44241

EDUCATION
Working Paper Series
Number - 11

Education, Skills, and Labor Market Outcomes: Evidence from Pakistan

by
Geeta Kingdon and
Måns Söderbom

May 2008

THE WORLD BANK

Education, Skills, and Labor Market Outcomes: Evidence from Pakistan

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).

This paper is a background study for an upcoming World Bank study on education labor market linkages and has been developed through consultations with Tazeen Fasih of the Human Development Network Education Department of the World Bank.

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 workshop participants at the World Bank, in particular Tazeen Fasih and Alonso Sánchez, and to Francis Teal, for useful comments on an earlier draft of the paper. We also thank Alonso Sánchez for excellent research assistance.

Introduction

Policy interest in education is linked to its potential to raise earnings and reduce poverty. This paper investigates the education-earnings relationship in Pakistan, drawing on the Pakistan Integrated Household Surveys of 1998–99 and 2001–02. The analysis has three main goals: to examine the labor market returns to education among waged, 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. Because data are available from two points in time, the paper also investigates how these returns have changed between the periods 1998–99 and 2001–02.

While wage employment has been the object of most existing analyses, it is typically a small, and often shrinking, part of the labor market in developing countries. The labor market benefits of education accrue both from the fact that education promotes a person's entry into lucrative occupations and, conditional on occupation, raises earnings. The objective is to ask whether education raises earnings within any given occupation and whether it also raises earnings indirectly by facilitating entry into well-paying occupations, such as waged work. This exercise will be accomplished by estimating multinomial logit models of occupational attainment and earnings functions for the different occupation groups. The rate of return to education is estimated by occupation and for different levels of education, the latter in order to see the shape of the education-earnings relationship. In estimating the returns to education, the paper also attempts to correct for selectivity and endogeneity biases.

In addition, the paper interrogates the role of cognitive skills in both occupational attainment and earnings determination. There is evidence in the literature that cognitive skills have economically large effects on individual earnings and national growth. This evidence suggests that workers' productivity depends not only on years of education acquired, but also on what is learned at school. Hanushek (2005) cites three U.S. studies that show quite consistently that a one-standard deviation increase in mathematics test performance at the end of high school in the United States translates into 12-percent higher annual earnings. Hanushek also cites three studies from the United Kingdom and Canada that show strong productivity returns to both numeracy and literacy skills. Substantial returns to cognitive skills also hold across the developing countries for which studies have been carried out, including Ghana, Kenya, Tanzania, Morocco, Pakistan, and South Africa. Hanushek and Zhang (2006) confirm significant economic returns to literacy for 13 countries on which literacy data were available. While a previous study already exists for Pakistan, the data analyzed here offer a number of advantages over previous data.[1]

Finally, the paper investigates the role of education along the earnings-distribution to shed light on whether the effect of education is to reduce or accentuate earnings inequality. Analysis is conducted separately by occupation, gender, and age group.

The paper is structured as follows. The first section provides details on the empirical framework of the analysis, focusing on the specifications and estimators used. In order to ensure that the results are comparable, the same techniques and specifications are used to analyze data from 1998–99 and 2001–02. The second section analyzes the 1998–99 data, which is divided into a short section describing the data; a section that investigates the role of education and skills in determining occupational outcomes (where a distinction is made between wage employment, non-farm self-employment, agriculture, unemployment, and the state of being out of the labor force); and a section that examines the relationships between earnings, education, and cognitive skills. The third section analyzes the 2001–02 data, following the same structure as the preceding section. The conclusion summarizes the main findings of the paper.

1 Analytical approach

It is widely believed that education affects people's economic status by raising their earnings in the labor market. It may raise earnings through a number of different channels, for example, by improving access to employment or, conditional on employment, by promoting entry into higher-paying occupations or industries. This paper explores both the

[1] The wage equation in the Pakistan study by Behrman, Ross, and Sabot (2002) uses 1989 data on 207 wage employees from three districts of Pakistan, although it also estimates other equations. The main advantage of this study is that it tested the cognitive skills of respondents using standardized achievement tests and may therefore have better cognitive skills data than that available in the Pakistan Integrated Household Survey (PIHS, 1998–99). The authors of the 2002 study find that cognitive skills have statistically significant payoffs in the labor market. While the PIHS provides only self-reported measures of the ability of respondents to read and do simple sums, it has the advantage of being (i) nationally representative, (ii) 10 years more recent, (iii) both a rural and urban sample, and (iv) a larger survey with much larger samples: the wage equations in the present study are fitted for about 5,000 men and 700 women. Finally, while Behrman and his colleagues focus on the total return to cognitive skills, they do not examine the possible role of skills in promoting entry into lucrative parts of the labor market.

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

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

where w_i is the real earnings of individual *i*, $\mathbf{x}_{\mathbf{i}}$ is a vector of worker characteristics excluding education, $\boldsymbol{\alpha}_{\mathbf{a}g}$ is a parameter vector, s_i is the years of education, $f_{ag}(\cdot)$ is the earnings-education profile, υ_i is a residual, and *a* and *g* denote age group and gender, respectively. The primary objective of this paper is to estimate the total returns to education and the variables included in the $\mathbf{x}_{\mathbf{i}}$ are selected accordingly. In particular, estimates of earnings regressions do not condition on variables that are determined by education because conditioning on such variables would change the interpretation of schooling effects. For example, it is likely that an important effect of education is to enable individuals to get high-wage jobs (e.g., managerial positions), enter certain high-wage sectors or firms, or generate job security and thus work experience. Consequently, estimates here do not condition on occupation, firm-level variables, work experience, or other variables sometimes seen on the right-hand side in earnings regressions.

Earnings regressions are similarly not conditioned on land in the agricultural earnings equation, or capital stock for the self-employed, because it is assumed that investment in these assets may be driven by education. It is acknowledged that this may be a strong assumption, especially, perhaps, for the agricultural sector in a country where land is often inherited (and where land may therefore drive education). The effect of including these asset variables in the regressions is therefore included in the discussion that follows. The analysis focuses, however, on regressions that include only a small set of control variables, with age and gender emphasized most. With respect to the effects of these variables on earnings, a fair deal of flexibility is allowed and all regressions are estimated separately both for men and women and for relatively young individuals (aged less than or equal to 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.

Estimation of the earnings-education profile $f_{ag}(\cdot)$ is the key purpose 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 where a quadratic term is added to the linear specification is considered, providing a convenient way to test 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 exists on earnings as well as on the explanatory variables. For individuals that are either self-employed or work in the agricultural sector, no earnings data are available at the individual level. Instead, earnings at the household level, which distinguish between earnings of the self-employed and agricultural workers, are used. 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, that is, to calculate the mean values of the explanatory variables within the household and labor market subsector (obviously, this operation is not performed for the wage employed, as individual-level data exists on their 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 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, a condition that 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 employment, self-employment, 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 in the earnings function (1),

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

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

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

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 λ_{ji} as an additional explanatory variable in the wage equation, so that

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

where z_{ij} is a set of variables explaining selection into occupation and γ are the associated coefficients. Thus, the probability of selection into each occupation type is first estimated by fitting a model of occupational attainment, based on which the selectivity term

(λ) is computed.[3] The coefficients on the lambda terms λ_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 the state of being out of the labor force. Occupational attainment is accordingly modeled using a multinomial logit equation.

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 function 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

---

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

estimating the earnings functions separately for occupation groups, since each of these may not be a random draw from the population. Finally, the analysis attempts to address the problem of endogeneity by using a family fixed effects model.[4]

The paper also estimates earnings functions by the quantile regression (QR) method. OLS regression models the *mean* of the conditional distribution of the dependent variable. However, if schooling affects the conditional distribution of the dependent variable differently at different points in the wage distribution, then quantile regressions are useful because they allow the contribution of schooling to vary along the distribution of the dependent variable. Thus estimation of the 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 investigates 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 are 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 Results for 1998–99

This section undertakes a detailed analysis of the Pakistan Integrated Household Survey (PIHS) 1998–99 data. Analysis is divided into three parts. First, details on the sample and summary statistics on key variables are provided. Second, the effects of education and cognitive skills on occupational outcome are examined, and third, the effects of education and cognitive skills on earnings, conditional on occupational outcome, are analyzed.

Data and descriptive statistics

Following a two-stage sampling strategy, the PIHS provided a nationally representative sample made up of around 16,000 households, which represented roughly 115,000 observations.[6] The household questionnaire was composed of a number of detailed modules on such characteristics as income, education, health, maternity, family planning, consumption, expenses, housing conditions, and available services. In addition, certain modules concentrated on household enterprises and agricultural activities, including associated expenses and revenues. These modules enabled the present analysis to define five occupation categories: wage employment, non-farm self-employment, agriculture, unemployment, and out of the labor force.

The construction of the earnings variable is an important issue. For individuals who are either unemployed or out of the labor force, a measure of earnings cannot be constructed. For self-employed and agricultural workers, earnings 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. Earnings of paid employees are, by contrast, derived from the sum of reported income—cash, other occupations, in kind, pensions, and so forth—from the income module.

---

[6] The authors are most grateful to Alonso Sánchez for his substantial input to this subsection.

Table A1.1[7] shows summary statistics for selected variables used in the analysis, both for the full sample, and for the five identified occupation categories. The sample consisted of individuals aged between 16 and 70 years who were not currently enrolled in school. Unemployed individuals are those seeking employment and available for it, while individuals who were out of labor force (OLF) are those not seeking employment (e.g., housewives and the retired). The labor force participation rate in Pakistan in 1998–99 was about 51 percent and the unemployment rate, 6 percent.

Table A1.1 shows that average earnings in the full sample were 30,277 Pakistan rupees, which corresponds to approximately US$600. There are significant differences in average earnings across the three job categories for which a measure of earnings could be constructed (not possible for nonworkers). Self-employed and wage-employed workers earn on average about 70 percent more than individuals working in the agricultural sector. This finding is mirrored by a similar differential in education: average years of education among agricultural workers is 2.5, whereas for the self-employed and wage employed, average education is between 4.5 and 5.4 years. It is worth noting that the average level of education among the OLF category was similar to that for agricultural workers. The pattern for literacy and numeracy skills is similar: 55 percent or more of individuals in self-employment, wage employment, and unemployment can read and write, and about 70 percent or more have basic math skills, while in agriculture and among the OLF category, less than 35 percent can read and write and less than 60 percent have basic math skills. Finally, although the mean earnings of the self-employed exceed the mean earnings of the wage employed, this is neither true for earnings expressed in natural logs (where the numbers imply that wage employment carries a 17-percent premium compared to self-employment) nor for median earnings. The latter finding is explained by the fact that the distribution of earnings differs across sectors, as can be seen the lower panel of table A1.1.

In summary, although five occupation categories are distinguished in the data, the main difference with regard to skills and earnings is between self-employed and wage-employed workers on the one hand, and agricultural workers and the OLF category, on the other. This suggests that skills matter a great deal in determining into which of these two

[7] For ease of reading, all tables and figures have been removed to appendices. Tables and figures associated with analysis of the 1998–99 data are found in appendix 1, and selectivity corrected tables for this data, in appendix 2. All tables and figures are first identified first by the number of their respective appendix: table A1.1 (appendix 1), table A2.1 (appendix 2). No appendices are included for the 2001–02 data because the findings of the two surveys are so similar.

broadly defined occupation groups individuals are sorted. While unemployed individuals possess the mean skill levels of waged and self-employed persons, they clearly queue for suitable job opportunities in the labor market.

Education and occupational attainment

As shown clearly in table A1.1, average earnings vary dramatically between individuals that are either self-employed or wage employed, on the one hand, and individuals that work in the agricultural sector, on the other. The table also shows that the average level of education and skills varies substantially between these two groups. It therefore seems very likely that one channel by which education raises incomes in Pakistan is by enabling individuals to get a job in a high-earnings sector. This section looks at the effects of education and skills on occupational outcome. From a policy point of view, the link between skills and labor market outcomes among the relatively young deserves special attention. Accordingly, the following subsection analyzes labor market outcomes for the young age group (aged 16–30 years) separately from that for the old age group (aged 31–70 years).

To understand the role played by skills and family background in this context, occupational outcome is modeled by means of a simple, parsimoniously specified multinomial logit. The explanatory variables are education, skills, and basic individual and family characteristics (age, marital status, number of young children in the household, and number of elderly people in the household), and province dummies. While the multinomial logit is a useful estimator in this context, one drawback is that the estimated coefficients are hard to interpret. Marginal effects and graphical analysis are therefore reported, based on the results of the multinomial logits (see appendix 2 for all underlying regression results).[8] Whenever education is included as an explanatory variable, literacy and numeracy variables are excluded, and vice versa, because 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.

First, the occupational outcomes are modeled for men and women, as well as for age group (young and old), with years of education used as the measure of skill. Table A1.2 shows marginal effects for number of children, number of elderly people in the household, and marital status. While these findings are not of central interest, it is perhaps worth noting

[8] All regressions are run separately for men and women.

that the number of children significantly reduces the likelihood that an individual is in (highly paid) wage employment for men, but somewhat surprisingly, not for women. One possible reason is that wage employment is a less flexible occupation (in terms of working hours, for example) than the other job categories considered here. For men, being married strongly increases the likelihood of working and reduces the likelihood of being unemployed or OLF. For women being married decreases the likelihood of working (except for older women in agriculture) and strongly increases the likelihood of being OLF.

Figure A1.1 illustrates the estimated association between years of education and the predicted likelihood of different occupational outcomes for young men (panel i) and young women (panel ii), evaluated at the sample mean values of the other explanatory variables in the model. Clearly, the likelihood of men being a wage employee is relatively invariant to the education level of the individual. By contrast, education is clearly associated with a lower likelihood of being involved in agricultural production. Strikingly, the likelihood of being a nonworker (i.e., either unemployed or OLF) *increases* with years of education. One possible reason for this result is that individuals with a great deal of education are willing to wait for a good job opportunity before taking paid employment. The likelihood of self-employment can be graphed as an inverse "U" with respect to education, peaking at about eight years of education.

For women the picture is very different indeed. Women with up to about eight years of education are unlikely to work. As education increases to the secondary level and beyond, however, the likelihood of wage employment increases quite dramatically. Indeed, according to these estimates, the likelihood that a woman with a university degree (approximately 16 years of education) has a waged job is approximately 0.50. Correspondingly, education has no relationship with the labor force participation of women until they have reached roughly 10–12 years of schooling, after which their participation rises sharply with education (i.e., the OLF curve falls sharply). It is thus very clear that education matters much more for women than for men in Pakistan in terms of determining the type of occupation.

Figure A1.2 plots the estimated occupation probabilities as a function of age, again for young persons (aged 16–30 years), holding all other explanatory variables fixed at sample mean values. This figure is informative of the nature of the transition from education to work. Perhaps the most interesting result here is that women enter gainful employment relatively late, only after about age 25. By contrast, between the ages of 15 and 25, men enter the labor force at a rapid rate so that by about age 25, almost all men are already labor force participants (i.e., the OLF curve falls sharply between ages 15 and 25). The relationship between age and participation in wage employment is a strikingly inverted "U" shape: up to about age 25, the likelihood of wage employment increases with age, but the relationship then becomes less strong. A similar though far less pronounced pattern is discernible in agricultural employment. The chances of self-employment rise throughout with age, but somewhat more steeply after about age 24. This result can possibly be explained by the fact that young people can only enter self-employment once they have accumulated some savings.

Figures A1.3 and A1.4 repeat the type of calculations illustrated in the previous two figures, only for older individuals (aged 31–70 years). In figure A1.3, a striking difference regarding the role of education is apparent for men: among the young, the likelihood of being a wage employee is by and large unresponsive to education. Highly educated young men are basically either wage employees or not gainfully employed (i.e., unemployed or OLF). By contrast, older men's likelihood of being wage employed is strongly responsive to education. Among older women the basic patterns are similar to those of the young.

Table A1.3 presents the marginal effects of basic literacy and numeracy on the likelihood of being in different labor market states. The descriptive statistics discussed earlier clearly established that wage employment and self-employment, not agriculture, are the well-paying parts of the labor market in Pakistan. Overall, table A1.3 shows that possession of literacy promotes entry into a well-paying part of the labor market, namely wage employment, for all groups except young men. In the older group, the effect is three times as large for men as for women. Literacy skills very strongly reduce the chances of ending up in the worst-paying part of the labor market, namely, agriculture; the effect is significantly higher for men than for women in both age groups. Somewhat surprisingly, however, being literate is associated with significantly *increased* chances of both being OLF and unemployed for all groups. Literate women either work in wage employment—which may be viewed as the respectable part of the labor market—or remain OLF (and to a lesser extent, unemployed). They perhaps remain OLF due to cultural norms or their greater efficiency in the production of home goods. A weak suggestion exists that literacy reduces both young and old women's entry into self-employment, but promotes that of young men.

Also somewhat unexpected, numeracy is not related to a worker's chances of being in wage employment, suggesting that many waged jobs are unskilled and thus do not require numerate individuals. For men, however, numeracy has a high association with a worker's chances of being self-employed. This finding could be explained either by the fact that numeracy promotes entry into self-employment (i.e., causation runs from being numerate to entering self-employment) or by the fact that people in self-employment end up becoming numerate (i.e., numeracy is learned on the job). Either way, there is no such positive relationship between numeracy and self-employment for women, suggesting that many self-employed women may be at a disadvantage. Numeracy skills also reduce the chances of being OLF for men, but being numerate is evidently not an escape route from the OLF state for women, a finding that could be due to cultural norms or differential earnings rewards of numeracy for men and women.

Of note, the marginal effects of cognitive skills on occupational outcomes are generally smaller in size for the young. For instance, while literacy reduces the chances of agricultural employment very substantially for both young and old men alike in the two respective samples, the relationship is significantly smaller in the young sample (–11.0 points, compared with –16.7 points for the old sample). Similarly the relationship between numeracy and the likelihood of self-employment for young men is less than half that for older men. When moving from the old sample to the young sample, the reduction in the size of the relationship is generally smaller for women than men.

Education and earnings

The basic relationship

Several authors have estimated returns to education in Pakistan; Aslam (2007) provides an annotated list of papers and their strengths and weaknesses. In line with much of the international literature on economic returns to education, these studies have estimated returns to education solely in wage employment. However, as seen in table A1.1, wage employment absorbs only about half of the total labor force, meaning that half of the labor force is engaged in self-employment, both agricultural and non-agricultural. What are the returns to education in this major part of the labor market? To the authors' knowledge, this question has not been addressed for Pakistan. The term "returns to education" is used here as it is commonly used 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 A1.4 presents basic OLS estimates of the Mincerian returns to education in Pakistan by occupation, gender, and age group, and shows that the returns to education are very precisely determined, even in cases where sample sizes are very small. As shown below, the pattern of returns to cognitive skills mirrors the pattern of returns to education, indicating a high correlation between schooling and skills.

It is clear that the returns to education are invariably statistically significantly greater for the older than the younger sample. In the older age group, the earnings premium associated with each extra year of schooling is significantly greater than in the young age group. A plausible explanation for this phenomenon is 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 (Knight, Sabot, and Hovey 1992). At the time the old sample got their jobs, primary completers were in more scarce supply and five to eight years of education may have been sufficient to obtain a white-collar job. People who obtained such jobs remain in them today. However, due to the rapid expansion of the supply of educated persons, young people (16–30 years) who complete grades 5 to 8 today may be fortunate to even get a low-paying, blue-collar waged job. For the uneducated, there is less scope for filtering down of occupations so that, over time, wages are compressed by education level. Thus the rate of return to education may be lower for younger workers because they perform different tasks, tasks for which education is less valuable than tasks performed by older persons with the same education level.

Table A1.4 also shows that returns to education are significantly and substantially greater for women than men in all occupations and in both age groups, with the exception of young women in agriculture.[9] The fact that returns to education in *wage* employment in Pakistan are about three to four times as high for women as for men (both young and old) could reflect the scarcity of educated women, combined with the existence of jobs that require (or which are largely reserved for) educated women, such as nursing and primary school teaching, which are predominantly female jobs in Pakistan. However, the reasons why women have a higher earnings premium than men in self-employment are less clear, even though the female premium is not so high in self-employment as in wage employment. For young men, on the other hand, returns to education are particularly low in agriculture and wage employment.

---

[9] When the sample is not divided into young and old age groups and pooled equations (not shown) are estimated, the return to each extra year of schooling in *wage* employment is 5.3 percent for men and 16.0 percent for women (i.e., three times higher), which is similar to estimates by gender in Pakistan based on PIHS 2001–02 data (Aslam 2006).

Interestingly, returns to education in agriculture are similar to those in other occupations, at least among the older age group, a finding also detected in Argentina (see Gallacher 2000), where the returns to education in agriculture for farms of average size are equal to the returns to education in wage employment.[10]

The existence of substantial returns to education in self-employment is welcome news for Pakistan because it suggests that education plays a poverty-reducing and productivity-enhancing role not only in wage employment—an increasingly shrinking sector in many labor markets—but also in other, potentially faster-growing sectors of the labor market. The gender pattern of returns is also welcome for women and provides them strong economic incentives to acquire schooling. Given that Pakistan has one of the world's largest (if not the largest) gender gaps in school enrollment and literacy, these strong labor market incentives can help redress those gaps, provided that the supply of schooling is ensured and credit constraints that impede girls' enrollment are removed. (Attendance-contingent cash subsidies, together with a female school stipend program, have virtually eliminated gender gaps in secondary school enrollment in Bangladesh).

However, even though the returns to education may be high for women, they actually earn much less than men in Pakistan. In other words, although the slope of the education-earnings relationship is three times as steep for women as for men, the intercept of the wage regression is much higher for men. Men enjoy earnings premiums at all levels of education, but particularly large ones at lower levels of education. This is clear from the graphs of predicted earnings in figures A1.5, A1.6, and A1.7, in which the slope of the education-earnings relationship is steeper, but the intercept is far lower, for women than for men. As Aslam (2007) shows, a large part of the gender gap in earnings is not explained by differences in men's and women's productivity endowments, such as education and experience, but by potential discrimination in the labor market. The education of women helps reduce the earnings gap, i.e., there is less gender discrimination among the educated in the Pakistan labor market. If Pakistan thus wishes to reduce gender gaps in education by improving women's incentives to acquire an education, it needs not only to improve school

[10] A rather dated review by Lockheed, Jamison, and Lau (1980) surveyed studies that used agricultural production functions to measure the effect of farmer education on farm output. Whereas in some countries the estimated return on primary education was high, a statistically significant effect of education was found in only 19 of the 37 data sets. The effect of education on rural productivity seemed to depend on whether there was a modernizing agricultural environment.

supply and ease credit constraints, but also to reform labor market policies in ways that reduce gender-differentiated treatment by employers.

The earnings equations for self-employed and agricultural workers were also estimated adding controls for productive assets. In the case of the self-employed, the log of the capital stock value (defined as the replacement value of buildings, plant, and equipment) per self-employed individual in the household is added, while for agricultural workers, the log of acres of land per individual engaged in agricultural production in the household is added. This addition means that the analysis moves from estimating reduced-form earnings equations towards estimating profit functions with controls for fixed inputs, a procedure that somewhat changes the interpretation of the results.

The results (not reported) indicate that controlling for the log of the capital stock has marginal effects (about one percentage point or less) on the coefficients on education for self-employed men, but for self-employed women, the coefficients are approximately halved. The coefficient on log capital is always statistically significant and varies between 0.12 and 0.17, except for old women, for whom it is 0.27. For agriculture, the coefficient on education falls by less than 0.01 for both young men and women and by about one-third for old men and women. The coefficient on log land is always significant and varies between 0.32 and 0.45, except for old women, for whom it is equal to 0.10.

How one interprets these results depends on the causal relationship between education and productive assets. If, on the one hand, assets depend on education (e.g., because education raises the marginal product of land, meaning educated farmers choose more land), then the earlier results (without controls for assets) can be interpreted as showing the total effect of education on earnings. If, on the other hand, education depends on assets (perhaps because land is inherited and parents with a lot of land ensure that their children get a lot of education), then the results with controls for land suggest the earlier results overestimate the effect of education on earnings. The truth is probably somewhere in between. Unfortunately, without more detailed data (e.g., information on assets at the time schooling decisions were made), it is difficult be more precise on this issue.

Extensions on the education-earnings relationship

Correcting returns estimates for endogeneity bias

As stated at the outset of this paper, OLS estimates of returns to education potentially suffer from sample selectivity bias and endogeneity bias. The analysis here attempts to address the former bias by employing the Heckman procedure. The multinomial logit equations in appendix 2 were used to calculate the selectivity terms, the results of which are presented in table A1.5. The selectivity term is statistically significant in 5 out of 12 earnings regressions. The introduction of the selection term generally reduces the returns to education and in three cases (waged young women and waged old men and women), this reduction is statistically significant. Since selectivity correction makes a difference in some cases, the selectivity corrected equations are preferred to OLS in this paper.

The problem of endogenous sample selection is akin to the problem of endogeneity (or "ability") bias discussed earlier in this paper. The endogeneity issue is addressed by estimating a household fixed effects earnings function for waged work. This cannot be estimated for self-employed or agricultural workers because no within-household variation exists in these cases. The results shown in table A1.6 yield similar results to those in table A1.5: returns to education fall in comparison with the OLS returns in table A1.4, although they generally fall more than when they are corrected for selectivity bias in table A1.5.[11] The household fixed effects approach is a powerful way to address endogeneity since the identification of the effect of education on earnings is derived only from within-family variation in earnings and education and accordingly nets out the effect of shared ability, akin to the twin-differencing approach. However, the reduction in estimated returns to education in table A1.6, compared with the OLS results in table A1.4, may represent more than simply a correction for endogeneity bias. The reduction may also represent measurement error bias, which is exacerbated in differenced models and biases coefficients downwards. For this reason, and because the household fixed effects results can be estimated only for the subsample of wage-employed persons, the selectivity corrected results are the preferred estimates in this paper.[12]

---

[11] Table A2.9 (appendix 2) presents household fixed effects estimates of the earnings function for waged workers by education *level* rather than *years* of education.

[12] The linear model for wage employees has also been estimated using two-stage least squares, the results which are summarized below. Young men: using father's and mother's education as instruments, and losing

Shape of the education-earnings relationship

What is the shape of the education-earnings relationship in different occupations? The analysis so far has imposed a linear relationship between "years of education" and earnings (Table A1.5). Table A1.7, estimated using the preferred sample selectivity corrected estimator, relaxes the implicit presumption of linearity by introducing quadratic terms in education. (Its OLS and household-fixed-effects counterparts are included in tables A2.9 and A2.10, respectively.) Table A1.7 shows no common pattern in the shape of the education-earnings relationship across occupations. In wage employment, the education-earnings relationship is convex for both old and young men; in agricultural employment, it is convex only for old men. The relationship is concave only for one group: for old women in wage employment. For all other groups, the relationship is evidently linear. Thus, the Pakistan labor market is not generally characterized by the commonly assumed concave relationship that implies diminishing returns to extra years of schooling.

The nonlinearities of the education-earnings relationship are explored further in table A1.8, which includes a dummy variable for each education level. The selectivity correction estimator is preferred, as before. OLS yields significantly higher coefficients compared with selectivity corrected estimates in several cases and is relegated to table A2.11; the household fixed effects results for the wage employed are shown in table A2.12. The base education category is "no education." The marginal return to each year of primary education, each year of middle education, and so forth, calculated from table A1.8, are set out in table A1.9. The latter table confirms certain patterns noted earlier. For instance, it shows that marginal returns to education are generally substantially lower for men than for women in

---

about 50 percent of observations in the process (see footnote 4), the coefficient on education rises from 0.033 (OLS, table A1.4) to 0.064 (significant at the 1-percent level), and the validity of the over-identifying restrictions is rejected at the 5-percent level; adding spouse's education to the instrument is not feasible, as too many observations are lost; using spouse's education as the only instrument, 60 percent of observations are lost and the coefficient rises to 0.068 (significant at the 1-percent level).

Young women: using father's and mother's education as instruments, 60 percent of observations are lost, the coefficient on education falls from 0.149 (OLS, table A1.4) to 0.137 (significant at the 1-percent level), and the validity of the over-identifying restrictions is accepted at the 10-percent level; adding spouse's education to the instrument is not feasible as too many observations are lost; using spouse's education as the only instrument, 60 percent of observations are lost and the coefficient rises to 0.18 (significant at the 1-percent level).

Old men: parental education cannot be used as an instrument, as too few individuals in this age group live with their parents; using spouse's education as the only instrument, 10 percent of observations are lost and the coefficient rises from 0.066 (OLS, table A1.4) to 0.102 (significant at the 1-percent level).

Old women: parental education cannot be used as an instrument, as too few individuals in this age group live with their parents; using spouse's education as the only instrument, 30 percent of observations are lost and the coefficient rises from 0.172 (OLS, table A1.4) to 0.184 (significant at the 1-percent level).

both wage employment and self-employment, although not in agriculture. It also shows that marginal returns are generally higher for the older age group than for the younger one, particularly so for waged women at primary and middle school levels. Among young men in waged employment, the marginal returns to education increase monotonically with education level so that an extra year of education is progressively more valuable when acquired at successively higher levels of education. This pattern also holds, somewhat more loosely, for young and old waged women, since their marginal return to education at the secondary school level is substantially higher than at the primary level.

For women, returns estimates beyond secondary school are typically insignificant because they are based on very small samples (few women in the sample had more than lower-secondary education). Taken together, the evidence of tables A1.7, A1.8, and A1.9 suggests that the education-earnings relationship in Pakistan is not concave in any of the occupations, that is, there is no evidence of diminishing marginal returns to education in the country. This finding is confirmed in figures A1.5, A1.6, and A1.7, which show the relationship between education and predicted earnings.

Earnings and cognitive skills

Table A1.10 shows OLS earnings functions and table A1.11, selectivity corrected earnings functions, by occupation, with cognitive skills measures on the left side. The first set of columns ("1. Wage employed") are individual-level earnings functions for waged workers, estimated separately for men and women. The next set of columns are earnings functions at the household level that account only for household members employed in a household enterprise (self-employment), and the third set of columns, earnings functions at the household level that account only for members in agricultural self-employment. Years of schooling is not included in the earnings functions because the goal was to estimate the total return to cognitive skills, irrespective of whether or not they were acquired through schooling.[13]

[13] A simple regression of years of education on literacy and numeracy, age, and age squared (pooled across age and gender groups) indicates that being literate is associated with 8.06 extra years education, while being numerate is associated with 0.3 extra years of education. A crude comparison of the coefficients on the cognitive skills variables to those on education reported earlier can thus be obtained by multiplying the education coefficients in the linear specifications by 8 (yielding an indirect estimate of the partial effect of literacy) and 0.3 (yielding an indirect estimate of the partial effect of numeracy). Note that this procedure

As table A1.11 shows, selectivity clearly matters in wage employment among both the young and old: the inclusion of the selectivity term significantly reduces the coefficients on the literacy skills variable (“can read and write”) among young women and among old men and women. Consequently, the selectivity corrected results are discussed here. (Household fixed effects results are reported in table A2.13A and show smaller effects than those in table A1.11, either because the fixed effects method provides a tighter upper bound on the skill effect than a selectivity correction approach, or because of attenuation bias in the fixed effects equation due to heightened measurement error.)

Table A1.11 shows strong returns to literacy in all occupations. In most cases, these returns are dramatically larger for women than for men, a result that is at least partly due to a scarcity premium—far fewer women than men are literate in Pakistan. Not only do fewer women than men have the years of schooling required to develop literacy skills, women are more likely than men to have attended poorer-quality schools.[14]

While literacy returns in waged work are only about one-fifth as large for men as for women, they are nevertheless substantial and statistically significant. In agriculture, literacy has striking payoffs for men. Literate men are significantly more productive than illiterate men (for women, the point estimate is large but not statistically significant), with the literacy return of young men double that of their literacy return in wage employment.

Significant positive returns to *numeracy* skills also accrue both to old men and women in agriculture. While they also accrue to old men in waged work, the size of this return is only-one third as large as that in agriculture. Among the young, returns to numeracy are confined to men in agriculture. The presence of productivity returns to literacy and numeracy skills for men suggests that Pakistani agriculture is not traditional: the ability to read and do simple calculations (that would allow a person to, for example, follow instructions on fertilizer packs) raises agricultural earnings. The lack of returns to skills in agriculture for women could arise because household males make farming decisions due to the gender division of roles in this traditional occupation.

---

will produce ballpark numbers only. A more rigorous approach would be to allow for different correlations between education and skills variables across the age and gender group.

[14] Aslam and Kingdon (2006) show that within the household, girls in Pakistan receive significantly lower educational expenditures than boys. Aslam (2007) finds that girls also face poorer-quality schooling than do boys both because they are significantly less likely to be sent to private schools than their brothers and because private schools are more effective than public schools in imparting cognitive skills to students. Her findings on the relative effectiveness of private and public schools are supported by other studies on Pakistan (Alderman, Orazem, and Paterno 2001; Andrabi, Das, and Khwaja 2002; Arif and Saqib 2003).

The educational decisions of today's children will depend much more on the observed pattern of returns to education and to skills among young adults rather than the old. That development of numeracy and literacy skills is a profitable investment for *young* men, even in agriculture, is cause for some optimism because it shows that rural males have a private economic incentive for acquiring cognitive skills in Pakistan. However, if the quality of education is low, it can take many years of schooling to develop literacy and numeracy. There is some support for this hypothesis. Young men's return to education in agriculture is statistically significant at the 5-percent level only from middle school onwards, suggesting that it takes eight years of schooling to acquire cognitive skills. This finding highlights the importance of the quality of schooling: the higher the quality of schooling, the greater the economic benefit of an extra year's education.

Heterogeneity in returns to education

While economists have generally estimated the average of the marginal returns to education, in actual fact, returns to education can be heterogeneous across people—a finding that has implications for the inequality-reducing role of education. However, distribution of the returns to education across the earnings spectrum is not known for Pakistan, as is true for most developing countries. This analysis therefore examines heterogeneity in the returns to education in Pakistan to ask whether some individuals benefit more from education than others and why, and then examines the inequality implications of the answer.

A literature now exists that investigates the pattern of returns to an additional year of education along the earnings distribution using quantile regression (QR) analysis. The results of this type of analysis suggest that in developed countries, the returns to education increase with quantiles (i.e., they 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 (i.e., returns to education are higher for lower earnings quantiles).[15]

If the returns to education increase as one goes from the lower to the higher end of the earnings distribution, this trend can be interpreted as indicating that ability and education

[15] 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).

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.

PIHS data from the 1998–99 round was used to estimate quantile regressions, which are reported in table A1.12. The results show that in wage employment, returns to education for women are highest in the lowest quantile of earnings (bottom quartile) and lowest in the highest earnings quantile (the top quartile). In other words, those with lower ability have higher rates of return to education. This finding is true for women in both the young and old age groups, suggesting that for women waged workers, education is inequality reducing because it reduces rather than increases wage differences between low- and high-ability individuals. There is no such pattern for males.

In self-employment among both young women and old men, education seems to be mildly inequality increasing. For self-employed young women, the return to education in the top quartile (of the conditional distribution of earnings) is nearly double that in the lowest quartile, although this difference is not statistically significant due to the imprecision of estimates based on a small sample size. For old men in self-employment, the return to education in the top quartile (7.2 percent) is 1.6 percentage points (or 28 percent) higher than that in the bottom quartile (5.6 percent), a difference that is statistically significant since both are very precisely determined. However, the size of the difference in returns is not economically large. It can thus be said that in agriculture and self-employment, there is no strong pattern of differential returns to education at different points of the conditional earnings distribution.

While women with lower ability have higher rates of return to education among both the young and old in wage employment, the extent of the difference in returns to education between the bottom and top quartiles of conditional earnings is significantly larger among old than young women. In other words, education is more inequality reducing in the older waged women's group than in the younger. The inequality-reducing role of education for women in wage employment is akin to a social externality of women's education and further boosts the already strong efficiency case for public subsidization of girls' schooling in Pakistan.

3 Results for 2001–02

This section analyzes data from the 2001–02 round of the Pakistan Integrated Household Survey (PIHS). The text follows the same structure used in the previous section. First, details on the sample are provided, including summary statistics on key variables. Second, the effects of education and cognitive skills on occupational outcome are examined, and third, the effects of education and cognitive skills on earnings, conditional on occupational outcome, are analyzed. The main purpose of this part of the analysis is to see if the key findings based on the earlier wave of data have changed, and if so, how. In the interest of brevity, the text concentrates mostly on the changes that occurred.[16]

Data and descriptive statistics

The sampling strategy for this survey was the same as that used for the 1998–99 survey, discussed earlier. The same procedures for defining occupations and calculating earnings were used for both rounds of the survey, which allow for adequate comparison of data.

Comparing summary statistics of the 2001–02 survey with those of the 1998–99 survey, there is clearly a good deal of similarity in the overall labor market picture in most respects. This finding is perhaps unsurprising, given the short three-year gap between the two survey rounds. The labor force participation rate in Pakistan remained the same, about 51 percent, and the distribution of the adult population into the different labor market states did not change greatly except for the proportion of the labor force employed in agriculture, which fell by 5 percentage points over the 3 years, from 30 percent to 25 percent. Correspondingly, the proportion of the labor force employed in wage employment rose by 2.5 percentage points; in self-employment, by 1.4 percentage points; and in unemployment, by 1.2 percentage points.

Average earnings in the full sample did not change in nominal terms (suggesting a fall in real terms), although this finding masks modest changes in the opposing direction in the mean earnings of self-employed and wage-employed groups. The very large difference in mean earnings between agricultural workers, on the one hand, and both self-employed and

[16] Tables and figures illustrating the 2001–02 data are not included in appendices because the analytical findings are very similar to those based on the 1988–89 data.

wage employed workers, on the other, remained in 2001–02. The hierarchy of average years of education by occupation also did not change between the two surveys, although mean years of education among the self-employed increased conspicuously and mean education in wage employment fell to some degree. These changes in education by occupation explain, at least in part, the evident reduction in mean earnings in wage employment and the rise in mean earnings in self-employment over the three years. The percentage of workers who were numerate increased appreciably over time in most occupations, but not the percentage of workers who were literate.

In summary, descriptive statistics show that the overall labor market picture did not change greatly. Relevant quantities moved in the expected directions, for example, the proportion of people employed in agriculture decreased, while mean education and cognitive skills rose and the mean number of children fell over time. As in 1998–99, it remained the case in 2001–02 that the main difference between the skills and earnings of the different occupational groups was between self-employed and wage-employed workers, on the one hand, and agricultural workers and people in the OLF category, on the other. This finding suggests that skills continue to matter a great deal in determining into which of these two broadly defined occupation groups individuals are sorted.

Education and occupational attainment

When occupational outcomes are modeled for men and women, and for age group (young and old), using years of education as the measure of skills, the results are remarkably similar to those shown in table A1.2 for the 1998–99 period. Differences suggest that the way in which household demographics impinge on occupational choice and/or outcomes seems to have become stronger over time. Several examples illustrate this hypothesis.[17]

The relationship between years of education and the predicted likelihood of different occupational outcomes for young men and young women, evaluated at the sample mean

---

[17] For example, the results for number of elderly persons and for marriage are stronger in 2001–02 than in 1998–99, both in terms of size and statistical significance. Similarly, the marginal effect of the number of elderly people on the chance of being in agriculture are mostly stronger in 2001–02. Again, the marginal effect of having elderly persons in the household on the chance of being wage employed are significantly stronger for men in 2001–02 than in 1998–99. Lastly, number of children in the household has a statistically significant positive effect on older men's chances of being OLF in 2001–02, but not in 1998–99. Of course, this result was not always the case and the data provide one or two counterexamples, for example, the marginal effect of the number of children on the chance of being self-employed are less strong in 2001–02, but this finding is a less common occurrence when comparing data for the two periods.

values of the other explanatory variables in the model, again shows that education is associated with a lower likelihood of being involved in agricultural production and a higher likelihood of being either OLF or self-employed and has an inverse "U" shape with respect to education. There is suggestion in the 2001–02 data that education weakly reduces young men's chances of being in wage employment. The extent to which education reduces young men's chances of being in agricultural employment has become more muted over time: the slope for agriculture with respect to education is visibly flatter than in the earlier period. For women the picture for 2001–02 is remarkably similar to that for 1998–99: women with up to about eight years of education are very unlikely to work, but education at the secondary level and beyond strongly raises the likelihood of wage employment. Education also matters more for women than men in terms of determining the type of occupation.

When estimated occupation probabilities are plotted as a function of age for young persons (aged 16–30 years), holding all other explanatory variables fixed at the sample mean values, women's age matters little to their labor force participation decision. At any age, women have only about a 20-percent chance of being in the labor force. By contrast, age matters strongly to men's decision to enter gainful employment, that is, the OLF curve falls sharply between ages 15 and 25. Age and men's waged work participation have a inverted "U"-shaped relationship. Comparing results to those of 1998–99 indicates that the relationship between age and the chances of self-employment was far steeper in 1999 and that women's occupational outcomes have become less responsive to age over time. This finding is particularly conspicuous for the relationship between age and the chances of being in wage employment and, to a lesser extent, OLF.

The 2001–02 data show that older men's likelihood of being wage employed rises strongly with education beyond five years of education, but older women's chances of wage employment rise with education only beyond 10 years of education. Education also deters entry into agriculture. Among older women, very high levels of education make it pretty certain that they will be in waged work; the relationship is steeper than for young women. Education beyond the secondary level also spurs older women to become labor force participants. The basic patterns for older women are similar to those in the earlier data set. With respect to the relationship between *age* and *occupation outcome* for older men and women, age raises the chances of wage employment for men, but has relatively little effect on their likelihood of entering other occupations, in contrast to the data for 1998–99, which showed that older men's chances of being OLF fell strongly with age. Among older women in 2001–02, age decreases the likelihood of being OLF (i.e., increases their chances of workforce participation) and increases the likelihood of both waged and agricultural work, although the latter relationship is flatter than in the data for the earlier period.

The marginal effects of literacy and numeracy on occupational outcomes in the 2001–02 data show that literacy clearly promotes entry into well-paying parts of the labor market, namely, wage employment (except for young men) and self-employment (among men). Literacy skills also greatly reduce the prospect of being in the lowest-paying part of the labor market, namely, agriculture. Numeracy skills strongly increase the probability of being in well-paid work, that is, both wage employment and self-employment. In comparison with the 1998–99 data, the relationship of *literacy* with occupational outcome generally fell in some cases (e.g., for the chances of most groups to enter wage employment) and rose in other cases, (e.g., old women's chances of being OLF). However, the relationship of *numeracy* with occupational attainment is mostly stronger in 2001–02 than in 1998–99. This finding is most conspicuous in wage employment and the OLF category, but it is also found among older workers in agriculture and among women in unemployment. The existence of strong positive relationships between cognitive skills (literacy and numeracy) and the likelihood of accessing better-paid occupations, and the fact that this relationship (particularly with numeracy skills) has become greater over time, suggests that there is competition in Pakistan for well-paid jobs and that skills increasingly play a bigger role in sorting people into different occupations or rationing the better-paying jobs.

Education and earnings

The basic relationship

Basic OLS estimates of the Mincerian returns to education in Pakistan by occupation, gender, and age group for 2001–02 reveal strikingly similar results for the wage-employed as those based on the earlier data set. For young men, the estimated coefficient on education is 0.033 (compared to 0.035 for 1998–99), while for young women, it is 0.144 (0.149 for 1998–99). Further, the education coefficient for old men is 0.066 (0.070 for 1998–99) and for old women, 0.183 (0.172 for 1998–99). All estimates are strongly significantly different from zero. It is thus clear that returns to education among the wage employed remain statistically significantly greater for the older than the younger group.

Among self-employed and agricultural workers, the changes in the education coefficient are somewhat larger, but in most cases the hypothesis holds that the education coefficient is constant across the two time periods. However, there are two conspicuous differences, both for old women. First, for self-employed old women, there is a statistically significant change in the estimated education coefficient, which falls from 0.17 to 0.06. Second, for old female agricultural workers, the education coefficient falls from 0.19 to zero; again, this difference is statistically significant. Exogenous events like rainfall probably have a larger impact on earnings among agricultural workers and the self-employed than on the earnings of the wage employed, which could partly explain the changes in the estimated returns for the former two occupations. However, it remains unclear why the returns to education for men would be less sensitive to such events than the returns for women. Also, looking at the results for the young, there is no uniform pattern in the change of returns. This issue deserves further investigation.[18]

Taking stock of these findings, it is notable that the returns for men remain lower than the returns for women. Only in the case of old agricultural workers is the return higher for men than for women. Nevertheless, women actually have much lower *levels* of earnings than men in Pakistan. In other words, although the slope of the education-earnings relationship is steeper for women than for men, the intercept of the wage regression is much higher for men.

Extensions on the education-earnings relationship

Correcting returns estimates for endogeneity bias

Multinomial logit equations were used to calculate selectivity terms for the 2001–02 data, which were statistically significant in 5 out of 12 earnings regressions. The introduction of the selection term generally reduces the return to education (with the exception of old women in agriculture, where the estimate of the education coefficient goes from zero without selectivity correction to 0.26 with selectivity correction). Thus the consequences of correcting for sample selection for this data set are very similar to the consequences of correcting for the 1998–99 data. Since selectivity correction makes a difference in some cases, this paper prefers the selectivity corrected equations to OLS.

[18] Results that are not reported here indicate that controlling for the log of the capital stock has only marginal effects (about one percentage point or less) on the coefficients on education for three out of the four self-employed age-gender categories considered in this analysis. The exception is old women, where the coefficient falls from 0.056 to zero. The coefficient on log capital is always statistically significant and varies between 0.10 and 0.11, except for old women, where it is equal to 0.30. For agriculture, the coefficient on education falls by 0.02 for young men, by 0.06 for young women, and by 0.05 for old men (there is virtually no effect for old women). The coefficient on log land is always significant and varies between 0.45 and 0.50, except for young women, for whom it is equal to 0.18.

The problem of endogenous sample selection is akin to the problem of endogeneity bias, as discussed previously. Allowing for household fixed effects in estimating the earnings function for the wage employed is an alternative way of addressing the endogeneity problem.[19] The fixed effects results for the 2001–02 data indicate lower returns to education than the OLS estimates. This is consistent with the selectivity corrected estimates and is exactly the same result found when selectivity correction was performed on the 1998–99 data. In fact, the fixed effects results for 2001–02 are very similar to those for 1998–99. It is possible, of course, that the reduction in estimated returns to education, compared with the OLS results, may be driven by measurement errors bias (see earlier discussion of this issue in the section on 1998–99 data). For this reason, and because the household fixed effects results can be estimated only for the subsample of wage employed persons, this paper continues to prefer selectivity corrected results.

Shape of the education-earnings relationship

Relaxing the linear relationship between education and earnings, quadratic terms were introduced in education for the 2001–02 data. The results show that the education-earnings relationship is convex for both old and young men in wage employment, as well as for young men and old women in self-employment. The relationship is significantly concave for young and old women in agriculture. For the other groups, a linear relationship cannot be ruled out.

---

[19] Two-stage least squares results for wage employees in the 2001–02 survey can be summarized as follows. Young men: using father's and mother's education as instruments, and losing about 40 percent of observations in the process (see footnote 4), the coefficient on education rises from 0.035 to 0.059 (significant at the 1-percent level), and the validity of the over-identifying restrictions is rejected at the 5-percent level; adding spouse's education to the instrument is not feasible as we 80 percent of observations would be lost; using spouse's education as the only instrument, 60 percent of observations are lost and the coefficient rises to 0.075 (significant at the 1=percent level).

Young women: using father's and mother's education as instruments, about 60 percent of observations are lost and the coefficient on education falls from 0.144 to 0.129 (significant at the 1-percent level), and the validity of the over-identifying restrictions is accepted at the 10-percent level; adding spouse's education to the instrument is not feasible, as 99 percent of the observations would be lost; using spouse's education as the only instrument, 55 percent of observations are lost and the coefficient rises to 0.18 (significant at the 1-percent level).

Old men: parental education cannot be used as an instrument, as too few individuals in this age group live with their parents; using spouse's education as the only instrument, 10 percent of observations are lost and the coefficient rises from 0.070 to 0.105 (significant at the 1-percent level). Old women: parental education cannot be used as an instrument, as too few individuals in this age group live with their parents; using spouse's education as the only instrument, 25 percent of observations are lost and the coefficient rises from 0.183 to 0.192 (significant at the 1-percent level).

With respect to other nonlinearities, the marginal returns to education are generally substantially lower for men than for women in both wage employment and self-employment, although not in agriculture.[20] Marginal returns are also generally higher for the older age group than for the younger one. Data of the later survey again suggest that the Pakistan labor market is not generally characterized by the commonly assumed concave relationship, which implies diminishing returns to extra years of schooling.

Earnings and cognitive skills

When OLS estimates for 2001–02 data are corrected for selectivity, the results indicate positive, and often high, returns to literacy in all cases except for women in agriculture (where returns are insignificant). In wage employment, and among young self-employed individuals, the returns to literacy are much larger for women than for men. This finding is similar to that found based on the 1998–99 data and could be due to a scarcity premium, since far fewer women than men are literate in Pakistan.

The results on numeracy skills are quite mixed. In four cases, the estimated coefficient on numeracy skills is actually negative and significant, suggesting that these skills *reduce* earnings. This counterintuitive result is due to the fact that numeracy and literacy skills are highly correlated. If the literacy variable is excluded, the coefficient on numeracy skills tends to rise substantially. In other words, the relationship between numeracy skills and earnings *conditional on literacy skills* is sometimes negative and significant, whereas the unconditional relationship is usually positive and often large. This interpretation is probably the best way to view the effects of numeracy.

Heterogeneity in returns to education

Quantile regressions were used to examine heterogeneity in the returns to education. For women in wage employment, the results indicate that returns to education are highest in the lowest quantile of earnings (bottom quartile) and lowest in the highest earnings quantile (top quartile). In other words, women with lower ability have higher rates of return to education. This is true for women in both the young and old age groups. This finding suggests that for women waged workers, education is inequality reducing, since it reduces

[20] It should be noted, however, that the marginal returns to education in agriculture for women are very imprecisely estimated.

wage differences between low- and high-ability individuals. A very similar pattern was found for the 1998-99 data. The results are less clear for the other occupation categories.

Conclusions

The labor market benefits of education accrue both from the fact that education promotes a person's entry into lucrative occupations and, conditional on occupation, raises earnings. The findings from the two rounds of Pakistan data (1998–99 and 2001–02) are remarkably similar. Education is found to play a very important role in occupational outcomes, but this role differs greatly between genders. Both younger and older women begin to take advantage of the benefits of education in earnest only after about 10 years of schooling, when they start to join the labor force and enter wage employment. Among young men, the likelihood of wage employment is unresponsive to education level; young men also increasingly quit the labor force or queue unemployed as their education level increases.

Education is also found to consistently and substantially raise earnings, conditional on occupation. Again, however, this relationship varies greatly by gender. Young men have very low marginal returns to education, particularly at lower levels of education. Across occupations, women's returns to education tend to be much higher than those of men, reflecting at least in part a scarcity premium (far fewer women than men are educated in Pakistan). Yet this potentially positive factor for women is counterbalanced by the fact that overall, men's earnings are much higher than those of women at all levels of education, with the gap being particularly large among persons with little to no education. This latter finding highlights the case for policies that discourage gender discrimination by employers in the labor market.

Investigating the education-earnings relationship, the paper found—contrary to conventional wisdom—that the shape of this relationship in Pakistan is not concave, with diminishing returns to education. In wage employment for men and for some worker groups in other occupations, the relationship is convex. The implications of this finding are considerable. Extant education and labor market policy is predicated on the assumption that returns to education are greatest at the primary level and progressively lower at secondary and tertiary levels. The Millennium Development Goals also presume that the completion of basic education will help realize the goal of cutting world poverty in half by 2015. If, however, the relationship of education and earnings is convex (or even linear), then increasing education by small amounts at low education levels will not raise earnings substantially and thus will not prove an effective means of helping people climb out of poverty.

While the findings of the estimated returns to education along the earnings distribution are mixed, one clear pattern is discernible. Among both young and old women in wage employment, education is inequality reducing, that is, lower-ability women have higher returns to education than higher-ability women. Given that education is associated mainly with *wage* employment for Pakistani women (see figure A1.3A(ii)), the fact that it plays an inequality-reducing role is welcome news. This effect can be viewed as a non-market "externality" of women's education that further strengthens the case for public investment in girls' schooling.

Finally, the paper examined relationships between numeracy and literacy, on the one hand, and occupational outcomes and earnings, on the other. It found that cognitive skills have big payoffs for both women and men in Pakistan. In particular, literacy promotes entry into lucrative parts of the labor market for both men and women, although the size of the relationship is bigger for men. Conditional on occupation, literacy is also associated with substantially higher earnings in both wage employment and self-employment for both men and women, although in this case the size of the relationship is significantly bigger for women than for men.

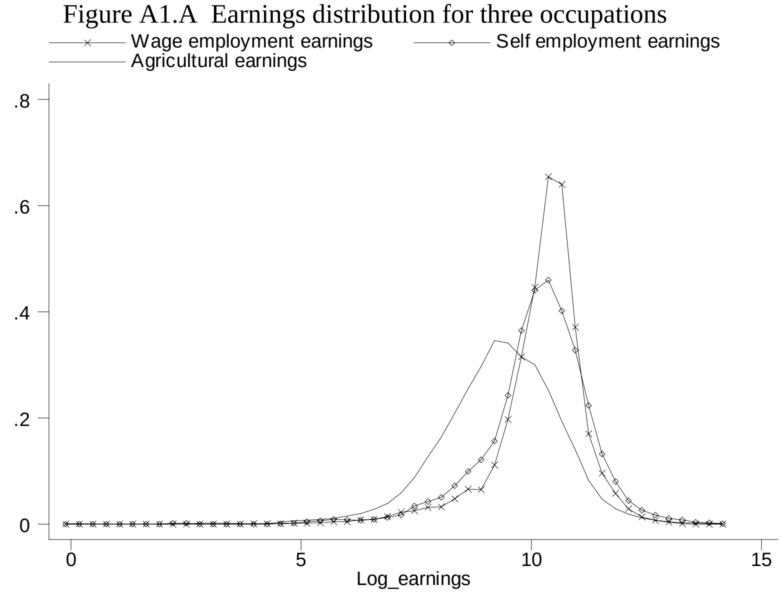
Appendix 1:

Data Analysis, 1998–99: Tables and Figures

Table A1.1
Full sample: Summary statistics by occupation (means and *medians*)

	All	**Self-employed**	**Agricult. employed**	**Wage employed**	**Unemp-loyed**	**Out of labor force**
Annual earnings	30277	36419	20674	35138	---	---
[*median*]	[*24125*]	[*28007*]	[*11681*]	[*31200*]		
Log earnings	9.78	9.91	9.29	10.08	---	---
[*median*]	[*10.09*]	[*10.24*]	[*9.37*]	[*10.35*]		
Years of education	3.35	4.58	2.46	5.41	4.83	2.39
Age	35.4	35.8	37.3	33.8	30.0	35.8
Proportion men	0.46	0.86	0.71	0.86	0.51	0.12
Math skills	0.61	0.78	0.58	0.76	0.68	0.52
Read & write skills	0.40	0.56	0.33	0.59	0.55	0.29
# children aged < 12 in household	2.63	2.68	2.76	2.49	2.33	2.66
# individuals aged > 65 in household	0.24	0.20	0.25	0.19	0.17	0.26
Proportion married	0.70	0.72	0.74	0.68	0.50	0.70
Observations	47804	3333	7066	11762	1413	24230
Earnings obs	22161	3333	7066	11762	0	0

Figure A1.A Earnings distribution for three occupations



Note: Earnings are measured in 1998/9 Pakistan rupees. The USD exchange rate over the sampling period is approximately 50. The figure shows kernel density estimates of the earnings distributions in agriculture, non-farm self-employment, and wage employment. Sampling weights were used for calculating means, but not for medians or kernel density estimates.

Table A1.2
Selected partial effects on the likelihood of occupational outcome, by gender and age group

	Young		Old	
	1. Men	2. Women	3. Men	4. Women
1. Self-employment				
# children aged < 12 in household	0.006	-0.001	0.004	-0.002
	(5.13)**	(1.22)	(3.08)**	(2.90)**
# individuals aged > 65 in household	-0.010	-0.005	-0.009	-0.006
	(1.43)	(1.94)+	(1.29)	(2.00)*
Individual is married	0.038	-0.009	0.021	-0.001
	(3.96)**	(5.36)**	(1.57)	(0.45)
2. Agriculture				
# children aged < 12 in household	0.008	0.001	0.006	0.002
	(5.37)**	(0.98)	(3.84)**	(2.17)*
# individuals aged > 65 in household	0.025	0.013	0.030	0.006
	(3.12)**	(2.89)**	(3.72)**	(1.15)
Individual is married	0.056	0.005	0.033	0.023
	(4.91)**	(0.80)	(2.14)*	(2.82)**
3. Wage employment				
# children aged < 12 in household	-0.019	-0.001	-0.010	-0.001
	(9.37)**	(1.42)	(5.61)**	(0.71)
# individuals aged > 65 in household	-0.011	-0.003	-0.026	-0.005
	(1.02)	(0.79)	(2.78)**	(1.20)
Individual is married	0.057	-0.051	0.093	-0.050
	(4.45)**	(14.57)**	(5.24)**	(10.66)**
4. Unemployed				
# children aged < 12 in household	0.001	-0.001	0.000	0.000
	(1.04)	(2.12)*	(0.01)	(0.65)
# individuals aged > 65 in household	0.002	-0.010	-0.001	-0.005
	(0.41)	(2.51)*	(0.30)	(1.65)+
Individual is married	-0.044	-0.002	-0.022	-0.003
	(11.39)**	(0.74)	(7.18)**	(0.83)
5. Out of labor force				
# children aged < 12 in household	0.003	0.003	0.000	0.000
	(2.38)**	(1.76)+	(0.01)	(0.30)
# individuals aged > 65 in household	-0.007	0.006	0.006	0.010
	(0.90)	(0.78)	(0.95)	(1.42)
Individual is married	-0.107	0.058	-0.125	0.032
	(21.98)**	(7.59)**	(15.55)**	(3.16)**

Note: These results are based on the multinomial logits reported in appendix 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

	Young		Old	
	1. Men	2. Women	3. Men	4. Women
1. Self-employment				
Can solve simple maths problem	0.028	-0.005	0.067	-0.001
	(2.18)**	(2.45)*	(5.95)**	(0.46)
Can read & write	0.020	-0.005	-0.002	-0.004
	(1.93)+	(2.09)*	(0.20)	(1.98)*
2. Agriculture				
Can solve simple maths problem	0.010	0.013	0.006	0.003
	(0.78)	(2.19)*	(0.60)	(0.59)
Can read & write	-0.110	-0.078	-0.167	-0.081
	(11.42)**	(25.77)**	(21.37)**	(29.38)**
3. Wage employment				
Can solve simple maths problem	-0.020	-0.003	-0.025	-0.003
	(1.14)	(0.47)	(1.90)+	(0.63)
Can read & write	0.017	0.031	0.119	0.041
	(1.15)	(4.05)**	(9.81)**	(4.80)**
4. Unemployed				
Can solve simple maths problem	0.010	0.001	-0.002	-0.006
	(0.95)	(0.26)	(0.64)	(2.00)*
Can read & write	0.030	0.014	0.009	0.008
	(3.16)**	(2.71)**	(1.99)*	(1.65)+
5. Out of labor force				
Can solve simple maths problem	-0.028	-0.005	-0.045	0.007
	(2.28)*	(0.59)	(5.94)**	(0.82)
Can read & write	0.042	0.038	0.041	0.036
	(3.53)**	(4.05)**	(4.69)**	(3.63)**

Note: These results are based on the multinomial logits reported in appendix 2.

Table A1.4
Earnings and years of schooling

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
A. Young						
Education	0.033	0.149	0.048	0.105	0.053	0.041
	(17.08)**	(20.02)**	(5.77)**	(3.39)**	(5.27)**	(1.17)
Age	0.165	0.021	0.043	0.130	0.152	0.331
	(6.31)**	(0.18)	(0.41)	(0.43)	(1.29)	(1.42)
Age squared	-0.002	0.001	0.000	-0.002	-0.001	-0.006
	(4.18)**	(0.24)	(0.08)	(0.30)	(0.56)	(1.28)
# Individuals	4844	732	1230	161	2027	973
B. Old						
Education	0.066	0.172	0.070	0.170	0.074	0.188
	(47.96)**	(28.99)**	(13.64)**	(6.92)**	(9.83)**	(4.07)**
Age	0.095	0.079	0.042	0.012	-0.019	0.016
	(11.98)**	(1.86)	(1.76)	(0.14)	(0.75)	(0.25)
Age squared	-0.001	-0.001	-0.001	0.000	0.000	-0.000
	(11.55)**	(1.68)	(2.10)*	(0.16)	(0.74)	(0.32)
# Individuals	5439	747	1783	159	2963	1103

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 A1.5
Earnings and years of schooling: Correcting for sample selection

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
A. Young						
Education	0.033	0.117	0.045	0.072	0.066	-0.157
	(16.94)**	(8.10)**	(5.33)**	(1.86)	(3.56)**	(1.23)
Selection term	-0.251	-0.600	-0.485	1.001	-1.086	2.595
	(2.12)*	(2.70)**	(1.81)	(1.41)	(2.51)*	(1.63)
# Individuals	4844	732	1230	161	2027	973
B. Old						
Education	0.038	0.145	0.071	0.180	0.067	0.257
	(14.59)**	(12.00)**	(13.74)**	(6.87)**	(4.04)**	(2.11)*
Selection term	-0.972	-0.484	0.271	-1.119	0.136	-0.634
	(11.97)**	(2.81)**	(1.41)	(1.12)	(0.48)	(0.61)
# Individuals	5439	747	1783	159	2963	1103

Note: Robust t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Age, age squared, and province dummy variables are included in all regressions.

Table A1.6
Earnings and years of schooling among the wage employed:
Controlling for household fixed effects

	Young Men	**Young Women**	**Old Men**	**Old Women**
Education	0.013	0.089	0.044	0.128
	(3.40)**	(13.70)**	(10.88)**	(18.37)**
# Individuals	4844	732	5439	747

Note: Absolute value of t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Age, age squared are included in all regressions.

Table A1.7
Earnings and years of schooling:
Correcting for sample selection, quadratic term included

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
A. Young						
Education	-0.005	0.100	0.005	0.126	0.056	-0.074
	(0.81)	(3.81)**	(0.14)	(1.67)	(1.69)	(0.53)
Education squared	0.003	0.002	0.003	-0.005	0.001	-0.021
	(6.61)**	(0.75)	(1.20)	(0.81)	(0.42)	(1.43)
Selection term	-0.078	-0.488	-0.754	1.242	-0.519	3.622
	(0.66)	(1.78)	(2.12)*	(1.60)	(1.23)	(2.09)*
# Individuals	4844	732	1230	161	2027	973
B. Old						
Education	0.012	0.231	0.039	0.034	0.029	0.337
	(2.81)**	(8.95)**	(1.19)	(0.40)	(1.34)	(2.22)*
Education squared	0.003	-0.009	0.002	0.011	0.006	-0.019
	(7.26)**	(3.48)**	(0.98)	(1.84)	(2.65)**	(0.89)
Selection term	-0.550	-1.115	-0.103	-0.861	-0.406	-0.286
	(5.24)**	(3.92)**	(0.24)	(0.85)	(1.17)	(0.26)
# Individuals	5439	747	1783	159	2963	1103

Note: Robust t-statistics in parentheses. $^{+}$ significant at 10% level; * significant at 5% level; ** significant at 1% level. Age, age squared, and province dummy variables are included in all regressions.

Table A1.8
Earnings and the level of schooling: Correcting for sample selection

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
A. Young						
Primary	0.096	0.388	0.123	0.855	0.224	-0.771
	(3.35)**	(2.06)*	(1.03)	(2.84)**	(1.75)	(1.66)
Middle school	0.175	0.447	0.258	2.304	0.431	-2.292
	(6.18)**	(2.17)*	(1.96)*	(3.95)**	(2.94)**	(2.26)*
Secondary	0.228	1.236	0.393	-1.206	0.697	-2.739
	(8.41)**	(8.05)**	(3.30)**	(2.01)*	(3.84)**	(2.21)*
Higher secondary	0.344	1.281	0.391	0.070	0.982	-3.603
	(9.90)**	(6.74)**	(2.48)*	(0.02)	(3.70)**	(1.25)
Tertiary	0.615	1.567	0.840	2.474	0.938	-5.726
	(16.91)**	(6.14)**	(4.41)**	(2.90)**	(2.33)*	(1.71)
Selection term	-0.127	-0.741	-0.685	0.458	-0.554	3.718
	(1.05)	(3.00)**	(2.02)*	(0.61)	(1.34)	(2.71)**
# Individuals	4844	732	1230	161	2027	973
B. Old						
Primary	0.179	0.600	0.102	0.851	0.257	0.897
	(8.94)**	(2.69)**	(1.13)	(1.76)	(2.97)**	(1.84)
Middle school	0.229	1.218	0.369	0.792	0.585	1.015
	(8.67)**	(7.77)**	(3.39)**	(1.57)	(4.22)**	(0.95)
Secondary	0.305	1.581	0.599	1.172	0.730	0.708
	(10.93)**	(12.29)**	(6.12)**	(2.36)*	(3.84)**	(0.40)
Higher secondary	0.469	1.470	1.008	3.110	0.791	-1.002
	(11.06)**	(9.68)**	(8.49)**	(1.75)	(2.33)*	(0.27)
Tertiary	0.674	1.486	1.074	3.253	1.982	---
	(13.50)**	(6.74)**	(8.84)**	(6.61)**	(5.00)**	---
Selection term	-0.867	-1.074	-0.126	-1.228	-0.256	0.245
	(8.66)**	(5.24)**	(0.36)	(1.19)	(0.79)	(0.25)
# Individuals	5439	747	1783	159	2963	1103

Note: Robust t-statistics in parentheses. * significant at 5% level; ** significant at 1% level. Age, age squared, and 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 = 1-5 years of education; middle school = 6-8 yrs; secondary = 9-10 yrs; higher secondary = 11-12 yrs; tertiary = 13+ years.

Table A1.9
Estimated return to an additional year of schooling, by level of education
(Using sample selectivity corrected earning function from Table 8A)

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
A. Young						
Primary	1.9 *	7.8 *	2.5	17.1 *	4.5	-15.4
Middle school	2.6 *	2.0	4.5	48.3 *	6.9	-50.7
Secondary	2.7 *	39.5 *	6.8	-175.5 *	13.3	-22.4
Higher secondary	5.8 *	2.3	-0.1	63.8	14.3	-43.2
Tertiary	9.0 *	9.5	15.0 *	80.1 *	-1.5	-70.8
B. Old						
Primary	3.6 *	12.0 *	2.0	17.0	5.1 *	17.9
Middle school	1.7 *	20.6 *	8.9 *	-2.0	10.9 *	3.9
Secondary	3.8 *	18.2 *	11.5 *	19.0	7.3	-15.4
Higher secondary	8.2 *	-5.6	20.5 *	96.9	3.1	-85.5
Tertiary	6.8 *	0.5	2.2	4.8	39.7 *	33.4

Note: The marginal return to a year of primary schooling is calculated as the coefficient on the primary school dummy variable divided by 5, since there are 5 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 6, 7 and 8); and so on for other levels of education. Only few women are in self-employment so sample sizes are very small, as seen in table A1.8.

* 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 below it. Among old men in self-employment, for instance, the return to each extra year of education at the middle level is significantly greater than the return to each extra year of education at the primary level and thus, 8.9 has a * by it, since in this case 8.9 is significantly higher than 2.0. Similarly, 20.5 is statistically significantly different from 11.5 (marginal return to higher secondary is significantly greater than that to secondary education) and hence 11.5 has a * by it. Men's returns are much more precisely determined due to larger sample sizes and thus, even seemingly small differences in marginal returns at different levels of education are significantly different from each other, e.g. in wage employment.

Table A1.10
Earnings, literacy and numeracy

	1. Wage employed		2. Self employed		3. Agriculture	
	Men	Women	Men	Women	Men	Women
A. Young						
Can solve simple	0.036	0.184	0.039	-0.433	0.339	0.077
maths problem	(1.06)	(1.13)	(0.28)	(1.35)	(2.48)*	(0.41)
Can read & write	0.216	1.393	0.371	1.053	0.271	0.209
	(7.17)**	(8.97)**	(3.34)**	(2.86)**	(2.23)*	(0.82)
Age	0.192	0.180	0.089	0.080	0.186	0.336
	(7.21)**	(1.39)	(0.82)	(0.26)	(1.57)	(1.43)
Age squared	-0.003	-0.002	-0.001	-0.001	-0.002	-0.006
	(4.93)**	(0.84)	(0.33)	(0.12)	(0.81)	(1.30)
# Individuals	4844	732	1230	161	2027	973
B. Old						
Can solve simple	0.076	0.047	0.132	0.208	0.341	0.356
maths problem	(3.22)**	(0.37)	(1.60)	(0.88)	(4.36)**	(2.34)*
Can read & write	0.486	1.901	0.454	1.285	0.251	0.445
	(22.65)**	(14.32)**	(6.86)**	(4.11)**	(3.26)**	(1.67)
Age	0.097	0.084	0.049	0.020	-0.017	0.016
	(11.21)**	(1.86)	(2.04)*	(0.22)	(0.65)	(0.25)
Age squared	-0.001	-0.001	-0.001	0.000	0.000	-0.000
	(11.11)**	(1.74)	(2.38)*	(0.04)	(0.59)	(0.33)
# Individuals	5439	747	1783	159	2963	1103

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

Table A1.11
Earnings, literacy and numeracy: Controlling for sample selection

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
A. Young						
Can solve simple	0.046	0.195	-0.025	-0.606	0.332	0.252
maths problem	(1.31)	(1.23)	(0.17)	(1.74)	(2.43)*	(1.06)
Can read & write	0.209	1.037	0.322	0.962	0.435	-0.995
	(6.94)**	(5.57)**	(2.83)**	(2.57)*	(2.66)**	(0.97)
Selection term	-0.255	-0.944	-0.669	1.041	-0.593	2.015
	(1.97)*	(3.82)**	(1.92)	(1.25)	(1.51)	(1.21)
# Individuals	4844	732	1230	161	2027	973
B. Old						
Can solve simple	0.106	0.084	0.228	0.264	0.335	0.354
maths problem	(4.28)**	(0.66)	(1.60)	(1.09)	(4.27)**	(2.32)*
Can read & write	0.352	1.536	0.450	1.379	0.389	0.655
	(10.44)**	(9.26)**	(6.78)**	(4.18)**	(2.49)*	(0.78)
Selection term	-0.620	-1.185	0.358	-1.009	-0.333	-0.291
	(5.05)**	(4.16)**	(0.83)	(0.97)	(1.00)	(0.26)
# Individuals	5439	747	1783	159	2963	1103

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

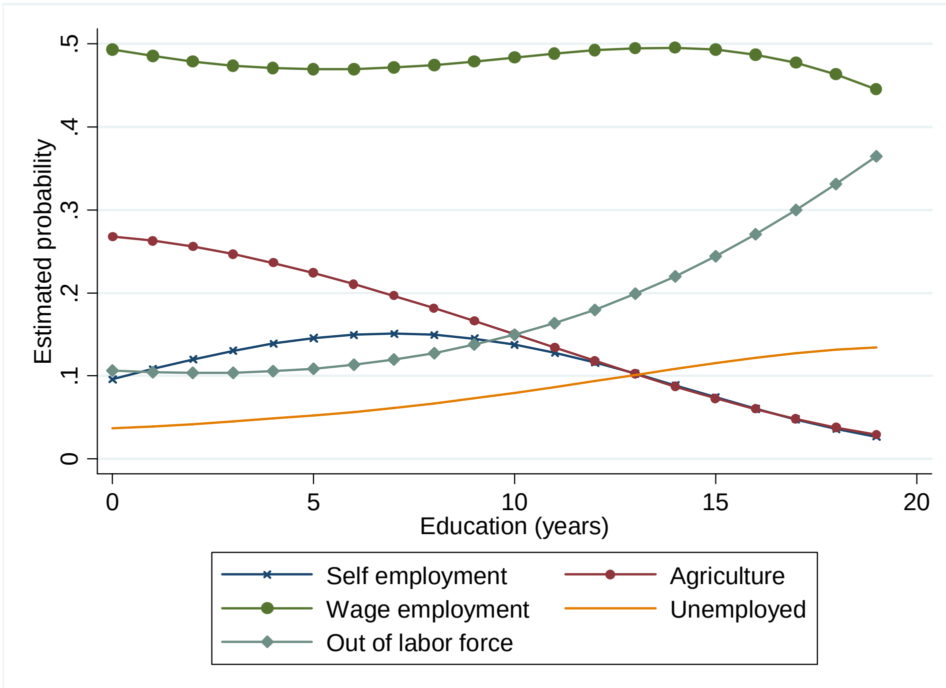
Table A1.12
Earnings and years of schooling: Quantile regressions

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
A. Young						
Education, P25	0.034	0.183	0.047	0.066	0.046	0.035
(low)	(14.06)**	(36.12)**	(5.03)**	(2.05)*	(3.84)**	(0.83)
Education, P50	0.031	0.162	0.041	0.090	0.043	0.031
(median)	(18.19)**	(44.54)**	(4.55)**	(3.07)**	(4.21)**	(0.87)
Education, P75	0.029	0.130	0.044	0.115	0.043	0.114
(high)	(13.20)**	(29.71)**	(4.58)**	(3.06)**	(4.71)**	(3.47)**
# Individuals	4844	732	1230	161	2027	973
B. Old						
Education, P25	0.061	0.213	0.056	0.175	0.066	0.134
(low)	(32.19)**	(37.51)**	(9.71)**	(8.09)**	(7.35)**	(2.48)*
Education, P50	0.056	0.170	0.064	0.190	0.064	0.133
(median)	(40.20)**	(44.39)**	(11.55)**	(7.68)**	(8.36)**	(2.83)**
Education, P75	0.061	0.125	0.072	0.178	0.066	0.190
(high)	(32.32)**	(26.20)**	(11.85)**	(7.01)**	(9.66)**	(5.33)**
# Individuals	5439	747	1783	159	2963	1103

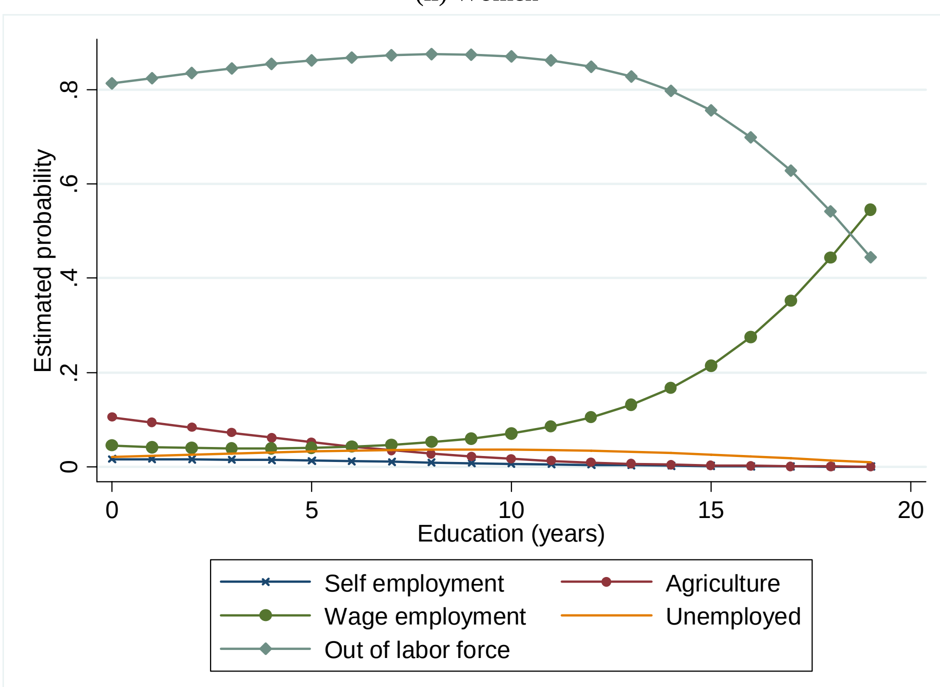
Note: Age, age squared, and province dummy variables are included in all regressions. Standard errors in parentheses.

Figure A1.1
Young individuals: Estimated probability of occupation and education

(i) Men



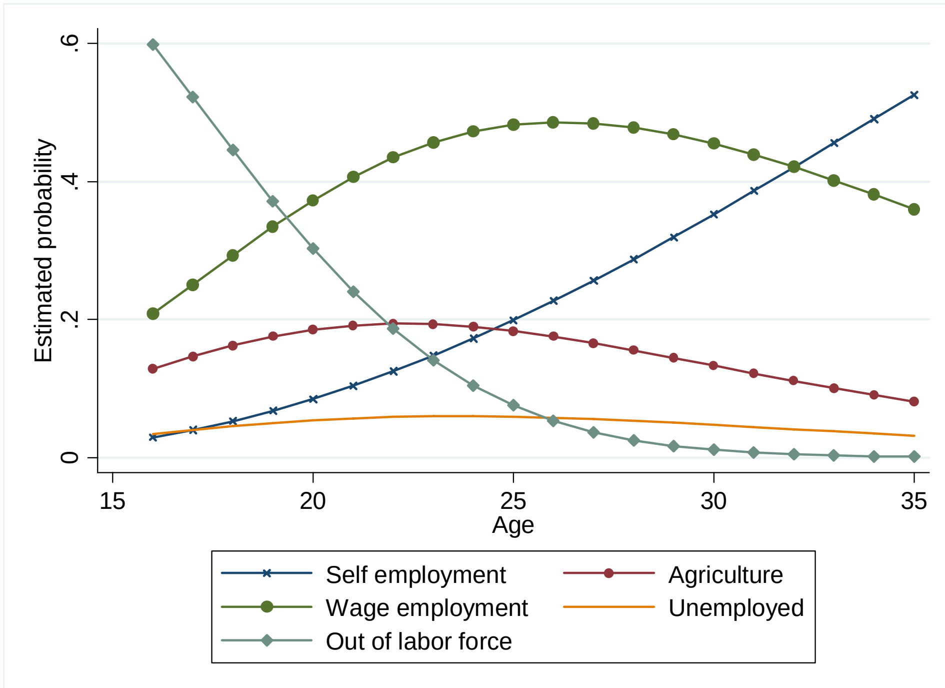
(ii) Women



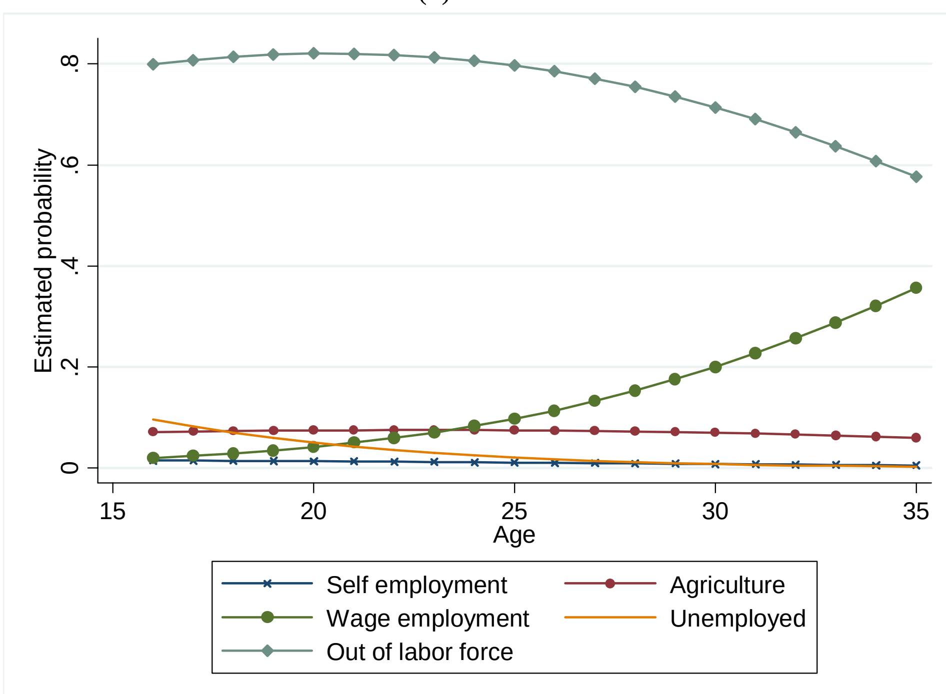
Note: These predictions are based on the multinomial logits reported in appendix 2.

Figure A1.2
Young individuals: Estimated probability of occupation and age

(i) Men



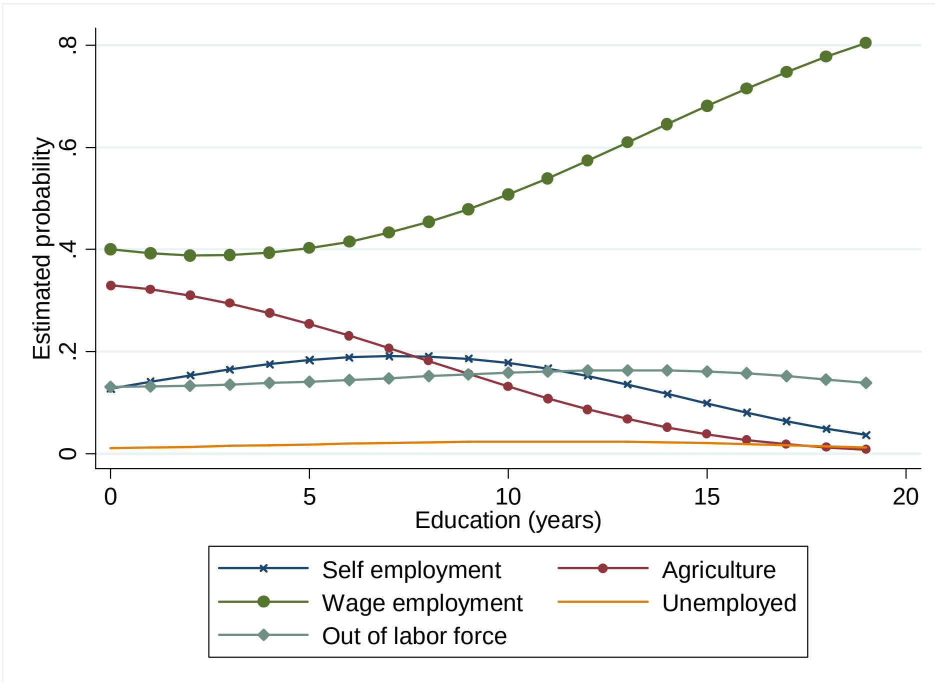
(ii) Women



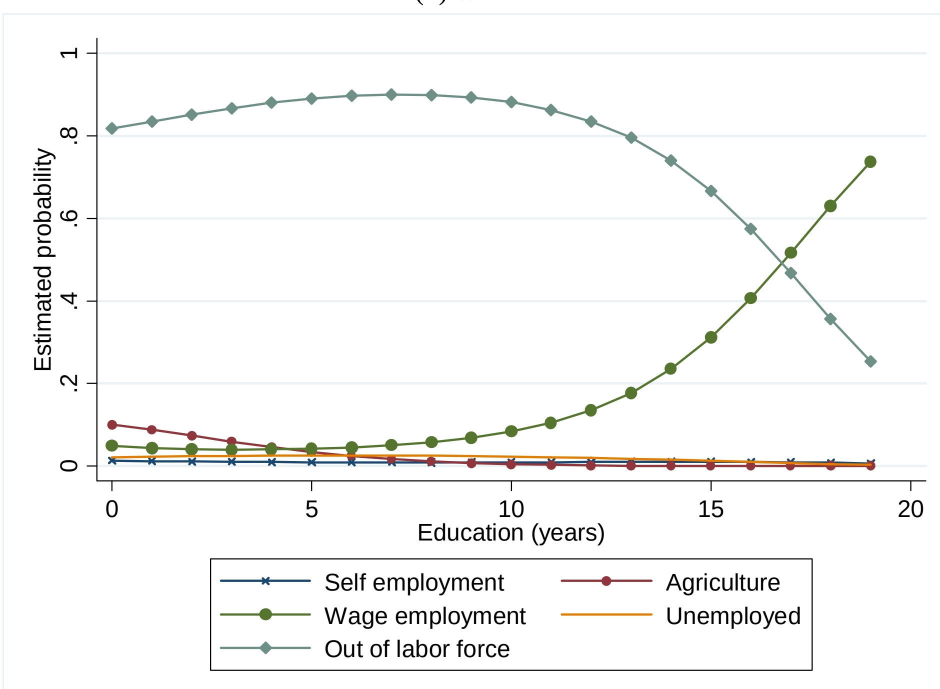
Note: These predictions are based on the multinomial logits reported in appendix 2.

Figure A1.3
Old individuals: Estimated probability of occupation and education

(i) Men



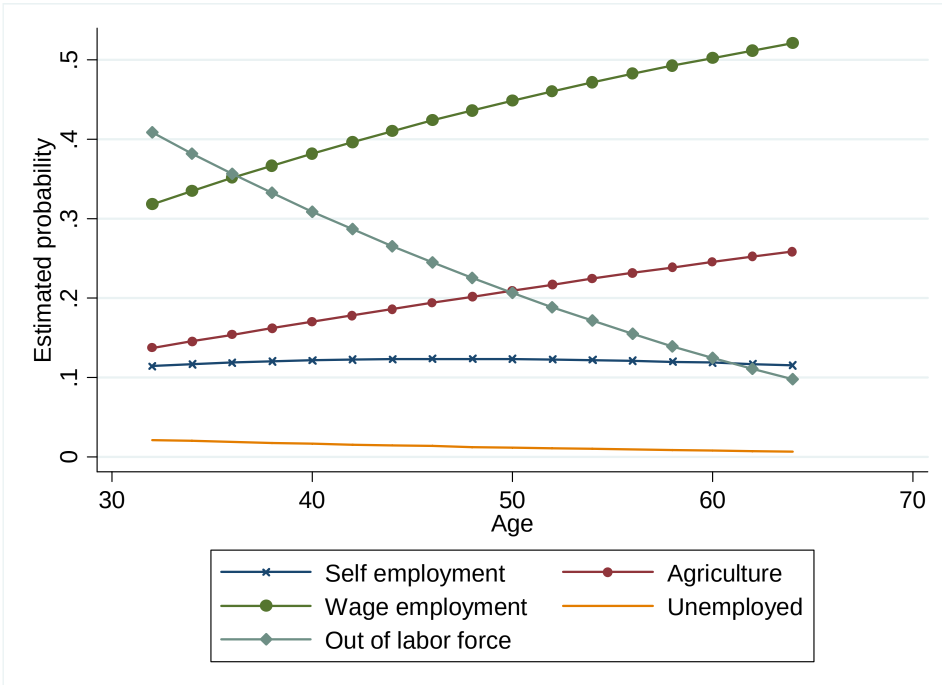
(ii) Women



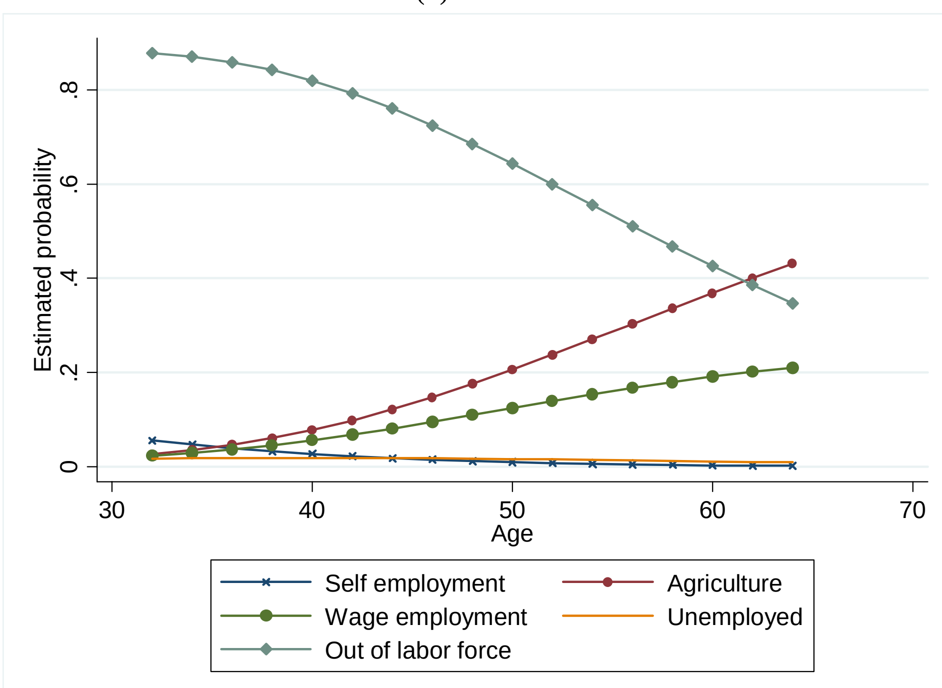
Note: These predictions are based on the multinomial logits reported in appendix 2.

Figure A1.4
Old individuals: Estimated probability of occupation and age

(i) Men

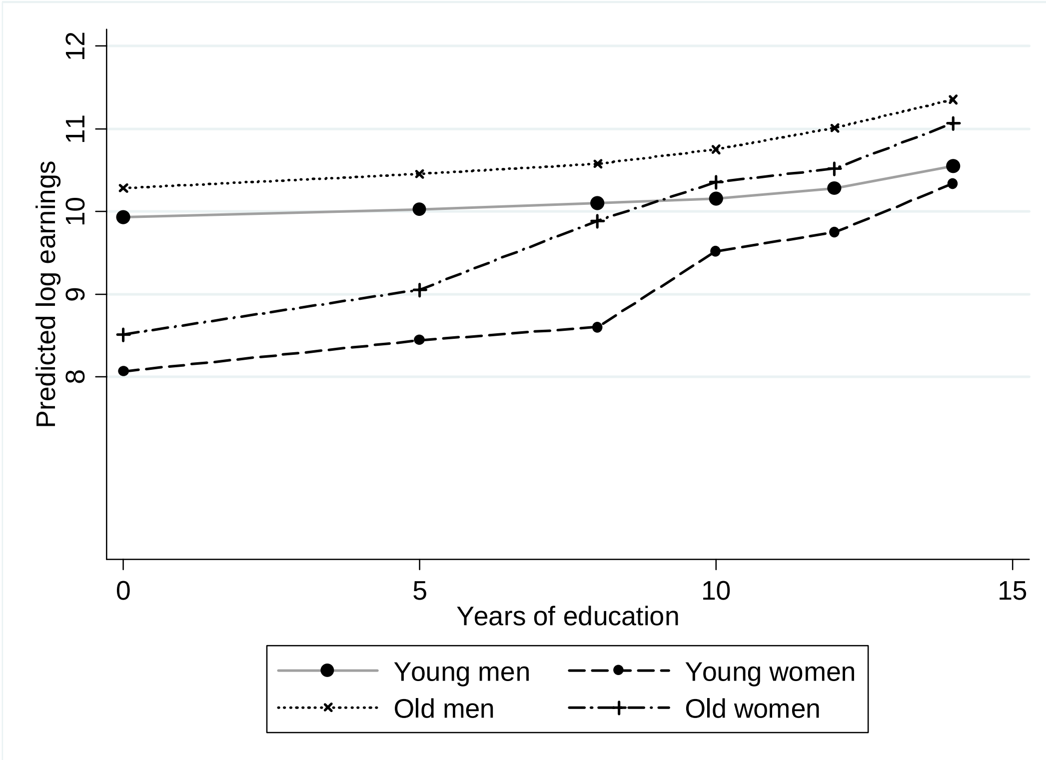


(ii) Women



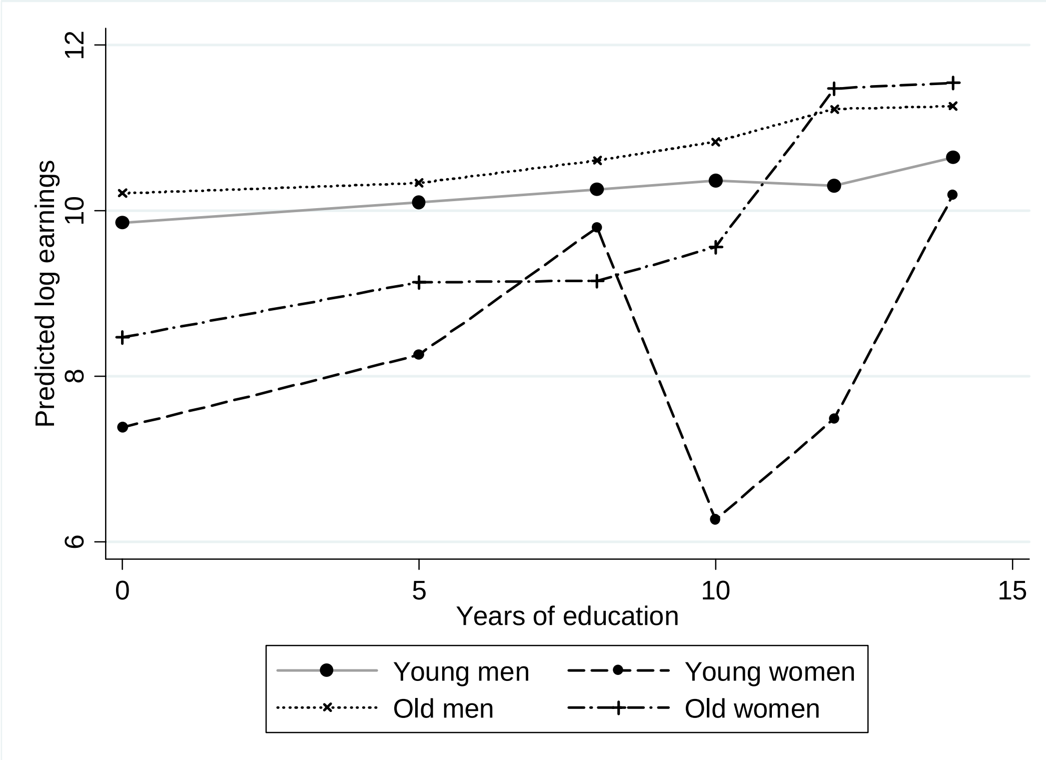
Note: These predictions are based on the multinomial logits reported in appendix 2.

Figure A1.5
Predicted earnings and level of education: Wage employed



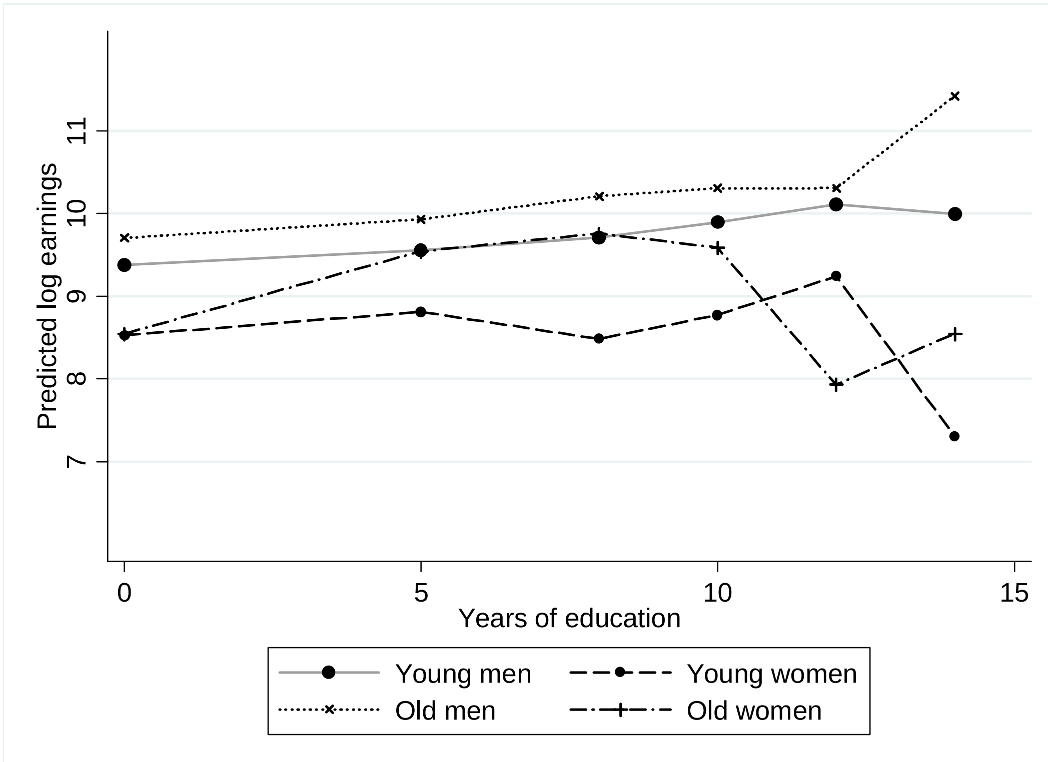
Note: These predictions are based on the results reported in table A2.10.

Figure A1.6
Predicted earnings and level of education: Self employed



Note: These predictions are based on the results reported in table A2.10.

Figure A1.7
Predicted earnings and level of education: Agriculture



Note: These predictions are based on the results reported in table A2.10.

Appendix 2:
Data Analysis, 1998–99: Selectivity Corrected Tables

Table A2.1
Multinomial logit estimates. Omitted category: Wage employment. Young men.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	0.151	0.005	0.081	-0.008
	(6.56)**	(0.27)	(2.72)**	(0.38)
Education squared	-0.011	-0.006	0.000	0.005
	(6.15)**	(3.77)**	(0.07)	(2.96)**
Age	0.152	-0.052	-0.023	-0.413
	(1.64)	(0.69)	(0.18)	(4.33)**
Age squared	-0.004	-0.000	-0.001	0.006
	(1.78)	(0.03)	(0.47)	(2.96)**
# of children in hh	0.093	0.082	0.058	0.068
under 12 years of age	(7.35)**	(7.67)**	(3.24)**	(5.01)**
# of elderly in hh over	-0.061	0.154	0.054	-0.030
65 years of age	(0.85)	(2.71)**	(0.56)	(0.40)
Married	0.177	0.153	-1.006	-1.269
	(2.17)*	(2.22)*	(7.66)**	(11.71)**
Observations	10004	10004	10004	10004

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

Table A2.2
Multinomial logit estimates. Omitted category: Wage employment. Young women.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	0.095	-0.013	0.217	0.110
	(1.30)	(0.26)	(4.52)**	(4.08)**
Education squared	-0.024	-0.023	-0.020	-0.015
	(3.30)**	(4.57)**	(5.26)**	(7.41)**
Age	-0.224	-0.183	-0.369	-0.189
	(0.92)	(1.27)	(1.97)*	(1.68)
Age squared	0.003	0.002	0.006	0.002
	(0.58)	(0.73)	(1.52)	(0.98)
# of children in hh	-0.027	0.038	-0.033	0.028
under 12 years of age	(0.59)	(1.71)	(1.02)	(1.54)
# of elderly in hh over	-0.351	0.255	-0.307	0.077
65 years of age	(1.59)	(2.38)*	(1.85)	(0.89)
Married	0.148	0.950	0.766	0.946
	(0.67)	(7.26)**	(4.57)**	(9.66)**
Observations	12765	12765	12765	12765

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

Table A2.3
Multinomial logit estimates. Omitted category: Wage employment. Old men.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	0.136	0.011	0.145	0.038
	(7.79)**	(0.60)	(3.27)**	(1.85)
Education squared	-0.013	-0.013	-0.009	-0.004
	(9.56)**	(8.26)**	(2.99)**	(2.95)**
Age	-0.022	-0.004	-0.079	-0.135
	(0.82)	(0.17)	(1.17)	(4.06)**
Age squared	0.000	0.000	0.001	0.003
	(1.67)	(2.02)*	(1.79)	(7.88)**
# of children in hh	0.052	0.052	0.023	0.029
under 12 years of age	(4.52)**	(5.13)**	(0.76)	(2.18)*
# of elderly in hh over	0.005	0.208	0.019	0.143
65 years of age	(0.08)	(3.91)**	(0.12)	(1.98)*
Married	-0.107	-0.166	-1.210	-1.201
	(0.89)	(1.62)	(5.68)**	(11.32)**
Observations	12037	12037	12037	12037

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

Table A2.4
Multinomial logit estimates. Omitted category: Wage employment. Old women.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	-0.002	-0.001	0.183	0.148
	(0.02)	(0.01)	(2.77)**	(4.72)**
Education squared	-0.009	-0.038	-0.023	-0.019
	(1.39)	(3.24)**	(3.99)**	(8.40)**
Age	-0.236	0.025	-0.120	-0.144
	(2.55)*	(0.46)	(1.60)	(3.39)**
Age squared	0.002	-0.000	0.001	0.002
	(2.53)*	(0.17)	(1.79)	(4.30)**
# of children in hh	-0.118	0.041	-0.005	0.014
under 12 years of age	(2.47)*	(1.80)	(0.16)	(0.74)
# of elderly in hh over	-0.343	0.172	-0.141	0.114
65 years of age	(1.53)	(1.62)	(0.85)	(1.32)
Married	0.619	1.064	0.588	0.777
	(2.39)*	(7.25)**	(2.84)**	(7.40)**
Observations	12998	12998	12998	12998

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

Table A2.5
Multinomial logit estimates. Omitted category: Wage employment. Young men.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	0.295	0.093	0.215	-0.178
maths problems	(2.39)*	(1.07)	(1.11)	(1.39)
Can read & write	0.128	-0.548	0.549	0.343
	(1.30)	(7.24)**	(3.56)**	(3.10)**
Age	0.125	-0.104	0.053	-0.346
	(1.35)	(1.39)	(0.42)	(3.66)**
Age squared	-0.003	0.001	-0.003	0.005
	(1.59)	(0.52)	(0.91)	(2.50)*
# of children in hh	0.095	0.085	0.053	0.065
under 12 years of age	(7.56)**	(7.97)**	(2.98)**	(4.75)**
# of elderly in hh over	-0.058	0.159	0.056	-0.028
65 years of age	(0.81)	(2.80)**	(0.58)	(0.38)
Married	0.196	0.186	-1.067	-1.331
	(2.43)*	(2.71)**	(8.15)**	(12.31)**
Observations	10004	10004	10004	10004

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

Table A2.6
Multinomial logit estimates. Omitted category: Wage employment. Young women.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	-0.372	0.237	0.095	0.049
maths problems	(1.50)	(1.68)	(0.46)	(0.41)
Can read & write	-0.911	-1.887	0.005	-0.487
	(3.53)**	(12.35)**	(0.03)	(4.29)**
Age	-0.479	-0.429	-0.556	-0.371
	(1.99)*	(3.02)**	(3.01)**	(3.38)**
Age squared	0.008	0.007	0.010	0.006
	(1.50)	(2.26)*	(2.39)*	(2.42)*
# of children in hh	0.003	0.066	-0.007	0.052
under 12 years of age	(0.08)	(2.96)**	(0.20)	(2.85)**
# of elderly in hh over	-0.376	0.226	-0.328	0.055
65 years of age	(1.70)	(2.14)*	(1.98)*	(0.65)
Married	0.305	1.131	0.917	1.074
	(1.38)	(8.77)**	(5.53)**	(11.19)**
Observations	12765	12765	12765	12765

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

Table A2.7
Multinomial logit estimates. Omitted category: Wage employment. Old men.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	0.566	0.065	-0.116	-0.359
maths problems	(6.28)**	(0.94)	(0.46)	(3.66)**
Can read & write	-0.295	-1.002	0.296	0.041
	(3.97)**	(15.29)**	(1.31)	(0.44)
Age	-0.023	-0.005	-0.078	-0.132
	(0.84)	(0.21)	(1.15)	(3.98)**
Age squared	0.000	0.001	0.001	0.003
	(1.74)	(2.21)*	(1.75)	(7.80)**
# of children in hh	0.061	0.063	0.026	0.034
under 12 years of age	(5.33)**	(6.37)**	(0.84)	(2.54)*
# of elderly in hh over	-0.023	0.155	0.013	0.124
65 years of age	(0.37)	(2.98)**	(0.08)	(1.72)
Married	-0.123	-0.179	-1.194	-1.203
	(1.04)	(1.76)	(5.62)**	(11.40)**
Observations	12037	12037	12037	12037

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

Table A2.8
Multinomial logit estimates. Omitted category: Wage employment. Old women.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	-0.025	0.102	-0.210	0.068
maths problems	(0.12)	(0.84)	(1.14)	(0.68)
Can read & write	-1.007	-2.304	-0.254	-0.565
	(3.71)**	(12.26)**	(1.20)	(5.19)**
Age	-0.220	0.043	-0.097	-0.126
	(2.39)*	(0.81)	(1.30)	(3.03)**
Age squared	0.002	-0.000	0.001	0.002
	(2.38)*	(0.46)	(1.53)	(4.02)**
# of children in hh	-0.101	0.062	0.019	0.034
under 12 years of age	(2.12)*	(2.76)**	(0.58)	(1.85)
# of elderly in hh over	-0.366	0.144	-0.172	0.086
65 years of age	(1.63)	(1.37)	(1.04)	(1.01)
Married	0.587	1.022	0.554	0.741
	(2.28)*	(7.03)**	(2.69)**	(7.21)**
Observations	12998	12998	12998	12998

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

Table A2.9
Earnings and years of schooling, Quadratic term included: OLS estimates

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
A. Young						
Education	-0.006	0.085	0.054	0.117	0.053	0.109
	(0.99)	(3.12)**	(2.12)*	(1.56)	(1.58)	(1.00)
Education squared	0.003	0.005	-0.001	-0.001	0.000	-0.009
	(7.09)**	(2.53)*	(0.26)	(0.17)	(0.02)	(0.66)
Age	0.165	0.017	0.046	0.134	0.152	0.340
	(6.30)**	(0.14)	(0.43)	(0.45)	(1.27)	(1.45)
Age squared	-0.002	0.001	0.000	-0.002	-0.001	-0.006
	(4.18)**	(0.23)	(0.06)	(0.31)	(0.56)	(1.31)
# Individuals	4844	732	1230	161	2027	973
B. Old						
Education	0.008	0.177	0.047	0.015	0.025	0.316
	(1.96)	(8.01)**	(3.12)**	(0.19)	(1.15)	(2.39)*
Education squared	0.004	-0.000	0.002	0.012	0.005	-0.020
	(14.46)**	(0.28)	(1.69)	(2.00)*	(2.43)*	(1.03)
Age	0.095	0.079	0.042	-0.003	-0.020	0.011
	(12.14)**	(1.86)	(1.80)	(0.03)	(0.76)	(0.17)
Age squared	-0.001	-0.001	-0.001	0.000	0.000	-0.000
	(11.80)**	(1.68)	(2.16)*	(0.32)	(0.74)	(0.24)
# Individuals	5439	747	1783	159	2963	1103

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 A2.10
Earnings and years of schooling among the wage employed:
Quadratic specification, with household fixed effects

	Young Men	Young Women	Old Men	Old Women
Education	-0.020	0.017	0.022	0.140
	(1.88)	(0.78)	(1.98)*	(5.11)**
Education squared	0.003	0.006	0.002	-0.000
	(3.42)**	(3.52)**	(2.41)*	(0.26)
# Individuals	4844	732	5439	747

Note: Absolute value of t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Age, age squared are included in all regressions.

Table A2.11
Earnings and the level of schooling, OLS estimates

	1. Wage employed		**2. Self employed**		**3. Agriculture**	
	Men	Women	Men	Women	Men	Women
A. Young						
Primary	0.091	0.377	0.242	0.874	0.180	0.284
	(3.28)**	(1.97)*	(2.33)*	(2.92)**	(1.44)	(1.12)
Middle school	0.170	0.538	0.401	2.412	0.335	-0.045
	(6.09)**	(2.57)*	(3.60)**	(4.27)**	(2.58)**	(0.08)
Secondary	0.226	1.452	0.505	-1.112	0.522	0.246
	(8.37)**	(11.31)**	(4.78)**	(1.91)	(4.13)**	(0.43)
Higher secondary	0.345	1.683	0.447	0.101	0.736	0.713
	(9.94)**	(14.10)**	(2.88)**	(0.03)	(3.86)**	(0.30)
Tertiary	0.620	2.274	0.789	2.809	0.622	-1.223
	(17.11)**	(18.56)**	(4.18)**	(4.26)**	(1.91)	(0.42)
Age	0.170	-0.010	0.052	0.263	0.138	0.351
	(6.52)**	(0.08)	(0.49)	(0.88)	(1.15)	(1.49)
Age squared	-0.002	0.001	0.000	-0.005	-0.001	-0.007
	(4.48)**	(0.47)	(0.01)	(0.79)	(0.43)	(1.36)
# Individuals	4844	732	1230	161	2027	973
B. Old						
Primary	0.173	0.540	0.124	0.665	0.227	0.995
	(8.63)**	(2.36)*	(1.81)	(1.45)	(2.92)**	(3.83)**
Middle school	0.293	1.376	0.396	0.682	0.507	1.211
	(11.63)**	(8.84)**	(5.00)**	(1.37)	(5.13)**	(1.65)
Secondary	0.468	1.841	0.623	1.092	0.604	1.041
	(22.55)**	(15.43)**	(8.48)**	(2.22)*	(5.65)**	(0.92)
Higher secondary	0.729	2.006	1.017	3.005	0.605	-0.616
	(23.47)**	(17.72)**	(8.82)**	(1.69)	(2.49)*	(0.18)
Tertiary	1.070	2.554	1.051	3.071	1.720	
	(45.33)**	(28.23)**	(9.86)**	(6.58)**	(8.02)**	
Age	0.097	0.085	0.041	0.014	-0.017	0.008
	(12.20)**	(2.01)*	(1.74)	(0.16)	(0.65)	(0.11)
Age squared	-0.001	-0.001	-0.001	0.000	0.000	-0.000
	(11.82)**	(1.83)	(2.09)*	(0.14)	(0.63)	(0.19)
# Individuals	5439	747	1783	159	2963	1103

Note: Robust t-statistics in parentheses. * 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 = 1-5 years of education; middle school = 6-8 yrs; secondary = 9-10 yrs; higher secondary = 11-12 yrs; tertiary = 13+ years.

Table A2.12
Earnings and the level of schooling among the wage employed: Controlling for household fixed effects

	Young Men	Young Women	Old Men	Old Women
Primary	0.042	0.012	0.215	0.249
	(0.90)	(0.10)	(3.99)**	(1.27)
Middle school	0.057	0.121	0.248	0.743
	(1.16)	(0.66)	(3.78)**	(2.56)*
Secondary	0.051	0.782	0.404	1.602
	(1.02)	(6.79)**	(7.04)**	(10.86)**
Higher secondary	0.161	1.090	0.542	1.666
	(2.48)*	(9.74)**	(6.20)**	(9.67)**
Tertiary	0.291	1.418	0.714	1.883
	(4.04)**	(12.27)**	(10.71)**	(16.03)**
# Individuals	4844	732	5439	747

Note: Absolute value of t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Age and age squared are controlled for, but the coefficients are not reported in order to conserve space. The omitted education category is no education. See notes to table A1.5 for information on how the education categories are defined.

Table A2.13
Earnings, literacy and numeracy among the wage employed:
With controls for household fixed effects

	Young Men	Young Women	Old Men	Old Women
Can solve simple maths problem	0.120	0.074	0.113	0.104
	(1.79)	(0.68)	(1.80)	(1.14)
Can read & write	-0.044	0.684	0.208	1.151
	(0.78)	(6.67)**	(3.76)**	(10.87)**
# Individuals	4844	732	5439	747

Note: Absolute value of t-statistics in parentheses. + significant at 10% level; * significant at 5% level; ** significant at 1% level. Age, age squared are included in all regressions.

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." *The Journal of Human Resources* 36, no. 2: 304–326.

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, Massachusetts.

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, no. 2: 405–458.

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.

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

Hanushek, E., and L. Woessmann. 2007. "The Role of School Improvement in Economic Development." Working Paper, no. 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, no. WP12664. NBER, Cambridge, Massachusetts.

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

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): 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.

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.

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

This paper investigates the education-earnings relationship in Pakistan, drawing on the Pakistan Integrated Household Surveys 1998–99 and 2001–02. The analysis has three main goals: to examine the labor market returns to education among waged, self-employed, and agricultural workers; to examine the labor market returns to the literacy and numeracy skills for these categories of workers; and to analyze the pattern of returns to education along the earnings distribution. The shape of the education-earnings relationship is also investigated. The analysis is conducted separately by gender and age group and attempts to address the usual biases when esti mating returns to education. Finally, the paper examines how key results have changed between the 1998–99 and 2001–02 surveys.