775578   .5707631    -3.11   0.002    -2.894254   -.6569033
      cityT3 |  -.6563304   .8508499    -0.77   0.440    -2.323965    1.011305
      cityT4 |  -2.211388   .8018383    -2.76   0.006    -3.782963   -.6398143
      cityT5 |  -1.584811   .7891059    -2.01   0.045     -3.13143   -.0381916
      cityT6 |  -.2896324    .749163    -0.39   0.699    -1.757965      1.1787
       _cons |     1014.8   750.4387     1.35   0.176    -456.0332    2485.632
-------------+----------------------------------------------------------------
public       |
        male |  -.5972426   .4558901    -1.31   0.190    -1.490771    .2962855
         age |   .1335665   .1509389     0.88   0.376    -.1622683    .4294013
       agesq |  -.0726103   .1766551    -0.41   0.681    -.4188478    .2736273
        educ |   .2644814   .2066158     1.28   0.201    -.1404782     .669441
      educsq |   .4552323   1.246232     0.37   0.715    -1.987338    2.897803
   kdistprim |   .0576885   .0853955     0.68   0.499    -.1096837    .2250606
    kdistsec |   .0695581   .0382476     1.82   0.069    -.0054058    .1445219
     momeduc |    .004732    .065306     0.07   0.942    -.1232655    .1327295
     dadeduc |   .0904871   .0505858     1.79   0.074    -.0086593    .1896334
   momformal |   .2364915   .7534257     0.31   0.754    -1.240196    1.713179
   dadformal |   .0764036   .4208799     0.18   0.856    -.7485059    .9013131
     married |   1.022381   .4201959     2.43   0.015     .198812     1.84595
    children |   -.181306   .0941846    -1.93   0.054    -.3659045    .0032924
      hhhead |   .6919723   .4511107     1.53   0.125    -.1921885    1.576133
       trend |  -.8652108    .337154    -2.57   0.010    -1.526021    -.204401
      cityT2 |  -1.347306   .5542582    -2.43   0.015    -2.433632     -.26098
      cityT3 |   .0168272   .7761988     0.02   0.983    -1.504495    1.538149
      cityT4 |  -.9109426    .667258    -1.37   0.172    -2.218744    .3968589
      cityT5 |   .0912407   .6804659     0.13   0.893    -1.242448    1.424929
      cityT6 |   .1942339   .7439718     0.26   0.794    -1.263924    1.652392
       _cons |   1727.336   675.5535     2.56   0.011     403.2758    3051.397
------------------------------------------------------------------------------
(occup==none is the base outcome)
```