EDUCATION
Working Paper Series
Number - 12

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

by
Geeta Kingdon and
Måns Söderbom

May 2008

THE WORLD BANK

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

by

Geeta Kingdon and Måns Söderbom

The Education Working Paper Series is produced by the Education Unit at the World Bank (HDNED). It provides an avenue for World Bank staff and consultants to publish and disseminate preliminary education findings to encourage discussion and exchange ideas within the World Bank and among the broader development community. Papers in this series are not formal World Bank publications. The findings, interpretations, and conclusions expressed in these papers are entirely those of the authors and should not be attributed in any manner to the World Bank, its affiliated organizations or to the members of its board of executive directors or the countries they represent. Copies of this publication may be obtained in hard copy through the Education Advisory Service (eservice@worldbank.org), and electronically through the World Bank Education website (www.worldbank.org/education).

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

Måns Söderbom, Department of Economics, University of Gothenburg, Sweden.


May 2008
Washington, D.C. -- USA

## Table of Contents

**Acknowledgments**

The authors are indebted to Tazeen Fasih and Alonso Sánchez for useful comments on an earlier draft of this paper, and to Alonso Sánchez, who provided excellent research assistance and wrote the data appendix.

## Introduction

The motivation and research questions for this case study remain the same as those used in a similar case study for Pakistan:[1] (i) to examine the labor market returns to education among 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 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 the returns to education not only in wage employment, but also in the two other major occupations—agriculture (which employed 64 percent of the labor 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 that it showcases how the availability of data over time enhances understanding of the poverty-reducing potential of education. Such data permits the decomposition of any increase in incomes due to changes in the average amount of education, as well as to underlying technical progress, over a given time period.

The current work adds value to Teal (2001) in five ways. First, it examines the role of education in facilitating entry into lucrative occupations by means of multinomial logit models of occupational attainment. This analysis is important because education plays a role in labor market success not only directly, by increasing earnings in any given occupation, but also indirectly, by promoting entry into well-paying occupations. Second, the paper examines the role of cognitive skills in labor market success, both in terms of occupational outcome and earnings. Third, it estimates returns to education along the earnings distribution by means of quantile regression analysis, which investigates whether the marginal return to education is greater at lower levels of earnings, that is, whether education ameliorates economic inequality or exacerbates it. Fourthly, the paper estimates returns to education by age group in order to examine whether the labor market rewards education differentially for younger and older workers.

[1] Geeta Kingdon, 2007, "Education, Skills, and Labor Market Outcomes: Evidence from Pakistan," background paper for Tazeen Fasih, 2008, "Education-Market Linkages," Human Development Network Education Department, World Bank, Washington, DC.

## 1 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 improving access to employment or, conditional on employment, promoting entry into higher-paying occupations or industries. This paper explores both the total effect of education on earnings and 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

$$\ln w_i = \boldsymbol{\alpha}_{\mathbf{ag}} \mathbf{x_i} + f_{ag}(s_i) + \upsilon_i \quad (1)$$

where $w_i$ is the real earnings of individual *i*, $\mathbf{x_i}$ is a vector of worker characteristics that exclude education, $\boldsymbol{\alpha}_{\mathbf{ag}}$ is a parameter vector, $s_i$ is the years of education, $f_{ag}(\cdot)$ is the earnings-education profile, $\upsilon_i$ is a residual, and *a* and *g* denote age group and gender, respectively.

The primary objective of this paper is to estimate the total returns to education and the variables included in $\mathbf{x_i}$ are selected accordingly. In particular, variables that are determined by education are not conditioned on in the earnings regressions, as this would change the interpretation of the effects of schooling. For example, it is likely that an important effect of education is to enable individuals to get high-wage jobs (e.g., managerial positions), enter certain high-wage sectors or firms, or generate job security and thus work experience. Consequently, occupation, firm-level variables, work experience, or other variables sometimes seen on the right-hand side in earnings regressions are not used as control variables. Instead, the analysis here is restricted to a small set of control variables, with age and gender most emphasized. With respect to the effects of these variables on earnings, a fair amount of flexibility is allowed and all regressions are estimated separately both for men and women and for relatively young individuals (aged less than 30 years) and relatively old ones (aged more than 30 years). Within each gender-age group, age is included as an additional control variable. Controls for province fixed effects are also included.

The estimation of the earnings-education profile $f_{ag}(\cdot)$ is key for the purposes of this paper. It focuses on two specifications: a standard linear model and a model with dummy variables for the 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 analyzing how returns to education differ across different levels of education. In addition, a model is considered in which 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 subsectors, namely, wage employment, self-employment, and agriculture. Among the wage employed, individual data on earnings as well as on the explanatory variables is available. For individuals that are either self-employed or work in the agricultural sector, no earnings data is available at the individual level. Instead, data exists on earnings at the household level, which distinguishes between earnings for self-employed and agricultural workers. In order to identify the parameters in (1), the explanatory variables need to be aggregated so that they are defined at the same level of aggregation as the dependent variable. Fortunately, this is a straightforward task. All that is required is to "collapse" the data (i.e., calculate mean values) on the explanatory variables within the household and labor market subsector (obviously this is not done for the wage employed, given that data exists for individual-level earnings).[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 potential sources of bias in the Ordinary Least Squares 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 subsamples of workers, each of which may not be a random draw from the population, which 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

[2] To give a concrete example, suppose a household has two agricultural workers and three self-employed individuals. Data exists only for the household on total earnings derived from agriculture and total earnings from self-employment, which means it is not possible to estimate the earnings equation at the individual level. Earnings per person in agriculture and self-employment are thus calculated and matched with sector-household specific averages of the explanatory variables.

education influences people's decision to participate in wage, self-, or agricultural employment—it is also needed for 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.

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

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

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

$$E\left(\ln w_{ij} \middle| \mathbf{x}_{\mathbf{i}j}, s_{ij}, m_{ij} = 1\right) = \boldsymbol{\alpha}_{\mathbf{a}gj}\mathbf{x}_{\mathbf{i}j} + f_{agj}\left(s_{ij}\right) + E\left(\upsilon_{ij} \middle| m_{ij} = 1\right) \qquad (2)$$

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 while ignoring sample selection will lead to biased estimates. For example, if more highly motivated or more ambitious people systematically select into particular occupations, for example, into waged work, then people in the waged subsample would, on average, be more motivated and ambitious than those in the rest of the population. Thus, $E\left(\upsilon_{ij} \middle| m_{ij} = 1\right)$ is not zero in this subsample, as the waged workers' subsample 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 $\lambda_{ji}$ as an additional explanatory variable in the wage equation, so that

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

where $z_{ij}$ is a set of variables explaining selection into occupation and $\gamma$ 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 ($\lambda$) is computed.[3] The coefficients on the lambda terms $\lambda_j$ is a measure of the bias due to nonrandom 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, the analysis in this paper considers five broad labor market states: wage employment, self-employment, agricultural employment, unemployment, and being out of the labor force. Occupational attainment is accordingly modeled using a multinomial logit model.

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 with both included independent variables and 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 of education on earnings.

The analysis attempts to address the problem of endogeneity by estimating a family fixed effects regression of earnings. To the extent that unobserved traits are shared within a family, their effect is 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 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***

The empirical strategy of the paper is as follows: first, the earnings functions for each occupation is estimated using the simple Ordinary Least Squares (OLS) model as the baseline. Then, enquiry is made as to whether significant sample selectivity bias exists due to estimating the earnings functions separately for occupation groups, since each of these

---

[3] The inverse Mill's ratio is defined as $\lambda_{ji} = \dfrac{\phi(H_{ij})}{\Phi(H_{ij})}$, 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.

groups may not be a random draw from the population. Finally, an attempt to address the problem of endogeneity is made by using a family fixed effects model.[4]

The paper also estimates earnings functions by the quantile regression (QR) method. With this method, the conditional percentiles of the earnings distribution – rather than the conditional mean - are modeled as a function of the explanatory variables. Such analysis may be very informative if schooling affects the conditional distribution of the dependent variable differently at different points in the wage distribution, because quantile regressions 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 percentage increase in earnings. Using quantile regressions, the paper will 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 those of lower percentiles (on the grounds that such observations have atypically high wages, given their characteristics), quantile regressions will be informative of the effect of education on earnings across individuals with varying ability.[5]

[4] Insufficient data is available to implement a credible instrumental variables approach, for example, there is no data on the supply of education at a young age (Card 1999). In fact, the closest available data to "instruments" (variables that affect years of schooling acquired, but do not affect earnings other than through their effect on years of education) is information on parental education, but this type of data is available only for the subsample of individuals cohabiting 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 (unobserved ability is probably inherited), it was decided not to instrument education using this variable.

[5] If it is assumed that education is exogenous, then the QR approach tells us the return to education for people with different levels of ability, but it cannot be assumed *a priori* that education is exogenous. Thus, it cannot be said 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 because they do not handle the problems of endogeneity bias.

## 2 Data and Descriptive Statistics

The Ghana survey data used in this study correspond to round four of the Ghana Living Standards Survey of 1998–99 (GLSS4).[6] The GLSS4, which was carried out over a one-year period, followed 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 was 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, certain modules concentrate on household enterprises and agricultural activities, including associated expenses and revenues.

The earnings variables for self-employed and agricultural workers are derived from the specialized modules on household enterprises and agricultural activities, respectively. A simple, yet comprehensive computation of recurring (nondurable) expenses and revenues—including produced or harvested goods consumed by the household—attributed to enterprise or agricultural endeavors 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 A1.1[7] shows summary statistics for selected variables, the full sample, and the same five occupation categories used previously in the Pakistan case study. As in the latter study, the present sample consists of persons aged between 16 and 70 years who are not currently enrolled in school. The state of unemployment is defined as individuals who seek employment and are available for it, while the state of "out of the labor force" is defined as individuals who do not seek employment, such as housewives and retired people. The labor force participation rate is about 72 percent in Ghana (compared with 51 percent in Pakistan) and the unemployment rate, about 2 percent (compared with 6 percent in Pakistan).

Average earnings in the full sample are Ghanaian cedis 1,350,471, which is equal to about US$575 (compared with Pakistan's US$600 for the same year). There is a huge difference in average earnings between agriculturally employed persons and those in either

[6] The authors are indebted to Alonso Sanchez for his substantial contributions to this section.

[7] For ease of reading, tables and figures have been removed to appendixes. Tables associated with the core analysis of the paper are found in appendix 1; figures associated with this analysis, in appendix 2; and selectivity corrected tables, in appendix 3. All tables and figures are identified first by the number of their respective appendix: Table A1.1 (appendix 1), Figure A2.1 (appendix 2), Table A3.1 (appendix 3).

wage or self-employment. Self-employed and wage-employed persons earn on average about 150 percent and 163 percent more, respectively, than do individuals working in agriculture. This interoccupational earnings difference is more than double that of Pakistan, where the corresponding figure is 70 percent. However, the mean can be a misleading measure of central tendency, since the earnings distribution is very skewed. Thus, table A1.1 also shows median earnings and log of earnings, with figure 1.A below the able showing the distribution of log earnings. The latter show a clear and pronounced hierarchy, with earnings in agriculture the lowest and in wage employment, the highest, with a huge fourfold difference. Median earnings in self-employment are about half those in wage employment. Figure A1.A also shows that wages are more narrowly distributed than earnings from self-employment and agricultural employment. The average number of years of education among workers 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 the respective levels in Pakistan. The pattern for literacy and numeracy skills is similar to that for education: while 70–85 percent of wage employed and unemployed persons are literate and numerate, the corresponding figures for self-employed persons are considerably lower (50–65 percent) and for agricultural workers, even lower (35–40 percent).

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

## 3 Education and Occupational Attainment

As in the case study for Pakistan, analysis in this paper distinguishes between the effects of education and skills on (i) occupational outcome and (ii) earnings, conditional on occupational outcome. This section looks at the first issue and the following section looks at the second. The five "occupation" categories examined are: self-employment, agriculture, wage employment, unemployment, and individuals out of the labor force (OLF). While from a policy point of view, the link between education and labor market outcomes among the relatively young deserves attention, the Ghana sample is not large enough to permit separate analysis by age group and gender. Labor market outcomes for all persons aged 16–70 years are therefore analyzed, mainly by gender, with the main tables by age group also presented. With respect to the latter, “young” is defined as persons aged 16–30 years and “aged,” persons 31– 70 years.

A simple multinomial logit model is used to examine the role of education, skills, and family background in determining occupational choices and/or outcomes. The model is set up, and uses the same explanatory variables, as for the Pakistan study. Whenever education is included as an explanatory variable, literacy and numeracy variables are excluded, and vice versa, since these dimensions of skills are highly correlated and the analysis here has no interest in documenting the effects of education conditional on literacy and numeracy skills or the other way around.

Table A1.2 shows the 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 both the number of children and number of elderly people significantly reduces men’s likelihood of being in wage employment (which is highly paid). Somewhat surprisingly, the same finding holds less strongly for women. This negative association could be because wage employment is a less flexible occupation (e.g., in terms of working hours), 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 A2.1 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 direct comparisons are not possible between Pakistan and Ghana (separate graphs were generated for young and old workers in Pakistan, but not in Ghana), it is clear that the role of education in occupational attainment in Ghana is extremely different to that in Pakistan. First, the relationship between education and occupational choice is far more similar for men and women in Ghana than it is in Pakistan, where it varies dramatically by gender, which likely reflects a difference in the perceived gender role of women in a predominantly traditional Asian society. Second, even for males alone, the relationship between education and occupation is very different between Ghana and Pakistan, suggesting that very different forces are operating in the labor markets of the two countries.

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, although for women, education beyond the secondary level reduces chances of being OLF.

Table A1.3 presents the marginal effects of basic literacy and numeracy on occupational attainment. As noted earlier, table A1.1 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 A1.3 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 learned on the job). Either way, there is no such positive relationship between numeracy and self-employment for men. Being numerate also strongly reduces the chances of ending up in agriculture for both men and women, but the size of this marginal effect is significantly smaller for men than for women. Gender differences in the relationship between skills and occupational outcomes could be explained by differential earnings rewards of numeracy for men and women, something that is explored in the next section.

## 4 Education and Earnings

***Basic relationship***

Table A1.4a presents basic OLS estimates of the Mincerian earnings function in Ghana by occupation and gender. Table A1.4b presents OLS estimates of the same equation by occupation and age group. Unlike the Pakistan case study, in which returns to education were very precisely determined for all 12 subgroups (3 occupations x 2 genders x 2 age groups) except women in agriculture, returns in Ghana are precisely determined only for the sample of waged workers, which include men and women, as well as young and old workers (i.e., four subgroups). The term "returns to education" is used here as is common in the literature, however, strictly speaking, the coefficient on the Mincerian earnings function is simply the gross earnings premium from an extra year of education and not the "return" to education, since it does not take the cost of education into account.

Table A1.4a shows that the average marginal wage return to education in Ghana is about 5 percent for both men and women. This finding contrasts strongly with the finding for Pakistan, where wage returns are three to five times higher for women (15–17 percent) than men (3–6 percent). The greater premium on education for women in Pakistan is likely to reflect, at least in part, the greater scarcity of educated women in that country than in Ghana, combined with the existence of predominantly "female" jobs which require educated women, such as nursing and teaching. Table A3.9 shows that among wage-employed individuals in Ghana, the average education of men is only about 5 percent higher than that of women, but in Pakistan, it is 28 percent higher. However, table A1.8 shows that the gender gap in returns to education is pro-female at the secondary and tertiary levels of education in Ghana.

While the average slope of the education-earnings profile is no steeper for women than for men in Ghana, this is not true for the intercept of the earnings equation. Table A1.4b, 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 percent in wage employment to 35 percent in self-employment and 7 percent 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, that is, their earnings do not catch up with those of men at higher levels of education. This finding contrasts with the case of Pakistan, where women have a lower intercept in the earnings function, but enjoy higher returns to education, meaning that the gender earnings gap is significantly lower at higher levels of education. Graphs of predicted earnings show this finding more clearly in figures A2.3 and A2.5.

Table A1.4b also shows that wage returns to education for the young are statistically equal to those for the old. Again, this finding contrasts with the finding in Pakistan, where returns for the young were 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). It is possible that the lack of this "filtering down" phenomenon in Ghana is due to a less rapid expansion in the supply of educated persons in Ghana than in Pakistan over the past 40 years, although this hypothesis cannot be tested due to the lack of appropriate data.

Returns to education in self-employment and agriculture in Ghana 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 percent in self-employment for men and 2 percent each for women, old workers in agriculture, and old workers in self-employment. These findings are similar to those of Teal (2001), although they are not strictly comparable since Teal's paper pools data from four household surveys rather than using data only from the 1998–99 survey. Teal finds that returns to education in wage employment are about 6 percent; in self-employment, 2.5 percent; and in agriculture, 1 percent. 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 for Argentina (Gallacher 1999, 2001), where returns to education in agriculture for farms of average size was found to be equal to returns to education in wage employment.[8]

The low returns to education in self-employment in Ghana are unfortunate because non-agricultural self-employment is the fastest-growing occupation in the country (Teal 2001). This finding 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

[8] It could be argued that land and assets matter to profits in both self-employment and agricultural employment and that they are likely positively correlated with education, so that the return to education could, in principle, be upwardly biased. In practice, however, returns to education are very low, even excluding land and assets. Including these variables is thus not an issue in these data.

very high proportion of the workforce (64 percent) in Ghana. Neither is the gender pattern of returns not favorable to women; in addition to having lower earnings than men in all occupations at zero education (the intercept of the earnings function being much higher for men than women), women do not enjoy a substantial returns-to-education premium over men in any occupation. That is, the slope of the earnings function is also, on average, not higher for women, so that higher levels of education do not lead to a statistically significant reduction in the gender earnings gap.

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 the country, since these two occupations together constitute 82.5 percent of the employed workforce. This somewhat pessimistic conclusion is moderated by the finding seen in section 3, that education plays a major part in sorting people into highly paid occupations.

***Extensions on the education-earnings relationship***

*Correcting returns estimates for endogeneity bias*

As discussed earlier in this paper, OLS estimates of returns to education potentially suffer from sample selectivity bias and endogeneity bias. This paper attempted to address the former by employing the Heckman procedure, explained in the first section. The multinomial logit equations in appendix 3 were used to calculate the selectivity terms for each occupation and worker group (see table A3.5). 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, OLS is preferred to the selectivity corrected equations in this paper, unlike the Pakistan case study, in which the selectivity term was significant in many of the earnings functions for different worker groups.

The endogeneity of schooling is addressed by estimating a household fixed effects earnings function for waged work.[9] This value cannot be estimated for self-employment and

[9] The analysis here also seeks to address the problem of the endogeneity of education by using a two-stage least squares (2SLS) technique to estimate the earnings function. Desirable instruments are not available,

agricultural employment because there is no within-household variation in these cases. The results in table A1.5 show that correction for endogeneity bias does not change wage returns to education significantly. Fixed effects returns to education for men are 6.6 percent and for women 4.1 percent; 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 equation lends confidence to this paper's working assumption that OLS results are closer to the true causal estimates of the effects of education on wages and justifies the presentation of OLS results for waged men over the selectivity corrected results of table A3.5.

*Shape of the education-earnings relationship*

The analysis so far has imposed a linear relationship between "years of education" and earnings in all occupations, but it is not inevitable that this relationship is linear. Table 1.6a relaxes the implicit presumption of linearity by introducing quadratic terms into education. The selectivity corrected counterpart of table A1.6a is Table A3.6.[10] The first table shows that in wage employment, the education-earnings relationship is strongly convex for both men and women. Thus, the Ghanaian labor 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 A1.6a also shows that this relationship is weakly convex for men in agricultural employment; it is concave only for one group: self-employed men. Table A1.6b estimates

---

for example, a rule that would change years of education in an exogenous way, or even a variable such as distance to nearest school, when an individual was of school-going age (see the last section of the conclusion of this paper). Spouse's education was used as an instrumental variable for waged workers' schooling, as this data was available much more commonly than father's or mother's education. The effect of applying this instrument was to raise male waged workers' return to education to 9.4 percent (t=5.6) from the OLS estimate of 5 percent in Table A1.4a, and to reduce women's return to 4.0 percent (t=1.7) from the OLS estimate of 5.9 percent in the same table. Women's return from the 2SLS estimates is almost identical to that from the household fixed effects equation. The fact that the 2SLS 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 the household fixed effects estimation is also higher than the OLS result, although the 2SLS result is appreciably higher. However, the authors do not trust the 2SLS estimates as much as the fixed effects estimates because the validity of the 2SLS results cannot be confirmed. In addition, a spouse's education is *a priori* a questionable instrument for worker education, since the theory of assortative mating suggests that like people marry each other, that is, a spouse's education is likely to be correlated with a worker's unobserved characteristics.

[10] The household fixed effects estimates of the earnings function for waged workers is not estimated with either a quadratic term or with education *level* rather than *years* of education due to the very small subsample of households that have two or more members employed in waged work.

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 nonlinearities of the education-earnings relationship are explored further in table A1.7, which includes a dummy variable for each main education level. (The selectivity correction estimator can be found in table A3.7.) The base education category is "no education." Table A1.7 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, each year of middle education, and so forth, are calculated from table A1.7 and set out in table A1.8. The latter table shows that in wage employment, marginal returns to tertiary education are lower for men (12.8 percent) than women (18 percent), but the 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, 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.

Taken together, tables A1.6, A1.7, and A1.8 suggest that, with the exception of self-employed men, the education-earnings relationship in Ghana is generally not concave. Figures A2.3, A2.4, and A2.5, which show the relationship between education and predicted earnings, confirm this finding. Figure A2.3 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 A2.4) and for both genders in agriculture (figure A2.5), neither of these findings 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 the 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 A1.9 shows earnings functions by occupation with measures of cognitive skills on the left. Years of schooling is not included in the earnings functions because the analysis seeks to estimate the total return to cognitive skills irrespective of whether they were acquired through schooling.[11] (Corresponding selectivity corrected equations are presented in table A3.8.) Results for household fixed effects are not reported because very few households in the sample had two or more wage employed members. Table A1.9a shows these results by gender and table A1.9b, by age.

Table A1.9a shows no significant returns to numeracy skills in any worker group, but reveals 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. Women's earnings premia from literacy are generally statistically no different than those of men. This result contrasts greatly with findings from the Pakistan case study, which in most cases revealed dramatically larger returns to literacy for women than men. This finding may be partly due to a greater scarcity premium for women in Pakistan than in Ghana,[12] although Ghana's gender gaps in cognitive skills are large enough in absolute terms that one would have expected a higher earnings premium to this skill, which is more scarce among women. One plausible explanation for this apparent puzzle could be that 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, it is not necessary to reserve certain types of jobs for persons of particular genders.

When the sample is divided by age group, the estimates of earnings functions in table 1.9b 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 and the old in self-employment. There are no significant returns to literacy or numeracy in agriculture for either age group, suggesting that

---

[11] When years of education are included, 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 finding might be taken to suggest that education does not have an impact on earnings by raising cognitive skills, the authors are reluctant to draw this inference since the cognitive skills variables here are simple 0/1 dummies, rather than a more informative continuous measure.
[12] Table A1.9 shows that while the gender gap in the percentage of persons with literacy skills in all occupations in Ghana is large, it is nevertheless smaller than in Pakistan. Fewer women than men in Ghana have the years of schooling required to develop literacy skills, although it is not known whether women are likely to have attended poorer-quality schools than men, as is the case in Pakistan (Aslam 2007).

Ghanaian agriculture is mainly traditional, that is, cognitive skills that would allow a person to, for example, follow instructions on fertilizer packs, do not raise agricultural earnings.

That literate and numerate young workers can command an earnings premium in certain segments of the labor 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, a reality that highlights the importance of the quality of schooling.

Comparing the returns to education with the returns 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, the former was regressed on the latter, with the finding 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 in Ghana. Table A1.9a 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 percent, for waged men and 0.29/12, or 2.5 percent, for waged women. This finding compares with a rate of return to education of about 5 percent for both men and women. In other words, the apparent return to cognitive skills is lower than the return to education, suggesting that a substantial element of rent is associated with education in Ghana (i.e., education is used partly as a device to signal ability).

*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 differences in the interest rate faced by different individuals (for instance, the interest rate based on wealth and/or assets) (Card 2001). The fact that returns to education can be heterogeneous across individuals has implications for the inequality-reducing role of education. To the authors' knowledge, distribution of the returns to education across the earnings spectrum has not been investigated for Ghana, as is the case for most developing countries (Patrinos, Ridao-Cano, and Sakellariou 2006). This paper examines heterogeneity in returns to education in order to ask whether some workers benefit more from education than others and why, and then examines the inequality implications of the answer.

The Pakistan case study previously alluded to the literature that investigates 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 returns to education increase with quantiles in developed countries (i.e., returns are higher for higher earnings quantiles), whereas the evidence is mixed in middle-income countries; in the few developing countries for which evidence exists, returns decrease with quantiles.[13] If the 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 A1.10 reports the quantile regression results. The top half of the table presents results for men and women. It shows that in wage employment, there is a consistent pattern of returns to education being different at different points of the conditional earnings distribution for both men and women. Returns to education are highest in the lowest quantile of earnings (bottom quartile) and lowest in the highest earnings quantile (top quartile). For both men and women, the difference between the top and bottom quartiles is statistically significant, although the size of the difference is nearly twice as big for women as for men. While a 1.6 point difference (5.8 percent minus 4.2 percent) 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 percent minus 5.3 percent) 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 the worker and occupation groups used for the present study,

---

[13] For Austria, Denmark, Finland, France, Ireland, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom, see Martins and Pereira (2004); for Latin American countries, see Patrinos, Ridao-Cano, and Sakellariou (2006); for South Africa, see Mwabu and Schultz (1996); for the United States, see Buchinsky (1998).

people with lower ability have higher rates of return to education, lending support to the notion of a substitution between ability and education. This finding suggests that among waged men and women, and among self-employed women, education is inequality reducing, since education decreases rather than increases the wage differences between low- and high-ability individuals. However, among self-employed men, education appears to be inequality increasing: the returns to education in the top earnings quartile is nearly double the returns at the median earnings quintile, which is weakly higher than the returns in the bottom quartile. No such patterns are discernible in agriculture.

The bottom half of table A1.10 presents QR results by age group, showing that among old waged workers, returns to education are highest for the bottom earnings quartile and lowest for the top earnings quartile, suggesting that education is inequality reducing among older waged workers, although the size of the difference is not economically large. There is no suggestion of any pattern of 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.

## 5 Conclusions

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

While education raises earnings *indirectly* by helping individuals gain entry into high-paying occupations, it has low *direct* effects on earnings. Results show that education raises earnings only modestly and 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 percent of the employed workforce. While it may seem that the economic incentives for acquiring schooling may be weak in Ghana, two considerations counter this assumption. First, education has large indirect effects by promoting entry into (well-paying) wage employment. Second, the returns to education mostly increase with education level in Ghana, so that there is an economic incentive to reach 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 (as well as for the young and old age groups). The gender gap in education is not large in Ghana and, in any case, the labor market does not appear to be segmented by gender. Does the fact that women reap economic rewards equal to those of men from education mean that girls will have the same economic incentives to acquire schooling as boys in Ghana? Unfortunately, this cannot be concluded from a mere examination of the returns to education for the two genders (i.e., 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, not only the slope). Not only are the returns to education no higher for women than for men (i.e., there is no scarcity premium for women's education), women's earnings overall are much lower than those of men (i.e., the intercept of the earnings function is much lower for women than for men). The gender gap in earnings, moreover, does not narrow significantly at higher levels of education. These findings are evidence of gender differentiation in the labor market, although to establish whether gender discrimination exists, further analysis and better data would be required, including accurate measures of labor market experience and the quality of schooling of men and women. In conclusion, there appears to be a prima facie case for policies that discourage gender-differentiated treatment by employers in the labor market.

In Ghana, as in Pakistan, the shape of the education-earnings relationship is not concave, with diminishing returns to education, as conventional wisdom suggests. Rather, in wage employment for both men and women (and to a lesser extent, in agriculture for men), the relationship is convex, that is, higher returns accrue only at 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 climb out of poverty.

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

Examining relationships between numeracy and literacy, on the one hand, and occupational outcomes and earnings, on the other, literacy and numeracy both strongly promote entry into the lucrative parts of the labor market for both men and women in Ghana. Conditional on occupation, literacy skills have moderately large payoffs for both genders, although these are confined mainly to wage employment. While there is a 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 are never statistically significant. While in some cases this finding may be attributed to the small sample, in other cases, sample size is clearly not the reason, for instance, in agriculture.

***Lessons for future research***

What has this research revealed about how household survey data can be used to analyze labor markets in developing countries? First, it has highlighted the importance of estimating returns to education separately by occupation, rather than estimating them only for wage employment, unlike in most applied labor economics literature for developing countries. Wage employment is typically a small, and increasingly, shrinking part of the labor market in many developing countries. The substantial difference in returns to education in different occupations that is documented by this paper highlights the importance of estimating the returns to education separately by occupation.

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

Third, a comparison of Pakistan and Ghana case studies makes clear that sample size matters. Large household surveys that furnish reasonably large samples of workers permits more disaggregated analysis and more nuanced understanding of the differing role of education for different worker groups. Large samples also allow the researcher to use more demanding econometric techniques, which 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 provide a much tighter upper bound on the true causal effect of education on earnings, since such functions net out many aspects of unobserved traits shared by family members. In the Ghana case study, the coefficients on education could not be identified using the quadratic and levels specifications in a family fixed effects equation because of small sample sizes.

Fourth, 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 has shown how these issues may be addressed. While this paper has used a household fixed effects methodology to address the problem of endogeneity bias (also known as "ability bias"), the instrumental variables method is an alternative technique. This alternative, however, requires "instruments," that is, 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 an individual was of school-going age is one such instrument used in the literature. It would be helpful if the need for such data becomes an important consideration when planning future labor market surveys. Of course, it would be ideal if data existed on an administrative rule change that affected years of education acquired and was exogenous from the point of each individual. For instance, a rise in the school-leaving age in the United Kingdom from 14 to 15 in 1947, and another from 15 to 16 in 1973, permitted Harmon and Walker (1995) to estimate the true causal return to education, uncontaminated by the effect of "ability bias."

## Appendix 1: Tables

**Table A1.1**
**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 |
| 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 U.S. dollars were $575. Read and write skills in own language or English.

**Figure A1.A Earnings distribution for three occupations**

**Table A1.2**
**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 tables A3.1 and A3.2. Robust t-statistics in parentheses. $^{+}$ significant at 10% level; * significant at 5% level; ** significant at 1% level.

**Table A1.3**
**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 tables A3.3 and A3.4. Can read or write I n native language or English = 1; else =0. $^{+}$ significant at 10% level; * significant at 5% level; ** significant at 1% level.

**Table A1.4a**
**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 A1.4b**
**Earnings and years of schooling, by age group**

| **Accessed September 2002.** | **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 A1.5**
**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 A1.6a**
**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 A1.6b**
**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 A1.4b.

**Table A1.7**
**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 A1.8**
**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 A1.9a**
**Earnings, literacy and numeracy, by gender**

| | 1. Wage employed | | 2. Self employed | | 3. Agriculture | |
|---|---|---|---|---|---|---|
| | Men | Women | Men | Women | Men | Women |
| Can solve simple maths problem | 0.165 | 0.112 | 0.231 | 0.179 | -0.217 | 0.209 |
| | (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 A1.9b**
**Earnings, literacy and numeracy, by age**

| | 1. Wage employed | | 2. Self employed | | 3. Agriculture | |
|---|---|---|---|---|---|---|
| | Young | Old | Young | Old | Young | Old |
| Can solve simple maths problem | 0.386 | 0.048 | -0.094 | 0.351 | 0.002 | 0.125 |
| | (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 A1.10**
**Earnings and years of schooling: Quantile regressions**

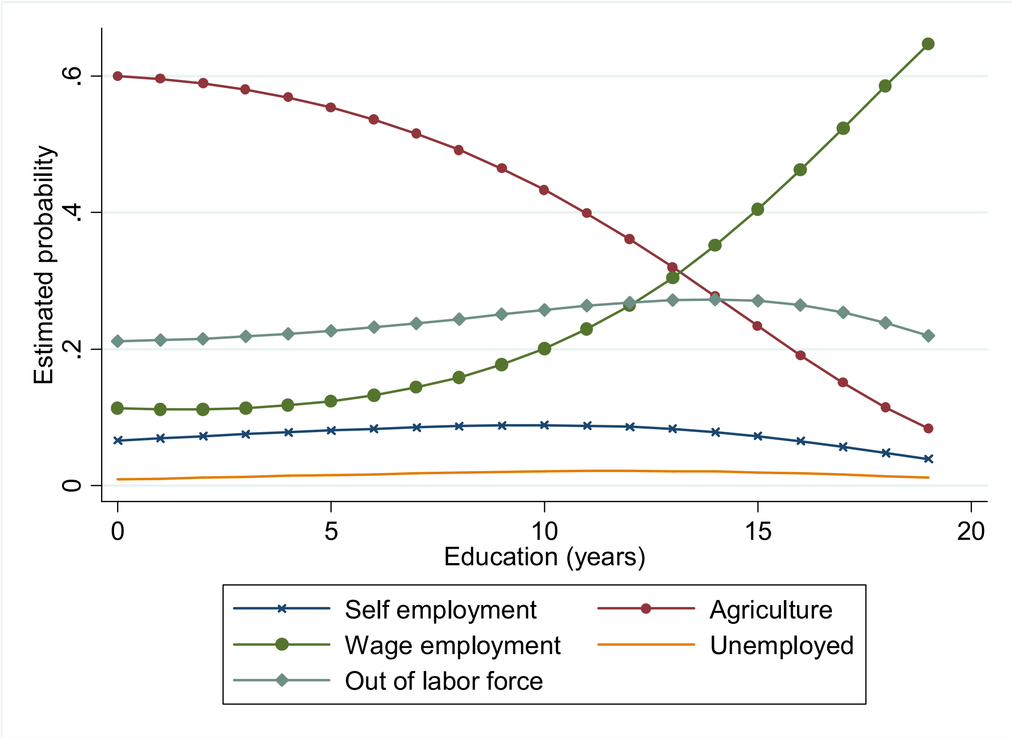
| | 1. Wage employed | 2. Self employed | 3. Agriculture |
|---|---|---|---|
| **Men** | | | |
| Education, | 0.058 | 0.033 | 0.014 |
| (25th percentile of earnings) | (7.34)** | (1.42) | (1.18) |
| Education | 0.049 | 0.042 | 0.013 |
| (50th percentile of earnings) | (13.63)** | (1.84) + | (1.10) |
| 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.

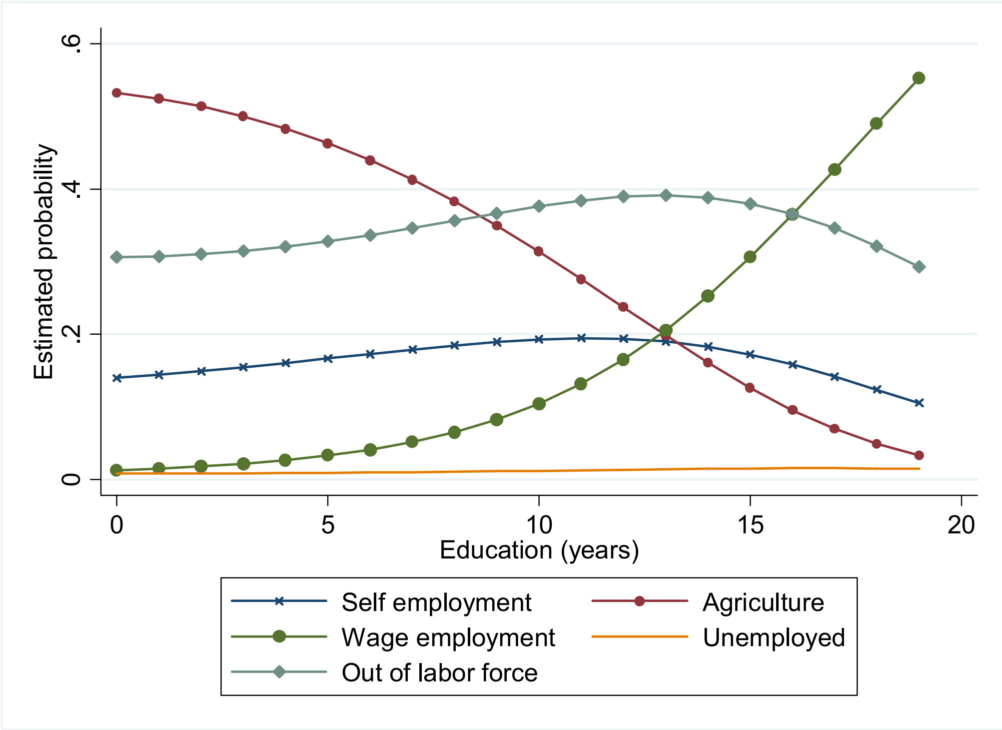
## Appendix 2: Figures

**Figure A2.1 Estimated probability of occupation and education**

(i) Men

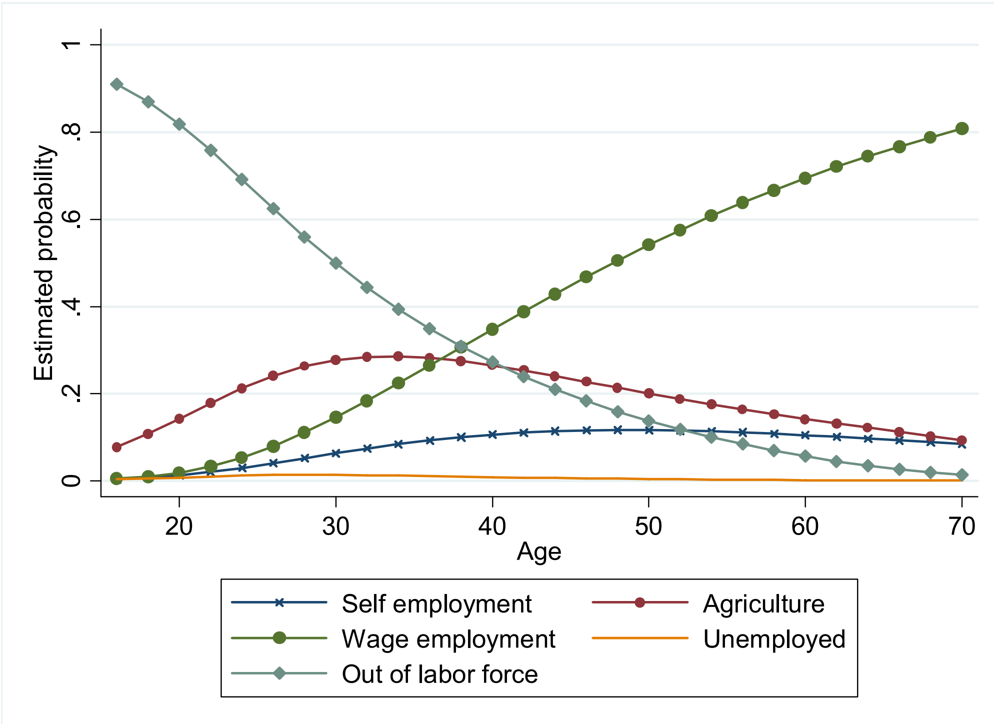



(ii) Women

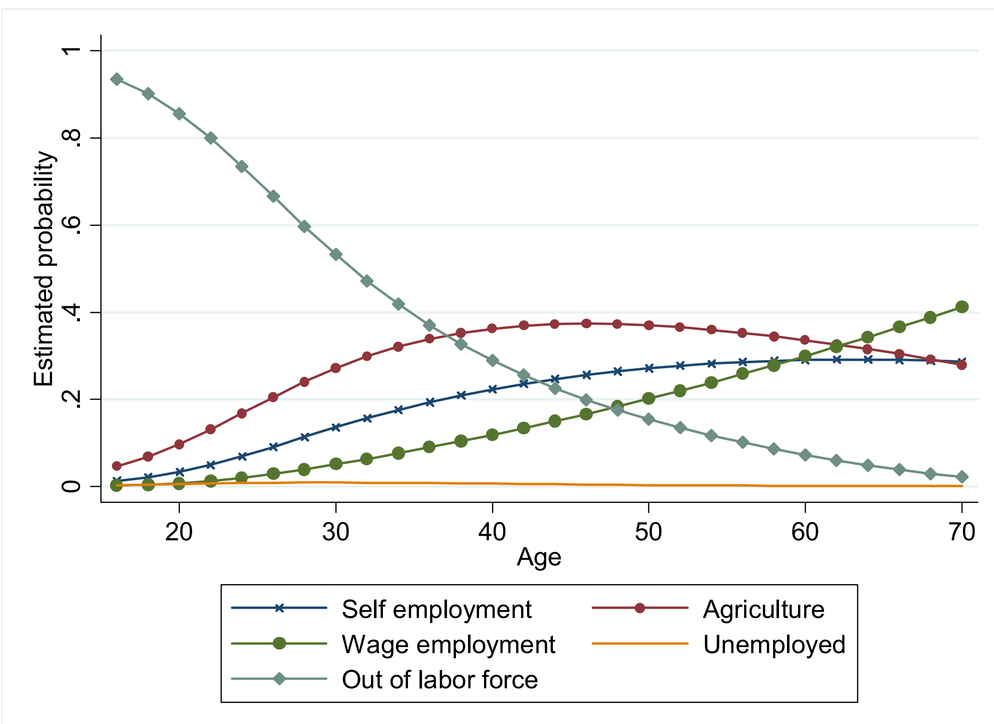



*Note*: These predictions are based on the multinomial logits reported in appendix 3.

**Figure A2.2**
**Estimated probability of occupation and age**
(i) Men

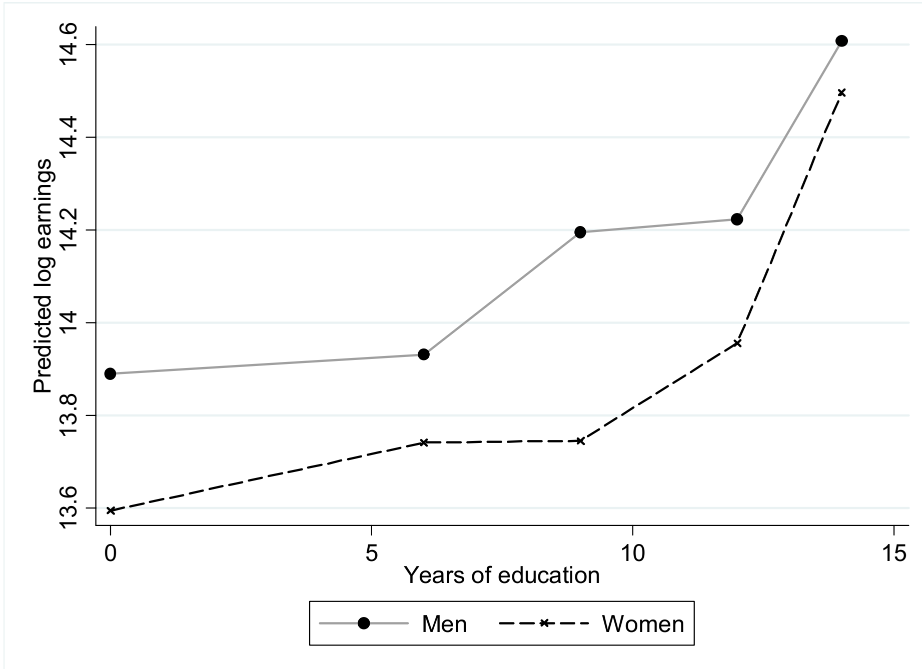



(ii) Women



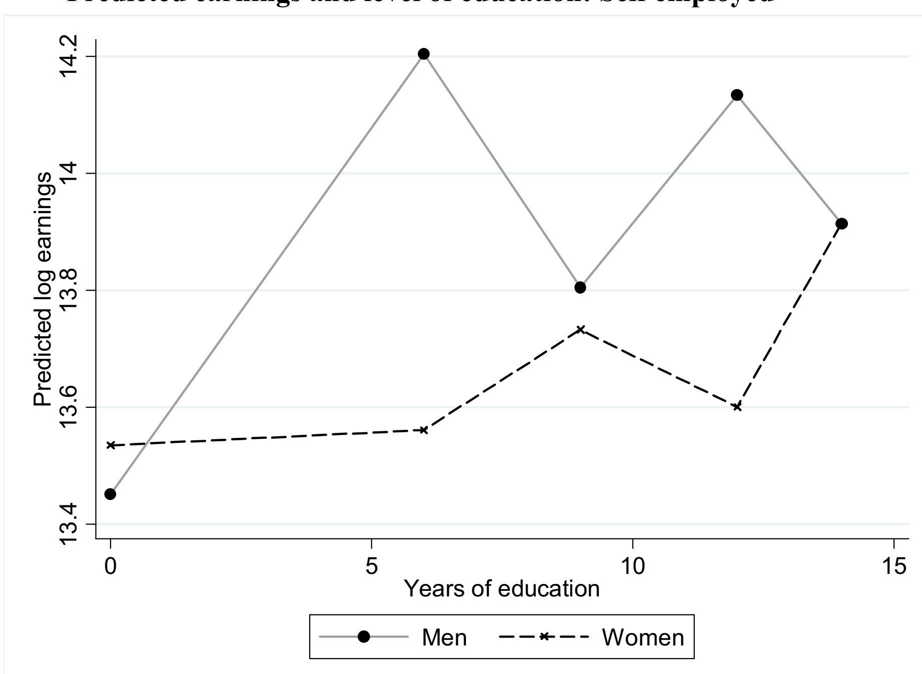

*Note*: These predictions are based on the multinomial logits reported in appendix 3.

**Figure A2.3**
**Predicted earnings and level of education: Wage employed**



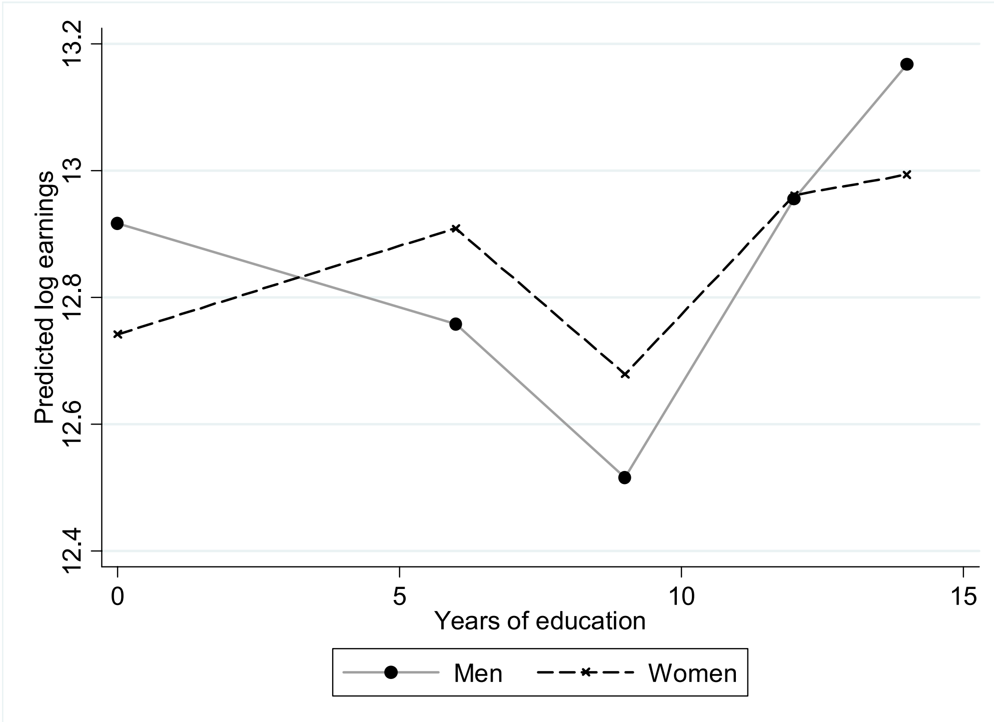

*Note*: These predictions are based on the results reported in table A3.7.

**Figure A2.4**
**Predicted earnings and level of education: Self employed**




*Note*: These predictions are based on the results reported in table A3.7.

**Figure A2.5**
**Predicted earnings and level of education: Agriculture**




*Note*: These predictions are based on the results reported in table A1.10.

## Appendix 3:
## Multinomial Logit Estimates and Selectivity Corrected Tables

**Table A3.1**
**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 A3.2**
**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 A3.3**
**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 |
| math 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 A3.4**
**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 A3.5 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 A3.6 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 A3.5. Number of observations also as in table A3.5. Age and its square included but not shown.

**Table A3.7 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 A3.5. Number of observations also as in table A3.5. 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 A3.8**
**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 maths problem | 0.031 | -0.196 | 0.251 | 0.197 | -0.221 | 0.185 |
| | (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 A3.9**
**Gender difference in years of education and in cognitive skills: Ghana and Pakistan**

| | | **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 |

**References**


Alderman, H., P.F. Orazem, and E.M. Paterno. 2001. "School Quality, School Cost, and the Public/Private School Choices of Low-income Households in Pakistan." *Journal of Human Resources* 36, no. 2: 304–26.

Andrabi, T., J. Das, and A. Khwaja. 2002. "The Rise of Private Schooling in Pakistan: Catering to the Urban Elite or Educating the Rural Poor?" World Bank, Washington, DC, and Harvard University, Cambridge, Massachussetts.

Appleton, Simon, and Arsene Balihuta. 1996. "Education and Agricultural Productivity in Uganda." *Journal of International Development* 8, no.3 (May–June): 415–44.

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* 26, no.1: 7-40.

Arif, G.M., and N. Saqib. 2003. "Production of Cognitive Life Skills in Public, Private, and NGO Schools in Pakistan." *Pakistan Development Review* 42, no. 1: 1–28.

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

Aslam, M. 2007. "Rates of Return to Education by Gender in Pakistan." CSAE Working Paper Series, GPRG-WPS-064. Department of Economics, Oxford University.

Behrman, J., D. Ross, and R. Sabot. 2002. "Improving the Quality Versus Increasing the Quantity of Schooling: Evidence from Rural Pakistan." University of Pennsylvania. Unpublished.

Buchinsky, M. 1994. "Changes in the U.S. Wage Structure 1963–1987: An Application of Quantile Regression." *Econometrica* 62: 405—58.

Gallacher, M. 1999. "Human Capital and Production Efficiency: Argentine Agriculture." CEMA Working Paper, no. 158. Universidad del CEMA, Buenos Aires, Argentina.

———. 2001. "Education as an Input in Agricultural Production: Argentina." CEMA Working Paper, no. 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: 267–90.

Hanushek, Eric A. 2005. "The Economics of School Quality." *German Economic Review* 6, no. 3 (August): 269–86.

Hanushek, Eric, 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, Working Paper WP12664. NBER, Cambridge, Massachussetts.

Harmon, C., and I. Walker. 1995. “Estimates of the Economic Return to Schooling for the United Kingdom.” *American Economic Review* 85, no. 5: 1278–86.

Heckman, J. 1979. “Sample Selection Bias as a Specification Error.” *Econometrica* 47, no. 1 (January): 153–62.

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

———. 2006b. “To Train or to Educate? Evidences from Tanzania.” GPRG Working Paper, no. 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* 1, no. 2 (August): 1: 192-205.

Koenker, R., and G. Bassett. 1978. “Regression Quantiles.” *Econometrica* 46, no. 1 (January): 33–50.

Lee, L. F. 1983. “Generalized Econometric Models with Sample Selectivity.” *Econometrica* 51, no. 2 (March): 507-512.

Machado, J., and J. Mata. 2000. “Box-Cox Quantile Regression and the Distribution of Firm Sizes.” *Journal of Applied Econometrics* 15, no. 3 (May/June): 253–74.

Patrinos, H., C. Ridao-Cano, and C. Sakellariou. 2006. “Estimating the Returns to Education: Accounting for Heterogeneity in Ability.” World Bank Policy Research Working Paper, no. 4040 (October). World Bank, Washington, DC.

Schultz, T., and G. Mwabu. 1998. “Labor Unions and the Distribution of Wages and Employment in South Africa.” *Industrial and Labor Relations Review* 51, no. 4: 680–703.

Teal, F. 2001. “Education, Incomes, Poverty and Inequality in Ghana in the 1990s.” CSAE Working Paper Series WPS/2001-21. Centre for the Study of African Economies, Department of Economics, University of Oxford.

THE WORLD BANK
1818 H Street, NW
Washington, DC 20433
www.worldbank.org/education

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 among 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. It also investigates the shape of the education-earnings relationship. Analysis is done separately by gender and age group and attempts to address the usual biases when estimating returns to education.