Education, skills, and labor market outcomes: evidence from Ghana

Kingdon, Geeta; Söderbom, Måns

Education, Skills, and Labor Market Outcomes: Evidence from Ghana*

Geeta Kingdon** and Måns Söderbom†

Draft

April 2007

Abstract

This paper investigates the education-earnings relationship in Ghana, drawing on the Ghana Living Standards Survey for 1998-99. The analysis has three main goals: to examine the labor market returns to education amongst wage-employed, self-employed and agricultural workers; to examine the labor market returns to literacy and numeracy skills for these categories of workers; and to analyze the pattern of returns to education along the earnings distribution. We also investigate the shape of the education-earnings relationship. The analysis is done separately by gender and age group, and attempts to address the usual biases when estimating returns to education.

Key words:
JEL Classification:

Acknowledgements: We are indebted to Tazeen Fasih and Alonso Sánchez for useful comments on an earlier draft of the paper, and to Alonso Sánchez who provided excellent research assistance and for writing the data appendix.

** Corresponding author. Centre for the Study of African Economies, Department of Economics, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK. Telephone number: +44 (0)1865 271065. Fax number: +44 (0)1865 281447. Email: geeta.kingdon@economics.ox.ac.uk.

† Centre for Study of African Economies, Department of Economics, University of Oxford, UK.

1 Introduction

The motivation and research questions for Ghana remain the same as for the Pakistan case study: (i) to examine the labor market returns to education amongst wage-employed, self-employed and agricultural workers; (ii) to examine the labor market returns to literacy and numeracy skills for these categories of workers; and (iii) to analyze the pattern of returns to education along the earnings distribution. In addition, in the case of Ghana the data permits us to examine the simple return to technical and vocational education and training.

In a wide-ranging paper on education, incomes, poverty and inequality, Teal (2001) estimates the returns to education in Ghana using four waves of data from 1988 to 1999. Unlike much of the international literature, his study estimates returns to education not only in wage employment but also in the other two major occupations – agriculture (which employed 64% of the labour force in 1998-99) and self-employment. Pooling the four rounds of Ghana data, he introduces Round dummies to examine how earnings changed over time in each occupation. A major contribution of the paper is to showcase how the availability of data over time can be used to learn about the poverty-reducing potential of education because it permits decomposition of any increase in incomes over time into that due to changes in the average amount of education and that due to underlying technical progress.

The current work adds value to Teal (2001) in five ways. Firstly, we examine the role of education in facilitating entry into lucrative occupations, by means of multinomial logit models of occupational attainment. This is important because, as noted before, education plays a role in labour market success not only directly by increasing earnings in any given occupation but also indirectly by promoting entry into the well-paying occupations. Secondly, we examine the role of cognitive skills in labour market success, both in terms of occupational outcome and earnings. Thirdly, we estimate returns to education along the earnings distribution by means of quantile regression analysis to ask whether the marginal return to education is greater at lower levels of earnings, i.e. whether education ameliorates economic inequality or exacerbates it. Fourthly, we estimate returns to education by age group to examine whether the labour market rewards education differentially for younger and older workers. Lastly, we estimate the simple return to technical and vocational education and training[1].

[1] When we have 2004-05 data from Ghana, we will extend the analysis to examine whether and how the relationship of education with labour market outcomes changed over the 6 year period 1999 to 2005.

2 Analytical approach

It is widely believed that education affects people's economic status by raising their earnings in the labor market. However, it may raise earnings through a number of different channels such as via improving access to employment or, conditional on employment, via promoting entry into higher paying occupations or industries. In this paper we explore both the total effect of education on earnings and also the role of education in occupational attainment since the latter is an important mechanism through which the market benefits of education are realized. The earnings function for wage employees is specified in general form as

$$\ln w_i = \boldsymbol{\alpha}_{\mathbf{ag}}\mathbf{x}_{\mathbf{i}} + f_{ag}(s_i) + \upsilon_i \tag{1}$$

where w_i is real earnings of individual *i*, $\mathbf{x}_{\mathbf{i}}$ is a vector of worker characteristics excluding education, $\boldsymbol{\alpha}_{\mathbf{ag}}$ is a parameter vector, s_i is the years of education, $f_{ag}(\cdot)$ is the earnings-education profile, υ_i is a residual, and *a* and *g* denote age group and gender, respectively. The primary objective in this paper is to estimate the total returns to education, and the variables included in the $\mathbf{x}_{\mathbf{i}}$ are selected accordingly. In particular, in estimating the earnings regressions we do not condition on variables that are determined by education, as conditioning on such variables would change the interpretation of the schooling effects. For example, it is likely that an important effect of education is to enable individuals to get high-wage jobs (e.g. managerial positions), get into certain high-wage sectors or firms, or to generate job security and thus work experience. Consequently, we do not condition on occupation, firm-level variables, work experience, or other variables sometimes seen on the right-hand side in earnings regressions. Instead, we restrict ourselves to a small set of control variables, where age and gender are those emphasised the most. With respect to the effects of these variables on earnings, we allow for a fair deal of flexibility and estimate all regressions separately for men and women, and separately for relatively young individuals (aged less than 30) and relatively old ones. Within each gender-age group, we include age as an additional control variable. We also include controls for province fixed effects.

Key for our purposes is the estimation of the earnings-education profile $f_{ag}(\cdot)$. We focus on two specifications: a standard linear model, and a model with dummy variables for highest level of education completed. The former is attractive partly because the results are straightforward to interpret, whereas the latter is an attractive way of analysing how returns to education differ across different levels of education. In addition, we also consider a model where a quadratic term is added to the linear specification. This is a convenient way of testing for nonlinearities in the earnings-education profile.

In the empirical analysis, earnings regressions are estimated based on data from three labor market sub-sectors, namely wage employment, self employment, and agriculture. Amongst the wage employed, we have individual data on earnings as well as on the explanatory variables. For individuals that are either self employed or work in the agricultural sector, we do not have earnings data at the individual level. Instead, we have earnings at the household level, distinguishing between earnings for self employed and earnings for agricultural workers. In order to identify the parameters in (1) we then need to aggregate the explanatory variables so that these are defined at the same level of aggregation as the dependent variable. Fortunately, this is a straightforward task. All we need to do is 'collapse' the data - i.e. calculate mean values - on the explanatory variables within household, and labor market sub-sector (obviously we do not do this for the wage employed, as we have individual level data on earnings for these individuals).[2] Thus, for agriculture and self employment, the estimable earnings equation is written

$$\ln \overline{w}_{hc} = \boldsymbol{\alpha}_{\mathbf{at}} \overline{\mathbf{x}}_{hc} + \left[\overline{f_{at}(s_i)}\right]_{hc} + \overline{\upsilon}_{hc} ,$$

where *hc* are household-category subscripts, and the bar-superscript indicates household-category averages.

Endogeneity bias

The two major sources of bias in the OLS estimate of the effect of education on earnings are sample selectivity bias and endogeneity (omitted variable) bias. Sample selectivity bias arises due to estimating the earnings function on separate sub-samples of workers, each of which may not be a random draw from the population. This violates a fundamental assumption of the least squares regression model. While modeling occupational outcomes is a useful exercise in its own right – suggesting the way in which education influences people's decision to participate in wage, self or agricultural employment – it is also needed for the consistent estimation of earnings functions. Modeling participation in different occupations is the first step of the Heckman procedure to correct for sample selectivity: probabilities predicted by the occupational choice model are used to derive the selectivity term that is used in the earnings function.

[2] To give a concrete example, suppose a household has two agricultural workers, and three self-employed individuals. There are data only on total earnings derived from agriculture, and the total earnings from self-employment, for the household, which means it is not possible to estimate the earnings equation at the individual level. What we do, then, is calculate earnings per person in agriculture, and in self employment, and match this information with sector-household specific averages of the explanatory variables.

Adding a subscript *j* to denote occupation-type to the earnings function (1),

$$\ln w_{ij} = \boldsymbol{\alpha}_{\mathbf{a}gj}\mathbf{x}_{\mathbf{ij}} + f_{agj}(s_{ij}) + \upsilon_{ij} \qquad (1')$$

it follows that the expected value of the dependent variable, conditional on the explanatory variables *x* and *s, and* selection into occupation *j*, is equal to

$$E(\ln w_{ij} | \mathbf{x}_{\mathbf{ij}}, s_{ij}, m_{ij} = 1) = \boldsymbol{\alpha}_{\mathbf{a}gj}\mathbf{x}_{\mathbf{ij}} + f_{agj}(s_{ij}) + E(\upsilon_{ij} | m_{ij} = 1),$$

where m_{ij} is a dummy variable equal to one if occupation *j* was selected and zero otherwise. The last term in (2) is not necessarily equal to zero in the sample of observations in sector *j*, in which case estimating the wage equation ignoring sample selection will lead to biased estimates. For example, if more highly motivated or more ambitious people systematically select into particular occupations – say, for example, into waged work – then people in the waged sub-sample would, on average, be more motivated and ambitious than those in the rest of the population. Thus, $E(\upsilon_{ij} | m_{ij} = 1)$ is not zero in this subsample, as the waged workers' sub-sample is not a random draw from the whole population. Least squares would therefore yield inconsistent parameter estimates. Following Heckman (1979) and Lee (1983), the earnings equations can be corrected for selectivity by including the inverse of Mills ratio λ_{ji} as an additional explanatory variable in the wage equation, so that

$$\ln w_{ij} = \boldsymbol{\alpha}_{\mathbf{a}gj}\mathbf{x}_{\mathbf{ij}} + f_{agj}(s_{ij}) + \theta_{agj}\lambda_{ij}(z_{ij}\gamma) + \varepsilon_{ij},$$

where z_{ij} is a set of variables explaining selection into occupation and γ are the associated coefficients. Thus, the probability of selection into each occupation-type is first estimated by fitting a model of occupational attainment, based on which the selectivity term (λ) computed.[3] The coefficients on the lambda terms λ_j will be a measure of the bias due to non-random sample selection. If these are statistically different from zero, the null hypothesis of 'no bias' is rejected. As will be discussed in the next section, we consider five broad labour market states – wage employment, self-employment, agricultural employment, unemployed, and individuals out of the labor force - and so occupational attainment is modeled using a multinomial logit equation.

---

[3] The inverse Mill's ratio is defined $\lambda_{ji} = \dfrac{\phi(H_{ij})}{\Phi(H_{ij})}$, where $H_{ij} = \Phi^{-1}(P_{ij})$, $\phi(.)$ is the standard normal density function, $\Phi(.)$ the normal distribution function, and P_{ij} is the estimated probability that the *i*th worker chooses the *j*th occupation.

Another way of expressing the problem of endogenous sample selection is as 'endogeneity' or omitted variable bias. Endogeneity bias arises if workers' unobserved traits, which are in the error term, are systematically correlated both with included independent variables and with the dependent variable (earnings). For instance, if worker ability is positively correlated with both education and earnings then any positive coefficient on education in the earnings function may simply reflect the cross-section correlation between ability on the one hand and both education and earnings on the other, rather than representing a causal effect from education onto earnings.

We will attempt to address the problem of endogeneity by estimating a family fixed effects regression of earnings. To the extent that unobserved traits are shared within the family, their effect will be netted out in a family differenced model. For instance, the error term 'difference in ability between members' will be zero if it is the case that ability is equal among members. While it is unlikely to be the case that unobserved traits are identical across family members, it is likely that they are much more similar within a family than across families and, as such, family fixed effects estimation gives an estimate of the return to education that reduces endogeneity bias without necessarily eliminating it entirely.

Empirical strategy

Our empirical strategy will be the following. We will first estimate the earnings functions for each occupation using the simple Ordinary Least Squares (OLS) model as the base line. Then, we will ask whether there is significant sample selectivity bias due to estimating the earnings functions separately for the occupation groups, since each of these may not be a random draw from the population. Finally we will attempt to address the problem of endogeneity by using a family fixed effects model.[4]

The paper will also estimate earnings functions by the quantile regression (QR) method. OLS regression models the mean of the conditional distribution of the dependent variable. However, if schooling affects the conditional distribution of the dependent variable differently at different points in the wage distribution, then quantile regressions are useful as they allow the contribution of schooling to vary along the distribution of the dependent variable. Thus, the estimation of returns to education using the QR method is more informative than merely being able to say that, on average, one more year of education results in a certain percent increase in earnings. Using quantile regressions we will

[4] We do not have data to implement a credible instrumental variables approach. We have no data on the supply of education at a young age (Card, 1999). In fact, the closest we have to instruments is information on parental education, but only for the sub-sample of individuals co-habiting with their parents at the time of the survey. Given the resulting large (and potentially endogenous) gaps in these data, and given that parental education is a dubious instrument anyway (unobserved ability is probably inherited), we decided against instrumenting education using this variable.

investigate how wages vary with education at the 25th (low), 50th (median) and 75th (high) percentiles of the distribution of earnings. To the extent that one is willing to interpret observations close to the 75th percentile as indicative of higher 'ability' than at lower percentiles (on the grounds that such observations have atypically high wages, given their characteristics), the quantile regressions will thus be informative of the effect of education on earnings across individuals with varying ability[5].

3 The Data and descriptive statistics[6]

The Ghana survey data used in this study correspond to round four of the Ghana Living Standards Survey of 1998/99 (GLSS4). The GLSS4, which was carried out over a one-year period, follows a two-stage sampling strategy to arrive at a nationally representative sample made up of about 26,000 individuals living in 5,998 households. The household questionnaire is composed of a number of detailed modules on such characteristics as education, health, employment, migration, housing, consumption and expenses, as well as information on credit, savings, assets, transfers and miscellaneous income. Additionally, there are modules that concentrate on household enterprises and agricultural activities—including associated expenses and revenues.

The earnings variables for self-employed and agricultural workers are derived from these specialized modules on household enterprises and agricultural activities respectively. A simple, yet comprehensive computation of recurring (non-durable) expenses and revenues—including produced or harvested goods consumed by the household—attributed to enterprise or agricultural endeavours is used to estimate earnings for these types of workers. The earnings of paid employees, however, are derived from the sum of reported income— both in cash and in kind —from the employment module.

Table 1 shows summary statistics for selected variables, for the full sample and for the same five occupation categories used before for the Pakistan case study. As before, our sample consists of persons aged between 16 and 70 not currently enrolled in school. Unemployment and out of the labor

[5] If we assume that education is exogenous then the QR approach tells us the return to education for people with different levels of ability, but *a priori* we cannot assume that education is exogenous. Thus, we cannot say that the return to education for, say, the 90th percentile gives the true return to education for high ability people, purged of ability bias. The same caution is given in Arias, Hallock and Sosa-Escudero (2001), who cite QR studies of returns to education (Buchinsky 1994; Machado and Mata 2000; Schultz and Mwabu 1999) and say that the results of these studies should be interpreted with caution since they do not handle the problems of endogeneity bias.

[6] We are indebted to Alonso Sanchez for making substantial contributions to this section.

force (OLF) state are defined as before. The labor force participation rate is about 72% (compared with 51% for Pakistan) and unemployment rate about 2% (compared with 6% for Pakistan).

Average earnings in the full sample are Ghanaian Cedis 1,350,471 which is equal to about USD 575 (compared with Pakistan's USD 600 for the same year). There is a huge difference in average earnings between agriculturally employed persons and those in either one of wage or self employment. Self-employed and wage-employed persons earn on average about 150% and 163% more (respectively) than individuals working in agriculture. This inter-occupation earnings difference is more than double that in Pakistan where the corresponding figure is 70%. However, the mean can be a misleading measure of central tendency since the earnings distribution is very skewed. Thus, Table 1 also shows median earnings and log of earnings and the Figure below Table 1 shows the distribution of log earnings. These show a clear and pronounced hierarchy, with earnings in agriculture the lowest and in wage employment the highest, with a huge four fold difference. Median earnings in self-employment are about half those in wage employment. The figure shows that wages are more narrowly distributed than self- and agricultural employment earnings. Average years of education in agriculture is 3.7, in self-employment 6.6 and in wage employment 10.5 years. All these education levels are significantly higher than in Pakistan. The pattern for literacy and numeracy skills is similar to that for education. While 70-85% of wage employed and unemployed persons are literate and numerate, the corresponding figures for self-employed persons are considerably lower (50-65%) and they are the lowest for agricultural workers (35-40%).

Thus, there is a clear hierarchy in the occupations with respect to education, skills and earnings: wage employment is at the top with the most well paid, best educated and the most literate and numerate workers; self-employment is next, with lower earnings, education and cognitive skills; and agriculture is last in all these three respects. This suggests that education and skills matter greatly for occupational attainment. While unemployed individuals possess the mean education and skill levels that are close to those of wage employed persons, they seem to queue for suitable job opportunities in the labor market. We now investigate the correlates of occupational outcome more in detail.

4. Education and occupational attainment

As before for Pakistan, we distinguish between the effects of education and skills on occupational outcome and on earnings conditional on occupational outcome. In this section we look at the first issue and Section 5 will look at the second. Our five 'occupations' are: self-employment, agriculture, wage employment, unemployment and individuals out of the labour force (OLF). While from a policy point of view, the link between education and labor market outcomes amongst the relatively young deserves attention, the Ghana sample is not large enough to permit separate analysis by age group as well as gender. We analyze labor market outcomes for all persons (16-70 year olds) mainly by gender but we also present the main tables by age group. For these, we define persons aged 16-30 as 'young' and those aged 31 to 70 as 'old'.

We use a simple multinomial logit model to examine the role of education, skills and family background in determining occupational choices/outcomes. The model is set up as before, with the same explanatory variables as for Pakistan. We report marginal effects of the model and present graphs based on the results, with all the underlying regression results presented in Appendix Tables A1 to A4. Whenever education is included as an explanatory variable, we exclude the literacy and numeracy variables, and vice versa since these dimensions of skills are highly correlated, and we have no interest in documenting the effects of education conditional on literacy and numeracy skills or the other way around.

Table 2 shows marginal effects of the multinomial logit equation for selected variables: number of children, number of elderly people in the household and marital status. It is conspicuous that number of children and number of elderly people significantly reduces men's likelihood of being in wage-employment (which is highly paid) though, somewhat surprisingly, less strongly for women. This negative association could be because wage-employment is a less flexible occupation (in terms of working hours for example) but it may also be because of unobserved preferences: the kinds of people who prefer wage employment may also have preferences for smaller families. For men, being married strongly increases the likelihood of waged work and reduces the likelihood of being unemployed, OLF or agriculturally employed. For women being married increases the likelihood of working in agriculture and reduces the likelihood of being OLF.

The relationship between years of education and the predicted likelihoods of being in different labor market states is presented graphically rather than via marginal effects. Figure 1 shows the estimated association for men (panel i) and women (panel ii), evaluated at the sample mean values of the other explanatory variables in the model. Even though we cannot compare directly (since for Pakistan, we have separate graphs for young and old, but not for Ghana), it is clear that the role of education in occupational attainment in Ghana is extremely different to that in Pakistan. Firstly, the relationship between education and occupational choice is far more similar for men and women in Ghana than in Pakistan, where it varies dramatically by gender, likely reflecting the difference in the perceived gender role of women in a predominantly Muslim Asian society. Secondly, even for males only, the relationship between education and occupation is very different between Ghana and Pakistan, suggesting that very different forces are operating in the two countries' labor markets.

In Ghana, education strongly and monotonically reduces the chances of being in agriculture and raises the chances of wage employment both for men and women. Chances of unemployment, OLF and self-employment are largely invariant with respect to education though for women, education beyond secondary level reduces chances of being OLF.

Table 3 presents the marginal effects of basic literacy and numeracy on occupational attainment. Table 1 showed that wage employment is the best paying part of the labor market, followed relatively closely by self-employment and that agriculture is a very low paid occupation. Table 3 shows that being literate strongly promotes entry into the best paying part of the labor market, namely wage employment, roughly equally for both men and women. Literacy also correspondingly reduces the chances of ending up in poorly paid agriculture, again roughly equally for both men and women. However, literacy is not associated with differentially greater or lower chances of being in other labor market states. Possession of numeracy skills powerfully raises men's likelihood of wage employment and women's likelihood of self-employment. The direction of causation in the latter relationship is unclear. It could either run from being numerate to entering self-employment (numeracy promotes entry into self-employment) or from self-employment to becoming numerate (people in self-employment get a lot of practice in counting money so numeracy is learnt on the job). Either way, there is no such positive relationship between numeracy and self-employment for men . Being numerate also strongly reduces chances of ending up in agriculture for both men and women but the size of this marginal effect is significantly smaller for men than women. Gender differences in the relationship between skills and occupational outcomes could be due to the earnings rewards of numeracy differing for men and women, something we explore in the next section.

5. Education and Earnings

5.1 The basic relationship

Table 4a presents basic OLS estimates of the Mincerian earnings function in Ghana, by occupation and gender. Table 4b presents OLS estimates of the same equation by occupation and age group. Unlike Pakistan, where returns to education were very precisely determined for all the 12 sub-groups (3 occupations x 2 genders x 2 age groups) except women in agriculture, in Ghana returns are precisely determined only for the sample of waged workers, men or women as well as young and old, i.e. for 4 sub-groups. While in common with the literature we use the term 'returns to education', strictly speaking the coefficient on the Mincerian earnings function is simply the gross earnings premium from an extra year of education and is not the 'return' to education since it does not take the cost of education into account.

Table 4a shows that the average of marginal wage returns to education in Ghana is about 5% for both men and women. This contrasts strongly with that for Pakistan where wage returns are three to five times higher for women (15-17%) than men (3-6%). This greater premium on education for women is likely to reflect, at least in part, the greater scarcity of educated women in Pakistan than in Ghana, combined with the existence of predominantly 'female' jobs which require educated women, such as nursing and teaching. Appendix Table A9 shows that among wage employed individuals in Ghana, men's average education is only about 5% higher than women's but in Pakistan, it is 28% higher. However, later in Table 8 we show that the gender gap in returns to education is pro-female at secondary and tertiary levels of education in Ghana.

While the average slope of the education earnings profile is no steeper for women than men in Ghana, how about the intercept of the earnings equation? Table 4b which estimates returns for the young and old separately includes a gender dummy variable. It shows that men enjoy a hefty earnings premium in all occupations, varying from a premium (averaging across the young and the old) of 19% in wage employment to 35% in self-employment and 7% in agriculture. Thus, not only do women not have a higher slope in the education earnings relationship, they also have a lower intercept in the earnings function, i.e. their earnings do not catch up with men at higher levels of education. This contrasts with the case of Pakistan where although women have a lower intercept in the earnings function, they enjoy higher returns to education so that the gender earnings gap is significantly lower at higher levels of education. The graphs of predicted earnings in Figures 3 to 5 show this more clearly.

Table 4b also shows that wage returns to education for the young are statistically equal to those for the old. Again, this contrasts with Pakistan where returns for the young were very significantly lower than those for the old. The lower returns to education for the young in Pakistan was explained by the so-called 'filtering down' of occupations, the process by which successive cohorts of workers at a particular education level enter less and less skilled jobs within a given occupation (Knight, Sabot and Hovey, 1992). The lack of this 'filtering down' phenomenon in Ghana could be due to a less rapid expansion in the supply of educated persons in Ghana than in Pakistan over the past 40 years, though unfortunately we cannot test this due to lack of appropriate data.

Returns to education in self-employment and in agriculture are significantly lower than in wage employment, for both genders and age groups. They are also at best weakly statistically significant. Mincerian returns to education are 3.5% in self-employment for men and 2% each for women and old workers in agriculture and for old workers in self-employment. These findings are similar to those in Teal (2001), though they are not strictly comparable since he pools data from four household surveys rather than using data only for 1998-99. He finds that returns to education in wage employment were about 6%, in self-employment 2.5% and in agriculture 1%. The finding that returns to education in agriculture are much lower than those in other occupations is closer to earlier findings for Africa (Appleton, 2000) than to findings in Gallacher (1999, 2001) who finds that in Argentina, returns to education in agriculture for farms of average size was equal to the returns to education in wage employment[7].

The low returns to education in self-employment in Ghana are unfortunate because non-agricultural self-employment is the fastest growing occupation in Ghana (Teal, 2001) and it means that education is not an effective means of increasing incomes and reducing poverty for the part of the working population that is growing most rapidly. The very low returns to education in agriculture are also lamentable because agriculture absorbs a very high proportion of the workforce (64%) in Ghana. The gender pattern of returns in Ghana is not favourable for women either because, as well as having lower earnings than men in all occupations at zero education (the intercept of the earnings function being much higher for men than women), they do not enjoy a substantial returns-to-education premium over men in any occupation, i.e. the slope of the earnings function is (on average) not higher for them either, so that higher levels of education do not lead to a statistically significant reduction in the gender earnings gap.

[7] It could be argued that land and assets matter to profits in self employment and agricultural employment and are likely positively correlated with education so that the return to education could, in principle, be upwardly biased. However, in practice returns to education are very low even excluding land and assets. Thus, including these variables is not an issue in these data.

In summary, results show that education raises earnings in Ghana but only modestly and only in wage employment. The low returns to education in self-employment and agriculture suggest that education does not directly promote economic mobility for the large majority of workers in Ghana since these two occupations together constitute 82.5% of the employed workforce. This somewhat pessimistic conclusion is moderated when we consider that, as seen in section 4, education plays a major part in sorting people into highly paid occupations.

5.2 Extensions on the education-earnings relationship

Correcting returns estimates for endogeneity bias

As is well known, OLS estimates of returns to education potentially suffer from sample selectivity bias and endogeneity bias. We attempted to address the former by employing the Heckman procedure, explained in Section 2. The multinomial logit equations in the Appendix tables were used to calculate the selectivity terms for each occupation and worker-group. The results are presented in Appendix Table A5. The selectivity term is statistically significant in only 1 out of the 6 earnings regressions. The introduction of the selection term generally reduces the return to education but this reduction is statistically significant only in the case of male waged workers. Since selectivity correction makes little difference in the majority of cases, we prefer the OLS to the selectivity corrected equations, unlike in the case of Pakistan where the selectivity term was significant in many of the earnings functions for the different worker groups.

We approach the endogeneity of schooling by estimating a household fixed effects earnings function for waged work[8]. We cannot estimate this for self- and agricultural-employment since there is no within-household variation in these cases. The results in Table 5 show that correction for endogeneity bias does not change wage returns to education significantly. Fixed effects returns to

[8] We also sought to address the problem of the endogeneity of education by estimating the earnings function using a two stage least squares technique. Desirable instruments are not available, such as some rule that would change years of education in an exogenous way, or even variable such as distance to nearest school when the individual was of school-going age. We used spouse's education as an instrumental variable for waged workers' schooling as this was available much more commonly than father's or mother's education. The effect of instrumenting was to raise male waged workers' return to education to 9.4% (t=5.6) from the OLS estimate of 5% in Table 4a, and to reduce women's return to 4.0% (t=1.7) from the OLS estimate of 5.9% in Table 4a. Women's return from IV estimation is almost identical to that from the household fixed effects equation. The fact that the IV estimate of returns to waged men's education is higher than the corresponding OLS estimate is consistent with the fact that men's return from household fixed effects estimation is also higher than the OLS result, though the IV result is appreciably higher. However, we do not trust the IV estimates as much as the fixed effects estimates since the validity of the IV cannot be tested with a test of over-identifying restrictions and a priori spouse's education is a questionable instrument for worker education since the theory of assortative mating suggests that like people marry each other, i.e. spouse's education is likely to be correlated with the worker's unobserved characteristics.

education for men are 6.6% and for women 4.1% and neither is statistically significantly different from their OLS counterparts. As the family fixed effects equation provides a tighter upper bound for the estimate of the return to education, this gives us some confidence that our OLS results are closer to the true causal estimates of the effects of education on wages, and justifies the presentation of the OLS results for waged men in preference to the selectivity corrected results of Appendix Table 5.

Shape of the education-earnings relationship

So far we have imposed a linear relationship between 'years of education' and earnings in all occupations but it is not inevitable that the relationship will be linear. Table 6a relaxes the implicit presumption of linearity by introducing quadratic terms in education. The selectivity corrected counterpart of Table 6a is in Appendix Table A6[9]. Table 6a shows that in wage employment, the education-earnings relationship is strongly convex for both men and women. Thus, the Ghanaian labour market is not generally characterized by the commonly assumed concave relationship which implies diminishing returns to extra years of schooling and for which evidence has been found in the past (Psacharopoulos, 1994). Table 6a also shows that in agricultural employment it is weakly convex for men. The relationship is concave only for one group: self employed men. Table 6b estimates earnings functions separately for young and old workers and shows strong convexity in returns to education for both young and old persons in wage employment but not in other occupations.

The non-linearities of the education-earnings relationship are explored further in Table 7 which includes a dummy variable for each main education level. The selectivity correction estimator is relegated to Appendix Table A7. The base education category is 'no education'. Table 7 shows that the coefficients on education level dummies rise monotonically for both men and women in wage employment but that statistically significant earnings premia in wage employment exist only for secondary and tertiary education, confirming convexity. The marginal returns to each year of primary education, to each year of middle education and so forth, calculated from Table 7, are set out in Table 8. It shows that in wage employment marginal returns to tertiary education are lower for men (12.8%) than women (18%), but this gender difference is not statistically significant. By contrast the returns to both secondary and tertiary education in self-employment differ statistically significantly between men and women: men have higher returns to secondary education than women

[9] We do not estimate household fixed effects estimates of the earnings function for wage workers with either a quadratic term or with education *level* rather than years of education due to the very small sub-sample of households that have two or more members employed in waged work.

but women have higher returns to tertiary education than men. Returns to education do not differ significantly for the two genders in agriculture at any level of education.

Tables 6 to 8 taken together suggest that, with the exception of self-employed men, the education-earnings relationship in Ghana is generally not concave. Figures 3, 4 and 5 which show the relationship between education and predicted earnings, confirm this. Figure 3 shows pronounced convexity in wage returns for both men and women. Women's somewhat higher returns at secondary and tertiary education levels imply that the gender gap for waged workers is narrowed at high levels of education. While there is some suggestion of convexity for women in self-employment (Figure 4) and for both genders in agriculture (Figure 5), neither of these is statistically robust.

Earnings and Cognitive Skills

Returns in education may accrue not so much to completed years of education per se but rather to cognitive skills acquired, presumably through schooling. Differences in quality of education between different regions within the country and between schools within a region can mean that a given number of years of education leads to different levels of cognitive skills development across individuals. Table 9 shows earnings functions by occupation with cognitive skills measures on the right hand side. Years of schooling is not included in the earnings functions because we wish to estimate the total return to cognitive skills irrespective of whether they were acquired through schooling or not[10]. Corresponding selectivity corrected equations are presented in Appendix Table 8. Household fixed effects results are not reported because of very few households with two or more wage employed members. Table 9a shows results by gender and 9b by age.

Table 9a shows no significant returns to numeracy skills in any worker group but it shows substantial returns to literacy in wage employment for both men and women. While literacy has sizeable positive point estimates for men in both self-employment and agriculture, the coefficients are not precisely determined. In general women's earnings premia from literacy are statistically no different to men's. This result contrasts greatly from that for Pakistan where, in most cases, the returns to literacy are dramatically larger for women than men. This may be partly due to a greater scarcity premium for

[10] When we include years of education, the cognitive skills variables are insignificant and the coefficient on education remains virtually unchanged compared with the specification without the literacy and numeracy dummy variables. While this might be taken to suggest that education does not have its impact on earnings through raising cognitive skills, we are reluctant to draw this inference since the cognitive skills variables here are simple 0/1 dummies rather than a more informative continuous measure.

women in Pakistan than in Ghana[11] though Ghana's gender gaps in cognitive skills large enough in absolute terms that one would have expected a higher earnings premium to this more scarce skill among women. One plausible explanation for this apparent puzzle could be if labor roles are not so gender differentiated in Ghana. If men and women are substitutable in most jobs and can work alongside each other rather than having to be segregated for social reasons, then it is not necessary to reserve certain types of jobs for persons of particular genders.

When we divide the sample by age group the estimates of earnings functions in Table 9b show that there are large payoffs to literacy for older workers in wage employment and to young workers in self-employment, but not for other worker groups. Earnings premia for numeracy skills also exist for the young in self-employment and the old in self-employment. There are no significant returns to literacy or numeracy in agriculture for either age group, suggesting that Ghanaian agriculture is mainly traditional in that cognitive skills that would allow a person to, for example, follow instructions on fertilizer packs does not raise agricultural earnings.

That literate and numerate young workers can command an earnings premium in certain segments of the labour market is welcome and should encourage demand for education and the development of cognitive skills in Ghana. However, if the quality of education is low, it can take many years of schooling to develop literacy and numeracy and this highlights the importance of quality of schooling.

Comparing the return to education with the return to literacy and numeracy is not straightforward since cognitive skills are measured as 0/1 variables while education is a much more continuous variable. To examine the relationship between cognitive skills and education, we regressed the former on the latter and found that each year of education increases the probability of being literate by 10 percentage points for men and by 8 percentage points for women. In other words it takes men 10 years and women 12 years of education to acquire literacy. Table 9a showed that the coefficients on the literacy variable are 0.35 and 0.29 respectively for waged men and women. Thus the implied 'return' to literacy (rendered on the same scale as education) is 0.35/10 or 3.5% for waged men and 0.29/12 or 2.5% for waged women. This compares with a rate of return to education of about 5% for both men and women. In other words, the apparent return to cognitive skills is lower than the return to education. This suggests there is a substantial element of rent associated with education: education is used partly as a device to signal ability.

[11] Appendix Table A9 shows that while the gender gap in percentage of persons with literacy skills in all occupations in Ghana is large, it is nevertheless smaller than in Pakistan. Fewer women than men have the years of schooling required to develop literacy skills, though it is not known whether women are likely to have attended poorer quality schools than men, as in Pakistan (Aslam, 2007).

Heterogeneity in returns to education

While the simple Mincerian earnings function supposes that the marginal return to education is the same for all individuals, this is a restrictive assumption. In practice economic returns to education can vary across people due to a number of unobserved factors such as ability, motivation and ambition as well as due to differences in interest rate faced by different individuals, based for instance on wealth/assets (Card, 2001). The fact that returns to education can be heterogeneous across individuals has implications for the inequality-reducing role of education. To our knowledge, the distribution of returns to education across the earnings spectrum has not been investigated for Ghana, as for most other developing countries (Patrinos, Ridao-Cano and Sakellariou, 2006). We examine heterogeneity in returns to education to ask whether some workers benefit more from education than others and why, and the inequality implications of that.

The Pakistan case study already alluded to the literature investigating the pattern of returns to an additional year of education along the earnings distribution using quantile regressions (QR). It noted the suggestion in this literature that in developed countries returns to education increase with quantiles (returns are higher for higher earnings quantiles), in middle-income countries the evidence is mixed, and in the few developing countries for which evidence exists, returns decrease with quantiles. If returns to education increase as one goes from the lower to the higher end of the earnings distribution, this can be interpreted as indicating that ability and education complement each other, with more able workers benefiting more (in terms of higher earnings) from additional investment in education. On the other hand, a negative relationship between ability and returns to education (decreasing returns with earnings quantiles) suggests substitutability between education and ability. Finally, if there is no distinct pattern, then average returns (in the absence of biases in their estimation) capture the overall profitability of education.

Table 10 reports the quantile regression results. The top half presents results for men and women. It shows that in wage employment, for both men and women, there is a consistent pattern of returns to education being different at different points of the conditional earnings distribution. Returns to education are highest in the lowest quantile of earnings (bottom quartile) and lowest in our highest earnings group (the top quartile). For both men and women the difference between the top and bottom quartiles is statistically significant, though the size of the difference is nearly twice as big for women as for men. While a 1.6 point difference (5.8% – 4.2%) in returns to education between the top and bottom earnings quartiles for waged males is not trivial, in the case of waged women, the difference of 2.8 points (8.1 – 5.3) is economically quite large. Similar results obtain for self-employed women, for whom returns to education in the top earnings quartile are significantly lower than those in the bottom earnings quartile, a difference of 8 points. Thus, in these worker and occupation groups, those with lower ability have higher rates of return to education, lending support to the notion of substitution between ability and education. This suggests that among waged men and women and among self-employed women, education is inequality reducing, since education lowers the wage differences between low and high ability individuals, rather than increasing them. However, among self-employed men, education appears to be inequality-increasing: the return to education in the top earnings quartile is nearly double the returns at median earnings and that is weakly higher than returns to education in the bottom quartile. There are no such patterns discernible in agriculture.

The bottom half of Table 10 presents results by age group. This shows that among old waged workers, returns to education are the highest for the bottom earnings quartile and lowest for the top earnings quartile, suggesting that education is inequality-reducing among older waged workers, though the size of the difference is not economically large. There is no suggestion of any pattern either increasing or decreasing returns to education by earnings quantile in any of the other five age-occupation categories.

The fact that education is inequality reducing in wage employment and among women in self-employment is welcome because it suggests that there is a social externality from education.

Returns to vocational education and training

There is currently a revival of the old debate about the relative efficacy of vocational versus general education in many countries. Several developed countries are strengthening their vocational education systems, ostensibly to cater for the less academic youth. The Ghana Living Standards Survey 1998-99 provides data on whether a worker had received any technical and vocational education and training (TVET). Table 1 shows that a non-negligible proportion of the employed workforce received TVET (5.9% of self-employed and 9.7% of wage employed, though only 1.4% of agriculturally employed) unlike in the case of Pakistan where less than 2% of employed persons had TVET. Table 11 presents results of the simple earnings function with a dummy variable included for TVET. The variable has a negative coefficient in the case of 4 out of 6 worker groups and when it has a positive coefficient, it is wholly statistically insignificant.

One explanation for TVET having a negative coefficient is that it is associated with entry into the less well paid narrowly-defined occupations (e.g. because people who opt for the vocational education stream have lower endowments of certain forms of ability) but that *within* a given occupation, those with TVET enjoy an earnings premium. In order to test this hypothesis, we re-estimated the wage equation for men and women including 7 occupation dummies (professional and managerial, clerical, sales, service, production, agricultural-wage-labour and skilled-crafts occupations). This is reported in Table 12. Even after the inclusion of occupation, however, the TVET variable has a negative coefficient of -0.139 (t=-1.45). This suggests that we have not been able to control for the relevant dimension of occupation since our occupation dummies are not detailed enough. Although the data provides finer occupational categories, our sample size is too small to include the full 2-digit occupation dummy variables. We conclude that the apparently negative relationship between TVET and earnings is not causal, i.e. TVET does not lower earnings, conditional on job. It is associated with entry into the lower paid jobs within broadly defined occupational categories.

Kahyarara and Teal (2006a) find that general education is more rewarding than vocational education in Tanzania where the marginal return to one year of education ranges between 4.8 and 17.5 percent compared to the return to one year of vocational education that ranges between 1.4 and 2.8 percent. This and Kahyarara and Teal (2006b) highlight the importance of panel data which enables the effects of unobservable worker and firm characteristics to be identified in assessing returns to both vocational education and training. Their results are stable even after they control for endogeneity, worker characteristics and firm fixed effects. Their research shows that the level of general education after which vocational education is acquired matters for returns to vocational education. While we do not have panel data, we did attempt to estimate returns to TVET acquired after different levels of general education. However, all the interaction terms between TVET and education were statistically insignificant, presumably because only less than 10% of the sample have any TVET. Since Kahyarara and Teal's data was a survey of manufacturing firms, a much higher percentage of workers had TVET in their data.

Finally, the relationship between apprenticeship training and labour market earnings (not shown) was very similar to that of TVET and earnings, i.e. the apprenticeship variable had a negative coefficient but again we do not draw any causal inferences from that, though it suggests that apprenticeship is not the route to higher wage employment in Ghana.

6. Conclusions

This paper has explored varied aspects of the education-earnings relationship in Ghana. It has examined (i) the role of education in occupational attainment; (ii) the role of education in earnings, conditional on occupation; (iii) the role of cognitive skills in both occupational attainment and earnings determination; (iv) the role of education in earnings at different points of the conditional earnings distribution and (v) the role of technical and vocational education and training in earnings determination. We have also examined the shape of the education earnings relationship.

The instrumental benefits of education arise both from its role in promoting a person's entry into the lucrative occupations and, conditional on occupation, from its role in raising earnings. Our results suggest that education plays a very important role in occupational outcome particularly the wage employment versus agriculture outcome, though it has less bearing on sorting into other labour market states.

While education raises earnings *indirectly* by helping individuals to gain entry into the high paying occupations, it has low *direct* effects on earnings. Results show that education raises earnings only modestly and that only in wage employment. It does not *directly* raise earnings for the large majority of workers in Ghana since returns to education in self-employment and agriculture are very low and since these two occupations together constitute 82.5% of the employed workforce. While it may seem that the economic incentives for acquiring schooling may we weak in Ghana, two considerations go against this. Firstly, education has large indirect effects via promoting entry into (well paying) wage employment. Secondly, the returns to education mostly increase with education level in Ghana so that there will be an economic incentive to reach the higher levels of education where returns are substantial.

Looking at whether the role of education differs for the two genders, the results show that, unlike the case of Pakistan, the marginal effect of education in occupational attainment is remarkably similar for the two genders. Again, in contrast to Pakistan, the relationship of education with conditional earnings is also virtually identical for the two genders (and also for the young and old age groups). The gender gap in education is not large in Ghana and in any case the labour market appears not to be segmented by gender. Does the fact that women's education reaps economic rewards equal to those from men's education mean that girls will have the same economic incentives to acquire schooling as boys in Ghana? Unfortunately we cannot conclude that from a mere examination of the returns to education for the two genders (the slopes of the earnings functions for men and women). One must also ask whether overall earnings are equal for men and women, i.e. one must also examine the intercept of the earnings function and not only the slope. Not only is women's return to education no higher than men's (no scarcity premium for women's education), overall women's earnings are much lower than men's (the intercept of the earnings function is a lot lower for women than men) and the gender gap in earnings does not narrow significantly at higher levels of education. This is evidence of gender differentiation in the labour market, though to establish whether this is gender discrimination one needs further analysis and better data, including accurate measures of labor market experience and quality of schooling of men and women. We conclude that there appears to be a prima facie case for policies that discourage gender differentiated treatment by employers in the labor market.

As for Pakistan, we find for Ghana too that the shape of the education earnings relationship is not concave, with diminishing returns to education, as conventional wisdom suggests. In wage employment for both men and women (and to a lesser extent in agriculture for men), the relationship is convex, i.e. the high returns accrue only at the higher (e.g. secondary and tertiary) levels of education. This means that increasing education by small amounts at low education levels will not raise earnings substantially and will not prove an effective means of helping people to climb out of poverty.

We estimated returns to education along the earnings distribution. There is a clear pattern in wage employment where education is inequality-reducing: among both genders, lower ability waged workers have higher returns to education than higher ability ones. As with Pakistan, the inequality reducing role of education dampens over time: it exists for older waged workers but not for the younger ones. In other occupations the pattern of returns along the distribution of conditional earnings is not so clearcut.

The paper also examined relationships between numeracy and literacy on the one hand and occupational outcomes and earnings on the other. We find that literacy and numeracy both strongly promote entry into the lucrative parts of the labor market for both men and women. Conditional on occupation, literacy skills have moderately large pay-offs for both genders in Ghana, though these are confined mainly to wage employment. While there is suggestion that literacy also raises earnings for men in self-employment and agriculture, these relationships are not statistically significant.

Possession of numeracy skills also have moderately high coefficients for various worker groups but they are never statistically significant. While in some cases this may be put down to the small sample problem, in other cases, this is clearly not the reason, for instance, in agriculture.

Lessons for future research

What have we learnt from this research about how household survey data can be used to analyse labour markets in developing countries? Firstly, the research has highlighted the importance of estimating returns to education separately in the different occupations rather than estimating them only in wage employment, unlike much of the existing applied labor economics literature for developing countries. Wage employment is typically a small and increasingly a shrinking part of the labor market in many developing countries. The substantial difference in returns to education in different occupations in this paper showcases the importance of estimating returns to education separately by occupation.

Secondly, the paper highlights the importance of recognising that even if education has low direct returns to education, it may raise earnings indirectly by facilitating entry into the lucrative occupations. The case of Ghana shows this to be true, though in Pakistan education has payoffs both indirectly in terms of improved occupational attainment and directly in terms of raising earnings substantially conditional on occupation.

Thirdly, the comparison of Pakistan and Ghana case studies makes it clear that sample size matters. Large household surveys that furnish reasonably large samples of workers can permit more disaggregated analysis and help to reach more nuanced understandings about the differing role of education for different worker groups. Large samples also allow the researcher to use some more demanding econometric techniques that permit more reliable inferences. For instance, large samples can yield a sufficient number of households with two or more members in a given occupation (e.g. wage employment) and thus enable estimation of family fixed effects earnings functions, which give a much tighter upper bound on the true causal effect of education on earnings since they net out many aspects of unobserved traits that are shared among members within a family. In our Ghana case study, due to small sample sizes, we could not identify the coefficients on education using the quadratic and levels specifications in a family fixed effects equation.

Fourthly, the study has highlighted the importance of paying attention to potential statistical biases when assessing the effect of education on labor market outcomes. It has explained sample selectivity and endogeneity biases in estimating the returns to education and shown how these issues may be addressed. While we have used a household fixed effects methodology for addressing the problem of endogeneity bias (also known as ‘ability bias’), an alternative technique is the instrumental variables method but for that we would have needed ‘instruments’, i.e. variables that affect years of schooling acquired but do not affect earnings other than through their effect on years of education. For instance, distance to school from home when the individual was of school-going age is one ‘instrument’ used in the literature. It is helpful if this becomes an important consideration when planning surveys for labor market analyses. Of course the ideal would be if data exist on some administrative rule change which affects the years of education acquired and which would be exogenous from the point of each individual. For instance, a rise in the school leaving age from 14 to 15 in 1947 and another from 15 to 16 in 1973 in the UK permitted Harmon and Walker (1995) to estimate the true causal return to education, uncontaminated by the effect of ‘ability bias’.

References

Alderman, H., Orazem, P.F. and Paterno, E.M. (2001), “School Quality, School Cost, and the Public/Private School Choices of Low-Income Households in Pakistan”, *Journal of Human Resources*, Vol. 36 (2), pp. 304-326.

Andrabi, T.,Das, J. and Khwaja, A. (2002), “The Rise of Private Schooling in Pakistan: Catering to the Urban Elite or Educating the Rural Poor?” World Bank and Harvard University.

Appleton, Simon and Arsene Balihuta (1996) “Education and agricultural productivity in Uganda”, *Journal of International Development*, Vol.8, No.3, pp.415-444, May-June 1996.

Arias, O., K. Hallock and W. Sosa-Escudero (2001) “Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data”, *Empirical Economics*, Fall 2000.

Arif, G.M. and Saqib, N. (2003), ‘Production of Cognitive Life Skills in Public, Private, and NGO Schools in Pakistan’, *Pakistan Development Review*, 42 (1), pp. 1-28.

Aslam, M. and Kingdon, G.G. (2006), ‘Gender and Household Education Expenditure in Pakistan’, CSAE Working Paper Series GPRG-WPS 025, University of Oxford. Forthcoming in *Applied Economics*.

Aslam, M. (2007), ‘Rates of Return to Education by Gender in Pakistan’, GPRG-WPS-064, Centre for the Study of African Economies, University of Oxford.

Behrman, J., D. Ross and R. Sabot (2002) “Improving the Quality Versus Increasing the Quantity of Schooling: Evidence from Rural Pakistan”, University of Pennsylvania, mimeo, 2002

Buchinsky M. (1994) “Changes in the U.S. wage structure 1963-1987: an application of quantile regression”, *Econometrica* 62: 405-458

Gallacher, M. (1999), “Human capital and production efficiency: Argentine agriculture”, CEMA Working Paper 158, Universidad del CEMA, Buenos Aires, Argentina.

Gallacher, M. (2001) “Education as an Input in Agricultural Production: Argentina”, CEMA Working Papers 189, Universidad del CEMA, Buenos Aires, Argentina.

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

Hanushek, Eric A (2005) "The Economics of School Quality." *German Economic Review* 6, no. 3 (August):269-286.

Hanushek, E. and L. Woessmann (2007) “The Role of School Improvement in Economic Development”, Working Paper PEPG 07-01, Program of Education Policy and Governance, Kennedy School of Government, Harvard University.

Hanushek, Eric A., and Lei Zhang. 2006. "Quality Consistent Estimates of International Returns to Skill." National Bureau of Economic Research, WP12664, Cambridge, MA, NBER November.

Harmon, C. and I. Walker (1995) “Estimates of the economic return to schooling for the United Kingdom”, *American Economic Review*, Vol. 85, No. 5, 1278-1286.

Heckman, J. (1979) Sample Selection Bias as a Specification Error. *Econometrica*, Vol. 47, No. 1, 153-162. Jan., 1979.

Kahyarara, G. and F. Teal (2006a) “General or Vocational Education? Evidences from the Returns to Education in Tanzanian Manufacturing Firms”, CSAE mimeo. University of Oxford.

Kahyarara, G. and F. Teal (2006b) “To train or to educate? Evidences from Tanzania”, GPRG Working Paper 051, Department of Economics, University of Oxford.

Knight, J., R. Sabot and D. Hovey (1992) “Is the Rate of Return to Primary Schooling Really 26 Percent?” *Journal of African Economies*.

Koenker R, Bassett G. (1978) Regression quantiles. *Econometrica* 46: 33-50

Lee, L F (1983) “Generalized Econometric Models with Sample Selectivity”, *Econometrica*, 51, No. 2, March 1983.

Machado J. and J. Mata (2000) Box-Cox quantile regression and the distribution of firm sizes. *Journal of Applied Econometrics* 15: 253-274

Patrinos, H., Ridao-Cano and C.Sakellariou (2006) “Estimating the returns to education: Accounting for heterogeneity in ability”, World Bank, World Bank Policy Research Working Paper 4040, October 2006

Schultz, T., and G. Mwabu (1998): “Labor unions and the distribution of wages and employment in South Africa,” *Industrial and Labor Relations Review*, 51(4), 680–703.

Teal, F. (2001) “Education, incomes, poverty and inequality in Ghana in the 1990s”, WPS/2001-21, Centre for the Study of African Economies, Department of Economics, University of Oxford. November.

Table 1
Ghana 1998-99. Full sample: Summary statistics by occupation

	All	Self-employed	Agricult. employed	Wage employed	Unemp-loyed	Out of labor force
Mean annual earnings (cedi)	1,350,471	2,096,982	839,476	2,211,669	0	0
Median annual earnings	629,254	963,250	423,775	1,703,000	0	0
Log earnings	13.28	13.78	12.82	14.22	0	0
Years of education	5.7	6.6	3.7	10.5	8.9	5.6
Vocational education %	3.5	5.9	1.4	9.7	8.5	2.9
Age	37.1	36.2	39.9	38.6	31.0	32.9
Proportion men	0.47	0.29	0.49	0.75	0.57	0.36
Math skills	0.54	0.65	0.39	0.86	0.81	0.56
Read & write skills	0.47	0.55	0.33	0.83	0.71	0.48
# children aged < 10 in household	1.52	1.41	1.73	1.16	1.01	1.46
# individuals aged > 70 in household	0.07	0.07	0.08	0.03	0.08	0.09
Proportion married	0.51	0.56	0.57	0.60	0.25	0.37
Observations	9613	1157	4438	1185	129	2704
Earning observations	6780	1157	4438	1185	0	0

Note: These are weighted means. The exchange rate on 30th September 1998 was USD 1 = Cedis 2350. Thus, annual mean earnings in USD were $575. Read and write skills in own language or English.

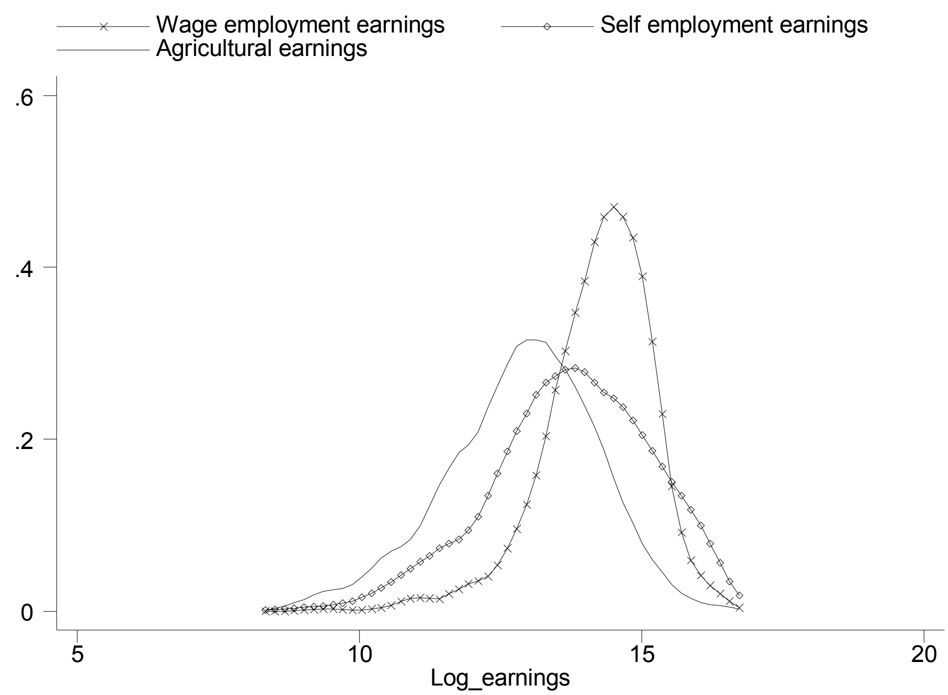


Table 2
Selected partial effects on the likelihood of occupational outcome, by gender and age group

	Men	Women
1. Self-employment		
# children aged < 10 in household	-0.006	-0.008
	(-1.98)*	(-2.23)*
# individuals aged > 70 in household	-0.004	-0.021
	(-0.25)	(-1.19)
Individual is married	0.019	0.003
	(1.68)$^{+}$	(0.23)
2. Agriculture		
# children aged < 10 in household	0.034	0.022
	(7.07)**	(5.13)**
# individuals aged > 70 in household	0.101	0.040
	(3.73)**	(2.05)*
Individual is married	-0.031	0.030
	(1.78)$^{+}$	(2.14)*
3. Wage employment		
# children aged < 10 in household	-0.033	-0.005
	(-7.41)**	(-1.77)$^{+}$
# individuals aged > 70 in household	-0.114	-0.018
	(-3.88)**	(-1.20)
Individual is married	0.081	-0.003
	(5.41)**	(-0.51)
4. Unemployed		
# children aged < 10 in household	-0.002	0.000
	(-1.32)	(-0.31)
# individuals aged > 70 in household	-0.002	0.000
	(-0.19)	(-0.01)
Individual is married	-0.009	-0.002
	(-2.60)**	(-0.85)
5. Out of labor force		
# children aged < 10 in household	0.007	-0.009
	(1.71)$^{+}$	(-1.99)*
# individuals aged > 70 in household	0.019	-0.002
	(0.86)	(-0.09)
Individual is married	-0.060	-0.027
	(-4.24)**	(-2.02)*

Note: These results are based on the multinomial logits reported in Appendix Tables 1 and 2. Robust t-statistics in parentheses. $^{+}$ significant at 10% level; * significant at 5% level; ** significant at 1% level.

Table 3
The partial effects of literacy and numeracy on occupational outcome, by gender and age group

	Men	Women
1. Self-employment		
Can solve simple maths problem	-0.001	0.101
	(-0.07)	(5.62)**
Can read & write	0.014	0.010
	(0.84)	0.65
2. Agriculture		
Can solve simple maths problem	-0.097	-0.160
	(-3.59)**	(-7.72)**
Can read & write	-0.183	-0.162
	(-7.15)**	(-7.37)**
3. Wage employment		
Can solve simple maths problem	0.117	0.014
	(4.48)**	1.05
Can read & write	0.142	0.131
	(5.92)**	(6.15)**
4. Unemployed		
Can solve simple maths problem	0.008	0.004
	(0.80)	(0.70)
Can read & write	0.003	-0.001
	(0.41)	(-0.17)
5. Out of labor force		
Can solve simple maths problem	-0.027	0.041
	(-1.20)	(1.94)$^{+}$
Can read & write	0.025	0.021
	(1.12)	(1.01)

Note: These results are based on the multinomial logits reported in Appendix Tables 3 and 4. Can read or write in native language or English = 1; else =0. $^{+}$ significant at 10% level; * significant at 5% level; ** significant at 1% level.

Table 4a
Earnings and years of schooling, by gender

	1. Wage employed		**2. Self-employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
Education	0.050	0.059	0.035	0.013	0.009	0.021
	(7.82)**	(5.90)**	(1.77)$^{+}$	(1.16)	(0.91)	(2.00)*
Age	0.109	0.121	0.067	0.118	0.075	0.038
	(5.11)**	(3.30)**	(1.45)	(4.57)**	(4.06)**	(2.06)*
Age squared	-0.001	-0.001	-0.001	-0.001	-0.001	-0.000
	(4.13)**	(2.54)*	(1.63)	(4.46)**	(3.72)**	(2.08)*
# individuals	898	287	338	819	2098	2340

Note: Robust t-statistics in parentheses. $^{+}$ significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions. The estimation method is OLS.

Table 4b
Earnings and years of schooling, by age group

	1. Wage employed		**2. Self-employed**		**3. Agriculture**	
	Young	Old	Young	Old	Young	Old
Education	0.048	0.055	0.012	0.022	0.009	0.019
	(5.20)**	(8.03)**	(0.75)	(1.77)$^{+}$	(0.70)	(2.34)*
Male	0.227	0.158	0.593	0.107	-0.041	0.184
	(2.14)*	(2.43)*	(3.27)**	(0.83)	(0.31)	(2.29)*
Age	-0.007	0.101	0.218	0.083	-0.234	0.010
	(0.04)	(2.44)*	(0.71)	(1.68)$^{+}$	(1.23)	(0.35)
Age squared	0.001	-0.001	-0.003	-0.001	0.006	-0.000
	(0.30)	(2.24)*	(0.56)	(2.01)*	(1.48)	(0.52)
# individuals	299	886	418	739	1351	3087

Robust t-statistics in parentheses. $^{+}$ significant at 10% level; * significant at 5% level; ** significant at 1% level. All regressions include province variables. Young persons are those aged 16-30. 'Old' are persons aged 31 to 70.

Table 5
Earnings and years of schooling among the wage employed, by gender:
Controlling for household fixed effects

	Men	**Women**
Education	0.066	0.041
	(3.02)**	(1.37)
# individuals	898	287

Note: Absolute value of t-statistics in parentheses. * significant at 5% level; ** significant at 1% level. Age, age squared are included in all regressions.

Table 6a
Earnings and years of schooling, by gender
Quadratic term included: OLS estimates

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
Education	0.010	0.006	0.131	-0.006	-0.038	0.011
	(0.53)	(0.23)	(2.30)*	(0.18)	(1.29)	(0.31)
Education squared	0.002	0.003	-0.007	0.002	0.004	0.001
	(2.52)*	(2.30)*	(1.80)+	(0.59)	(1.68)+	(0.32)
Age	0.112	0.127	0.054	0.118	0.077	0.039
	(5.21)**	(3.44)**	(1.16)	(4.57)**	(4.19)**	(2.09)*
Age squared	-0.001	-0.001	-0.001	-0.001	-0.001	-0.000
	(4.24)**	(2.67)**	(1.26)	(4.47)**	(3.89)**	(2.11)*
# Individuals	898	287	338	819	2098	2340
Mean years of education	10.5	10.1	8.5	5.7	5.3	2.4
Return to education (at mean education)	5.2	6.7	1.2	1.7	0.4	1.6

Note: Robust t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions. The estimation method is OLS.

Table 6b
Earnings and years of schooling, by age group,
quadratic term included: OLS estimates

	1. Wage employed		**2. Self-employed**		**3. Agriculture**	
	Young	Old	Young	Old	Young	Old
Education	-0.016	0.019	0.034	0.035	0.008	-0.021
	(0.64)	(0.93)	(0.72)	(0.96)	(0.19)	(0.81)
Education squared	0.004	0.002	-0.002	-0.001	0.000	0.004
	(2.83)**	(2.16)*	(0.49)	(0.39)	(0.05)	(1.64)
Male	0.235	0.144	0.598	0.106	-0.040	0.186
	(2.26)*	(2.21)*	(3.29)**	(0.82)	(0.31)	(2.32)*
Age	-0.042	0.108	0.235	0.082	-0.236	0.011
	(0.22)	(2.61)**	(0.76)	(1.68)	(1.23)	(0.40)
Age squared	0.002	-0.001	-0.004	-0.001	0.006	-0.000
	(0.47)	(2.41)*	(0.61)	(1.99)*	(1.49)	(0.61)
# individuals	299	886	418	739	1351	3087
Mean years of education	9.8	10.6	7.3	6.0	4.8	3.4
Return to education (at mean education)	6.2	6.1	0.5	2.3	0.8	0.6

Robust t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. All regressions include province dummy variables. Young and old are defined as in Table 4b.

Table 7
Earnings and the level of schooling, OLS estimates

	1. Wage employed		2. Self employed		3. Agriculture	
	Men	Women	Men	Women	Men	Women
Primary	0.042	0.147	0.752	0.026	-0.159	0.167
	(0.24)	(0.66)	(2.26)*	(0.17)	(1.11)	(1.32)
Middle school	0.306	0.151	0.354	0.198	-0.401	-0.063
	(1.10)	(0.34)	(0.57)	(0.58)	(1.51)	(0.21)
Secondary	0.335	0.362	0.682	0.065	0.038	0.219
	(2.84)**	(2.09)*	(2.65)**	(0.50)	(0.33)	(1.82)+
Tertiary	0.718	0.902	0.462	0.380	0.251	0.252
	(6.03)**	(5.16)**	(1.38)	(1.74)+	(1.36)	(0.74)
Age	0.115	0.126	0.041	0.123	0.064	0.034
	(5.08)**	(3.11)**	(0.82)	(4.53)**	(3.12)**	(1.75)+
Age squared	-0.001	-0.001	-0.001	-0.001	-0.001	-0.000
	(4.20)**	(2.43)*	(0.95)	(4.45)**	(2.96)**	(1.79)+
# Individuals	898	287	338	819	2098	2340

Note: Robust t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions. The estimation method is OLS. The omitted education category is no education. The education levels are defined as follows: primary = 1-6 years of education; middle school = 7-9 yrs; secondary = 10-12 yrs; tertiary = 13+ years.

Table 8
Estimated return to an additional year of schooling, by level of education
(Using OLS earning function from Table 7)

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
Primary	0.7	2.5	12.5 *	0.4	-2.7	2.8
Middle school	8.8	0.1	-13.3	5.7	-8.1	-7.7
Secondary	0.2	7.0	10.9	-4.4	14.6	9.4
Tertiary	12.8 **	18.0 **	-7.3	10.5 +	7.1	1.1

Note: The marginal return to a year of primary schooling is calculated as the coefficient on the primary school dummy variable divided by 6, since there are 6 years in the primary school cycle. The marginal return to a year of middle level schooling is calculated as the coefficient on the middle school dummy minus the coefficient on the primary school dummy, divided by 3 since there are 3 years in the middle school cycle (grades 7, 8 and 9); and so on for other levels of education. Both secondary and tertiary levels of education are assumed to be 3 year long cycles. * indicates that the marginal return to education at a given *level* of education is statistically significantly different (at the 5% level) from the marginal return at the education level immediately lower than it. Among men in self-employment, for instance, the return to each extra year of education at the primary level is significantly greater than the return to zero years of education a and thus, 12.5 has a * by it. Similarly, 18.0 is statistically significantly different from 7.0 (marginal return to tertiary education is significantly greater than that to secondary education) and hence 18.0 has a ** by it. The coefficients on the education level dummies are not precisely determined and thus, even seemingly large differences in marginal returns at different levels of education are not significantly different from each other, e.g. below tertiary level in wage employment.

Table 9a
Earnings, literacy and numeracy, by gender

	1. Wage employed		2. Self employed		3. Agriculture	
	Men	Women	Men	Women	Men	Women
Can solve simple	0.165	0.112	0.231	0.179	-0.217	0.209
maths problem	(1.09)	(0.63)	(0.72)	(1.12)	(1.36)	(1.57)
Can read & write	0.354	0.289	0.427	-0.044	0.226	-0.010
	(2.72)**	(1.66)+	(1.51)	(0.29)	(1.48)	(0.07)
Age	0.115	0.141	0.053	0.118	0.075	0.038
	(5.25)**	(3.67)**	(1.14)	(4.55)**	(4.02)**	(2.03)*
Age squared	-0.001	-0.001	-0.001	-0.001	-0.001	-0.000
	(4.39)**	(2.97)**	(1.31)	(4.42)**	(3.76)**	(2.05)*
# individuals	898	287	338	819	2098	2340

Note: Robust t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions. The estimation method is OLS.

Table 9b
Earnings, literacy and numeracy, by age

	1. Wage employed		2. Self employed		3. Agriculture	
	Young	Old	Young	Old	Young	Old
Can solve simple	0.386	0.048	-0.094	0.351	0.002	0.125
maths problem	(1.65)+	(0.37)	(0.41)	(1.89)+	(0.01)	(1.00)
Can read & write	0.082	0.462	0.367	-0.052	0.154	0.052
	(0.38)	(4.01)**	(1.78)+	(0.30)	(0.90)	(0.41)
Age	0.078	0.093	0.183	0.082	-0.222	0.003
	(0.41)	(2.12)*	(0.60)	(1.68)+	(1.17)	(0.11)
Age squared	-0.000	-0.001	-0.003	-0.001	0.006	-0.000
	(0.07)	(2.00)*	(0.46)	(1.97)*	(1.44)	(0.32)
# individuals	898	287	338	819	2098	2340

Note: Robust t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions. The estimation method is OLS.

Table 10
Earnings and years of schooling: Quantile regressions

	1. Wage employed	2. Self employed	3. Agriculture
Men			
Education,	0.058	0.033	0.014
(25th percentile of earnings)	(7.34)**	(1.42)	(1.18)
Education	0.049	0.042	0.013
(50th percentile of earnings)	(13.63)**	(1.84) +	(1.10)
Education	0.042	0.079	-0.006
(75th percentile of earnings)	(7.46)**	(3.53)**	(0.51)
N	898	338	2098
Women			
Education,	0.081	0.034	0.013
(25th percentile of earnings)	(5.76)**	(2.60)**	(0.99)
Education	0.066	-0.002	0.026
(50th percentile of earnings)	(9.60)**	(0.14)	(2.14)*
Education	0.053	-0.046	0.032
(75th percentile of earnings)	(4.92)**	(3.63)**	(2.66)**
N	287	819	2340
Young workers			
Education,	0.051	0.044	-0.005
(25th percentile of earnings)	(4.22)**	(2.10)*	(0.24)
Education	0.047	0.013	0.013
(50th percentile of earnings)	(5.69)**	(0.69)	(1.05)
Education	0.050	0.011	0.004
(75th percentile of earnings)	(6.11)**	(0.58)	(0.23)
N	299	418	1351
Old workers			
Education,	0.060	0.033	0.025
(25th percentile of earnings)	(7.32)**	(2.20)*	(2.00)*
Education	0.056	0.018	0.032
(50th percentile of earnings)	(11.56)**	(1.23)	(4.19)**
Education	0.047	0.027	0.024
(75th percentile of earnings)	(9.07)**	(1.90)+	(2.42)*
N	886	739	3087

Note: Age, age squared, and province dummy variables are included in all regressions. t-values in parentheses.

Table 11
Earnings and Technical and Vocational Education and Training (TVET)

	1. Wage employed		**2. Self-employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
Education	0.012	0.005	0.109	-0.016	-0.021	0.007
	(0.68)	(0.17)	(2.11)*	(-0.48)	(-1.02)	(0.27)
Education squared	0.002	0.003	-0.005	0.003	0.002	0.001
	(2.50)*	(2.11)*	(-1.57)	(1.14)	(1.19)	(0.23)
Age	0.113	0.127	0.030	0.098	0.053	0.030
	(6.36)**	(4.62)**	(0.71)	(4.16)**	(4.70)**	(2.76)**
Age squared	-0.001	-0.001	0.000	-0.001	-0.001	0.000
	(-5.42)**	(-3.69)**	(-0.84)	(-3.97)**	(-4.37)**	(-2.84)**
TVET	-0.126	0.109	-0.008	-0.488	0.139	-0.140
	(-1.22)	(0.63)	(-0.03)	(-1.79)	(0.70)	(-0.43)
# individuals	898	287	338	819	2098	2340

Note: Robust t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions. The estimation method is OLS. TVET is technical and vocational education and training.

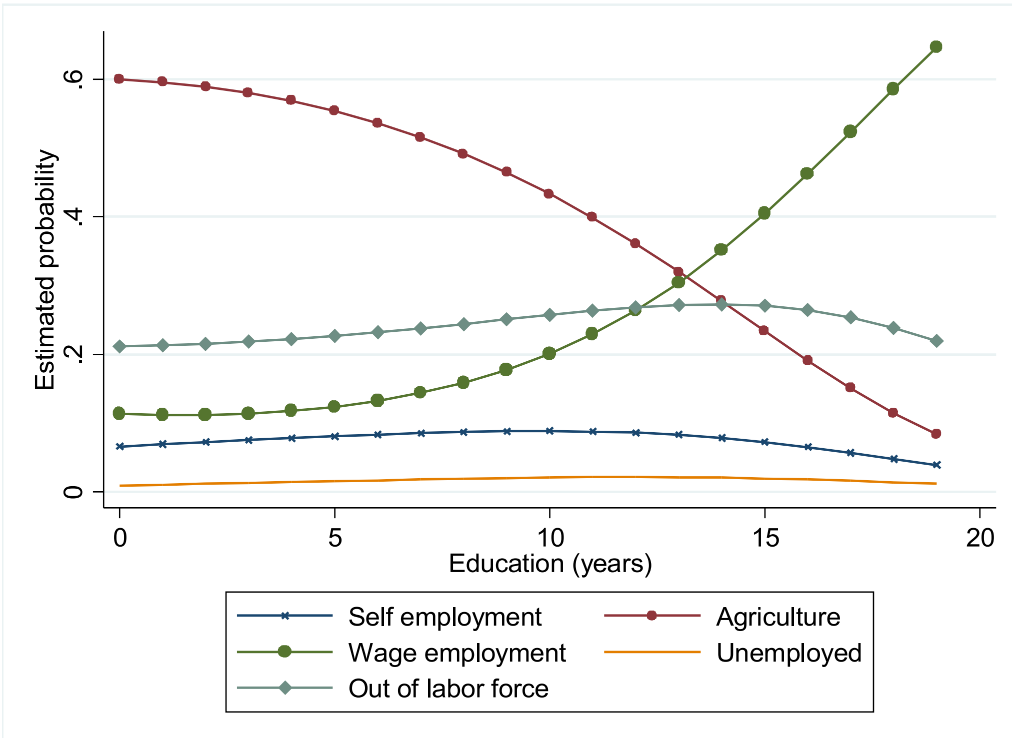
Table 12
Earnings and Technical and Vocational Education and Training (TVET)
with occupation dummies, Wage employment only

	Men			**Women**		
Education	-0.002	-0.08		-0.014	-0.48	
Education squared	0.003	2.71	**	0.003	2.54	**
Age	0.109	5.14	**	0.114	3.08	**
Age squared	-0.001	-4.21	**	-0.001	-2.39	*
TVET	-0.140	-1.45		0.099	0.65	
Professional/managerial	0.141	1.08		0.106	0.53	
Clerical	0.254	2.13	*	0.403	1.81	+
Sales	0.056	0.40		-0.094	-0.39	
Service	0.113	1.03		-0.110	-0.56	
Production	0.685	5.55	**	-0.005	-0.02	
Agricultural	-0.070	-0.42		-0.374	-1.19	
Skilled craftsmen	0.352	2.59	**	---	---	
_cons	11.204	26.78	**	11.372	17.28	**
N	898			287		
R-square	0.2279			0.3653		

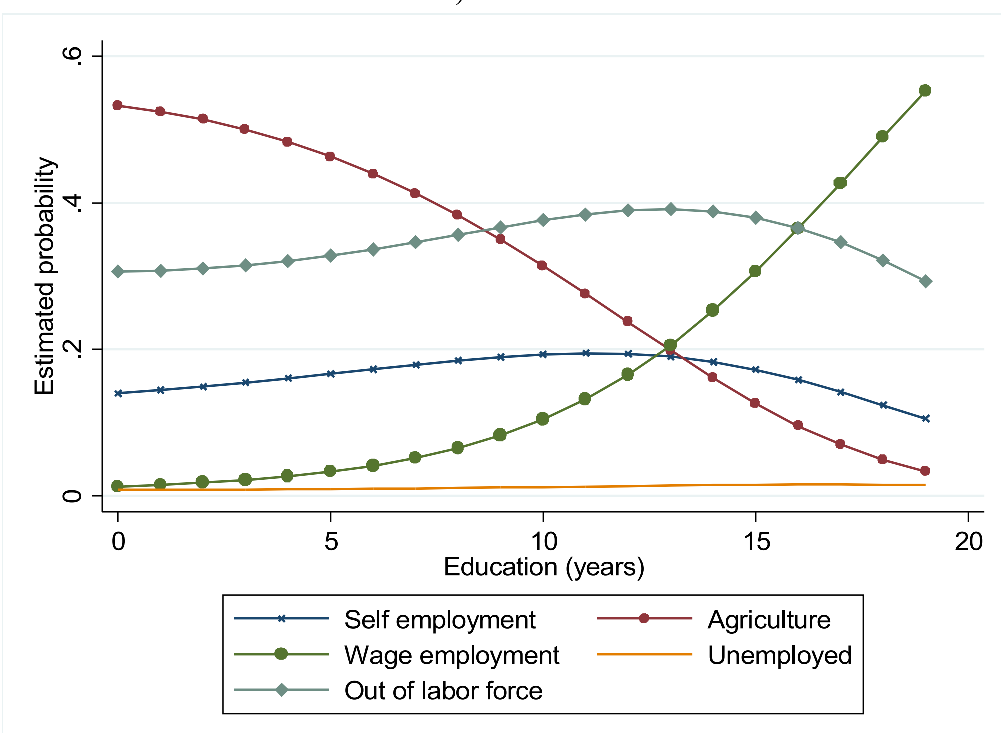
Note: The same as in Table 11. The omitted occupation is laborers.

Figure 1
Estimated probability of occupation and education

i) Men



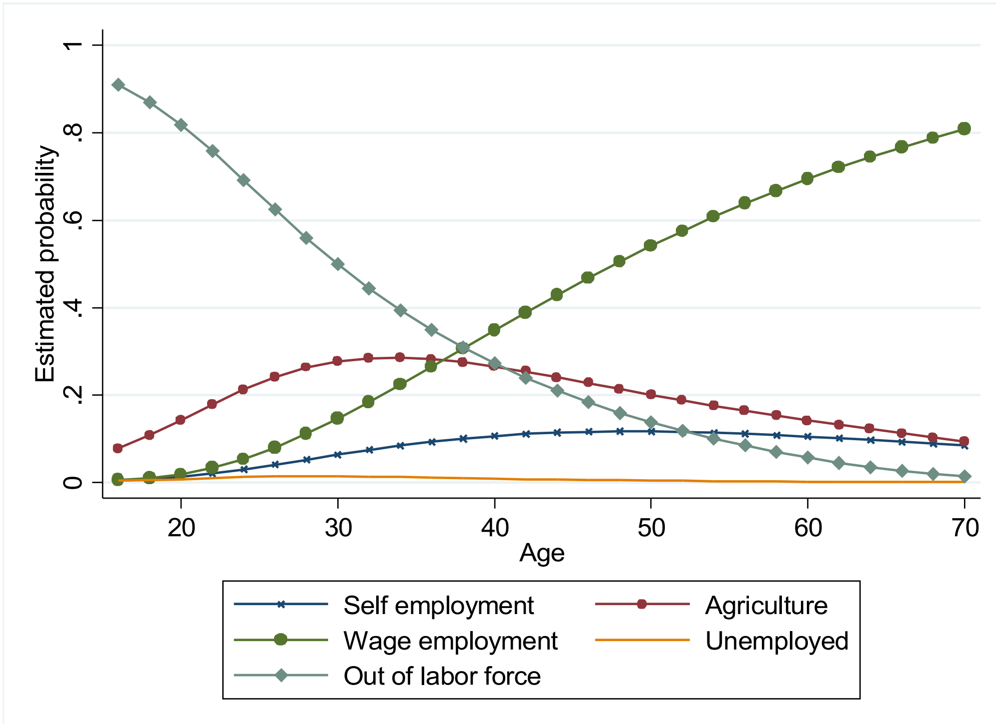
ii) Women



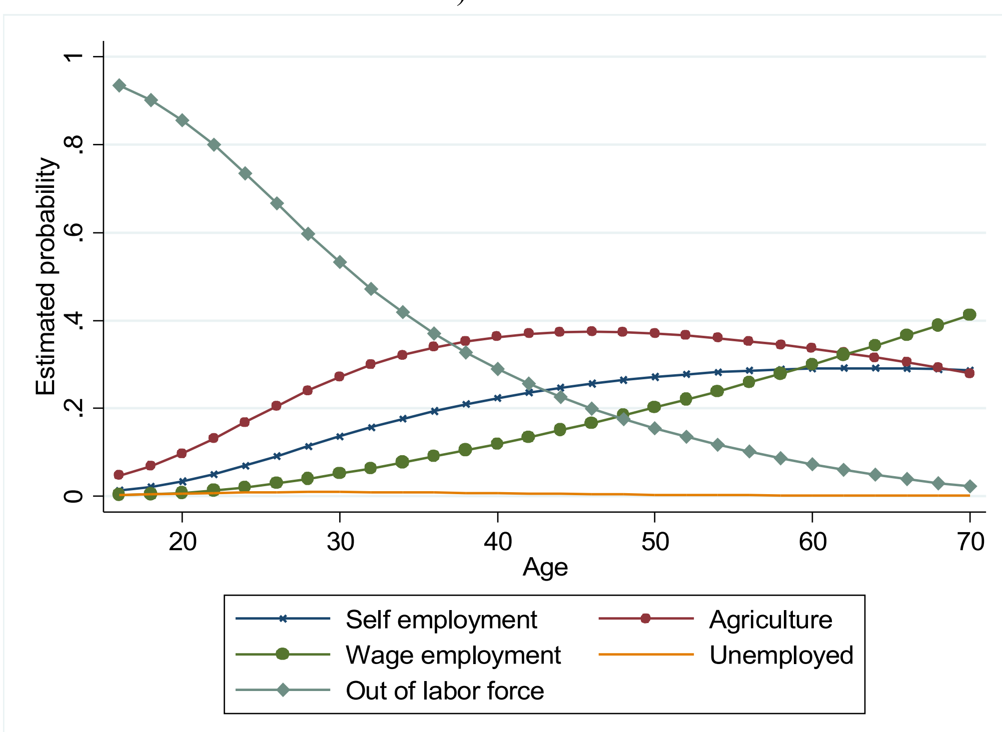
Note: These predictions are based on the multinomial logits reported in Appendix 1.

Figure 2
Estimated probability of occupation and age

i) Men

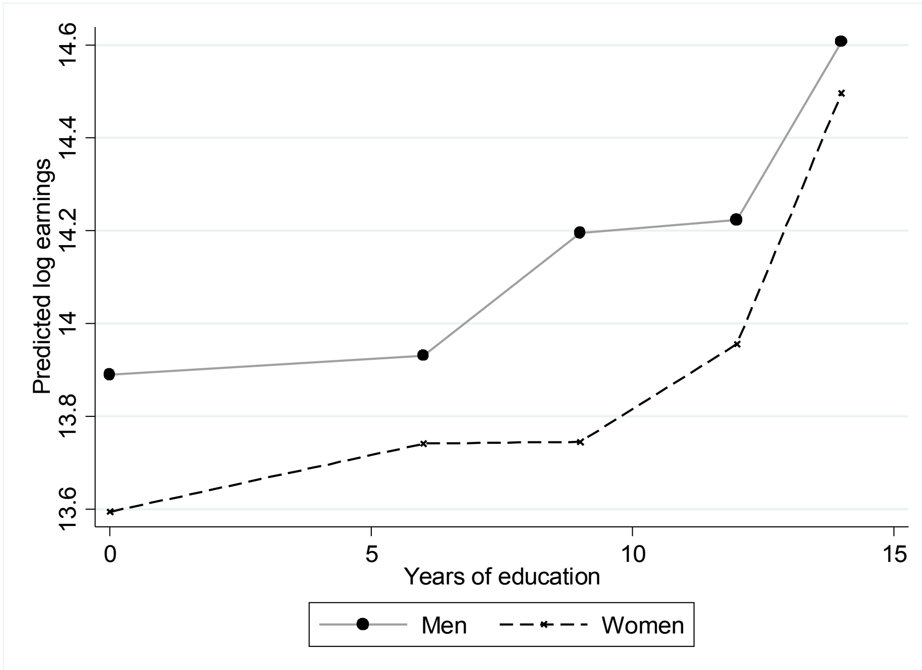


ii) Women



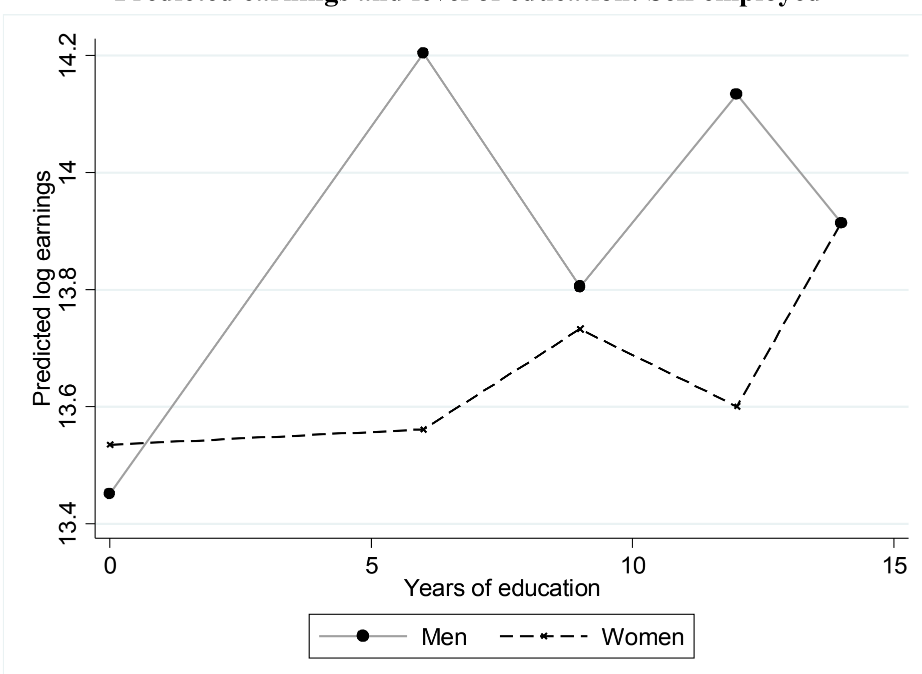
Note: These predictions are based on the multinomial logits reported in Appendix 1.

Figure 3
Predicted earnings and level of education: Wage employed



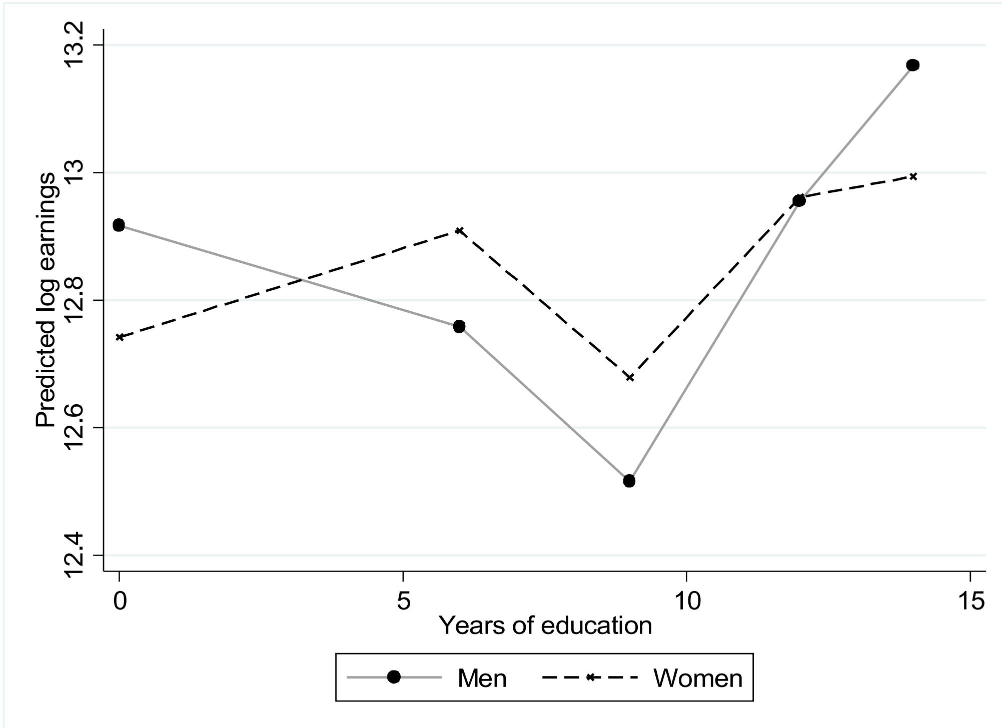
Note: These predictions are based on the results reported in Table A7.

Figure 4
Predicted earnings and level of education: Self employed



Note: These predictions are based on the results reported in Table A7.

Figure 5
Predicted earnings and level of education: Agriculture



Note: These predictions are based on the results reported in Table 10.

Appendix 1

Table A1
Multinomial logit estimates for Men (Omitted category: Wage employment)

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	0.075	0.019	0.160	0.032
	(1.87)+	(0.67)	(1.79)+	(1.03)
Education squared	-0.011	-0.013	-0.012	-0.008
	(4.24)**	(6.74)**	(2.45)*	(4.08)**
Age	-0.048	-0.178	-0.161	-0.420
	(1.29)	(6.98)**	(2.47)*	(15.61)**
Age squared	0.000	0.002	0.002	0.005
	(0.92)	(7.20)**	(1.88)+	(15.23)**
# of children in hh	0.048	0.215	-0.056	0.212
under 10 years of age	(0.88)	(5.63)**	(0.47)	(4.94)**
# of elderly in hh over	0.456	0.714	0.377	0.683
70 years of age	(1.37)	(2.90)**	(0.61)	(2.62)**
Married	-0.091	-0.535	-1.059	-0.856
	(0.55)	(4.53)**	(3.21)**	(6.30)**
Observations	4438	4438	4438	4438

Absolute value of z-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions.

Table A2
Multinomial logit estimates for Women (Omitted category: Wage employment)

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	-0.133	-0.182	-0.155	-0.166
	(2.96)**	(3.87)**	(1.71)+	(3.85)**
Education squared	-0.005	-0.010	-0.001	-0.003
	(1.64)	(2.97)**	(0.10)	(0.97)
Age	-0.054	-0.107	-0.147	-0.361
	(1.33)	(2.79)**	(1.80)+	(9.48)**
Age squared	0.000	0.001	0.001	0.004
	(0.79)	(2.64)**	(0.94)	(8.44)**
# of children in hh	0.154	0.279	0.028	0.191
under 10 years of age	(2.39)*	(4.50)**	(0.21)	(3.07)**
# of elderly in hh over	0.637	0.898	0.788	0.765
70 years of age	(1.98)*	(2.89)**	(1.50)	(2.46)*
Married	0.211	0.181	-0.517	-0.018
	(1.34)	(1.18)	(1.42)	(0.12)
Observations	5175	5175	5175	5175

Absolute value of z-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions.

Table A3
Multinomial logit estimates for Men (Omitted category: Wage employment)

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	-0.388	-0.580	0.401	-0.550
maths problems	(1.36)	(3.00)**	(0.70)	(2.53)*
Can read & write	-0.204	-0.837	-0.256	-0.267
	(0.79)	(4.80)**	(0.60)	(1.36)
Age	-0.049	-0.181	-0.156	-0.417
	(1.34)	(7.40)**	(2.42)*	(16.06)**
Age squared	0.000	0.002	0.001	0.005
	(0.99)	(7.73)**	(1.83) +	(15.64)**
# of children in hh	0.075	0.247	-0.034	0.238
under 10 years of age	(1.39)	(6.63)**	(0.29)	(5.63)**
# of elderly in hh over	0.416	0.636	0.359	0.635
70 years of age	(1.27)	(2.68)**	(0.59)	(2.49)*
Married	-0.128	-0.578	-1.090	-0.900
	(0.79)	(5.09)**	(3.31)**	(6.73)**
Observations	4438	4438	4438	4438

Absolute value of z-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions.

Table A4
Multinomial logit estimates for Women (Omitted category: Wage employment)

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	0.004	-0.732	0.412	-0.193
maths problems	(0.01)	(2.68)**	(0.85)	(0.70)
Can read & write	-1.784	-2.122	-2.064	-1.697
	(6.63)**	(8.11)**	(4.71)**	(6.45)**
Age	-0.067	-0.123	-0.152	-0.375
	(1.73) +	(3.36)**	(1.90) +	(10.32)**
Age squared	0.001	0.002	0.001	0.004
	(1.35)	(3.39)**	(1.10)	(9.36)**
# of children in hh	0.176	0.297	0.039	0.214
under 10 years of age	(2.77)**	(4.86)**	(0.29)	(3.47)**
# of elderly in hh over	0.588	0.825	0.720	0.710
70 years of age	(1.84)	(2.68)**	(1.37)	(2.31)*
Married	0.228	0.224	-0.514	0.001
	(1.48)	(1.50)	(1.41)	(0.01)
Observations	5175	5175	5175	5175

Absolute value of z-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province dummy variables are included in all regressions.

Table A5 Earnings and education, selectivity corrected

	1. Wage employment		2. Self-employment		3. Agriculture	
	Men	Women	Men	Women	Men	Women
Education	0.018	0.025	0.036	0.017	0.023	0.024
	(1.27)	(0.72)	(1.74)+	(1.30)	(1.28)	(1.11)
Age	0.057	0.097	0.070	0.139	0.063	0.035
	(2.01)*	(2.43)*	(0.95)	(3.22)**	(2.87)**	(1.11)
Age squared	-0.001	-0.001	-0.001	-0.002	-0.001	-0.000
	(1.58)	(1.95)+	(1.04)	(3.06)**	(2.72)**	(1.17)
Selectivity term	-0.488	-0.334	0.048	0.292	-0.394	-0.065
	(2.61)**	(1.04)	(0.06)	(0.60)	(0.94)	(0.15)
Observations	898	287	338	819	2098	2340

Robust t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Province variables included but not shown. The identifying variables exclude marriage status as that was statistically significant in the earnings function and is therefore not a good identifying exclusion restriction.

Table A6 Earnings and education with quadratic term, selectivity corrected

	Wage employment		Self-employment		Agriculture	
	Men	Women	Men	Women	Men	Women
Education	0.008	0.006	0.205	0.002	-0.037	0.012
	(0.40)	(0.18)	(2.75)**	(0.04)	(1.25)	(0.34)
Education squared	0.001	0.003	-0.011	0.001	0.007	0.001
	(1.07)	(1.89)	(2.36)*	(0.39)	(2.59)**	(0.38)
Selectivity term	-0.300	0.002	1.466	0.193	-1.081	-0.099
	(1.19)	(0.00)	(1.53)	(0.36)	(2.17)*	(0.20)

Note : As in Table A5. Number of observations also as in Table A5. Age and its square included but not shown.

Table A7 Earnings and education with education levels, selectivity corrected

	Wage employment		Self-employment		Agriculture	
	Men	Women	Men	Women	Men	Women
Primary	-0.038	-0.052	0.950	0.051	-0.063	0.149
	(0.21)	(0.23)	(2.62)**	(0.30)	(0.41)	(1.01)
Middle	0.186	-0.150	0.466	0.224	-0.168	-0.100
	(0.67)	(0.32)	(0.75)	(0.64)	(0.58)	(0.29)
Secondary	0.092	-0.033	0.897	0.099	0.298	0.167
	(0.69)	(0.13)	(2.99)**	(0.62)	(1.66)	(0.72)
Tertiary	0.173	0.208	0.538	0.403	0.752	0.149
	(0.96)	(0.54)	(1.58)	(1.75)	(2.32)*	(0.29)
Selectivity term	-0.573	-0.426	1.209	0.195	-0.819	0.127
	(3.75)**	(1.95)	(1.39)	(0.37)	(1.88)	(0.28)

Note: As in Table A5. Number of observations also as in Table A5. Age and its square included but not shown. The education levels are defined as: primary = 1-6 years of education; middle school = 7-9 yrs; secondary = 10-12 yrs; tertiary = 13+ years.

Table A8
Earnings, literacy and numeracy: Controlling for sample selection

	1. Wage employed		2. Self employed		3. Agriculture	
	Men	Women	Men	Women	Men	Women
Can solve simple	0.031	-0.196	0.251	0.197	-0.221	0.185
maths problem	(0.20)	(1.06)	(0.78)	(1.20)	(1.35)	(1.30)
Can read & write	0.110	-0.001	0.445	-0.031	0.221	-0.039
	(0.80)	(0.01)	(1.56)	(0.20)	(1.38)	(0.25)
Selection term	-0.646	-0.647	0.387	0.235	0.028	0.144
	(6.88)**	(6.03)**	(0.51)	(0.49)	(0.10)	(0.50)
# Individuals	898	287	338	819	2098	2340

Note: Robust t-statistics in parentheses. * significant at 5% level; ** significant at 1% level. Province dummy variables and age are included. The identifying variables exclude marriage status as that was statistically significant in the earnings function and is therefore not a good identifying exclusion restriction.

Table A9
Gender difference in years of education and in cognitive skills

		Men	Women	Absolute difference	Percentage difference
Years of education					
Ghana	Self-employment	8.51	5.66	2.85	50.4
	Agriculture	5.34	2.4	2.94	122.5
	Wage employment	10.54	10.05	0.49	4.9
	Average % gender gap	7.15	4.16	2.99	71.9
Pakistan	Self-employment	5.56	1.92	3.64	189.6
	Agriculture	2.99	0.46	2.53	550.0
	Wage employment	5.95	4.64	1.31	28.2
	Average % gender gap	5.17	2.11	3.06	145.3
Reading and writing skills					
Ghana	Self-employment	77.8	45.8	32.0	69.9
	Agriculture	50.3	17.6	32.7	185.8
	Wage employment	84.1	81.5	2.6	3.2
	Average % gender gap				90.9
Pakistan	Self-employment	66.0	24.0	42.0	175.0
	Agriculture	43.0	10.0	33.0	330.0
	Wage employment	65.0	43.0	22.0	51.2
	Average % gender gap	59.0	26.0	33.0	126.9

.