Education, skills, and labour market outcomes: evidence from Pakistan

Kingdon, Geeta; Söderbom, Måns

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

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

Draft

Geeta Kingdon** and Måns Söderbom†

May 2007

Abstract

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 amongst wage-employed, self-employed and agricultural workers; to examine the labor market returns to literacy and numeracy skills for these categories of workers; and to analyze the pattern of returns to education along the earnings distribution. We also investigate the shape of the education-earnings relationship. The analysis is done separately by gender and age group, and attempts to address the usual biases when estimating returns to education. Finally, we investigate how key results have changed between 1998/99 and 2001/02.

---

* This is a background paper for an upcoming World Bank study on education labor market linkages. We 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 thank Alonso Sánchez for excellent research assistance.

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

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

Part I

Introduction

The 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 1998/99 and 2001/02. The analysis has three main goals: to examine the labor market returns to education amongst wage-employees, 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 we have data from two points in time, we also investigate how these returns have changed between 1998/99 and 2001/02.

The paper will examine returns to education not only amongst the wage-employed, but also amongst self-employed and agricultural workers. While wage employment has been the object of most existing analyses, it is typically a small and often shrinking part of the labour market in developing countries. The labour market benefits of education accrue both from education promoting a person's entry into the lucrative occupations and, conditional on occupation, by raising earnings. The objective is to ask whether education raises earnings within any given occupation and whether it also raises earnings indirectly via 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. We will estimate the rate of return to education by occupation and for different levels of education, the latter to see the shape of the education-earnings relationship. In estimating the returns to education, we will attempt to correct for selectivity and endogeneity biases.

The paper will also interrogate the role of cognitive skills in both occupational attainment and earnings determination. There is evidence 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 learnt at school. This literature is summarised in Hanushek (2005). He cites 3 US studies showing quite consistently that a one standard deviation increase in mathematics test performance at the end of high school in the US translates into 12 per cent higher annual earnings. Hanushek also cites three studies from the UK and Canada showing 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, i.e. in 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 study already exists for Pakistan, our data offer a number of advantages over the previously used data[1].

Finally, the paper will investigate the role of education along the earnings distribution. This will enable us to say whether the effect of education is to reduce or accentuate earnings inequality. The analysis is done separately by occupation, gender by age group.

The remainder of the paper is structured as follows. Part II provides details on our empirical framework, focusing on the specifications and the estimators adopted. We use exactly the same techniques and specifications in the analysis of the data from 1998/99 as for 2001/02, in order to ensure the results are comparable. Part III contains our analysis of the 1998/99 data, which is divided into a short section describing the data; a section investigating the role of education and skills in determining occupational outcomes (where we distinguish between wage employment, non-farm self employment, agriculture, unemployment, and out of the labour force); and a section examining the relationships between earnings and education and cognitive skills. Part IV contains our analysis of the 2001/02 data, following the same structure as in Part III. Part V concludes.

[1] The wage equation in the Pakistan study by Behrman et. al. (2002) uses 1989 data on 207 wage employees from 3 districts of Pakistan, though 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 as such may have better cognitive skills data than that available to us in the Pakistan Integrated Household Survey (PIHS, 1998-99). They find that cognitive skills have statistically significant pay-offs in the labor market. While the PIHS only provides self-reported measures on whether the respondent can 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) having much larger samples: our wage equations are fitted for about 5000 men and 700 women. Finally while Behrman et al focus on the total return to cognitive skills, they do not examine the possible role of skills in promoting entry into the lucrative parts of the labor market.

Part II
Analytical approach

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

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

where w_i is real earnings of individual *i*, $\mathbf{x_i}$ is a vector of worker characteristics excluding education, $\boldsymbol{\alpha}_{\mathbf{ag}}$ is a parameter vector, s_i is the years of education, $f_{ag}(\cdot)$ is the earnings-education profile, υ_i is a residual, and *a* and *g* denote age group and gender, respectively. The primary objective in this paper is to estimate the total returns to education, and the variables included in the $\mathbf{x_i}$ are selected accordingly. In particular, in estimating the earnings regressions we do not condition on variables that are determined by education, as conditioning on such variables would change the interpretation of the schooling effects. For example, it is likely that an important effect of education is to enable individuals to get high-wage jobs (e.g. managerial positions), get into certain high-wage sectors or firms, or to generate job security and thus work experience. Consequently, we do not condition on occupation, firm-level variables, work experience, or other variables sometimes seen on the right-hand side in earnings regressions. We also do not condition on land in the agricultural earnings equation, or capital stock for the self-employed, because we assume investment in these assets may be driven by education. We acknowledge that this assumption may be strong, especially perhaps for the agricultural sector where land is often inherited (and where land may therefore drive education). We therefore discuss the effects on the results of including these asset variables in the regressions. We focus, however, on regressions that include only a small set of control variables, where age and gender are those emphasised the most. With respect to the effects of these variables on earnings, we allow for a fair deal of

flexibility and estimate all regressions separately for men and women, and separately for relatively young individuals (aged less than 30) and relatively old ones. Within each gender-age group, we include age as an additional control variable. We also include controls for province fixed effects.

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

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

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

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

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

Endogeneity bias

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

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

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

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

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

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

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

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

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

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

Empirical strategy

---

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

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

The paper will also estimate earnings functions by the quantile regression (QR) method. OLS regression models the mean of the conditional distribution of the dependent variable. However, if schooling affects the conditional distribution of the dependent variable differently at different points in the wage distribution, then quantile regressions are useful as they allow the contribution of schooling to vary along the distribution of the dependent variable. Thus, the estimation of returns to education using the QR method is more informative than merely being able to say that, on average, one more year of education results in a certain percent increase in earnings. Using quantile regressions we will investigate how wages vary with education at the 25^{th} (low), 50^{th} (median) and 75^{th} (high) percentiles of the distribution of earnings. To the extent that one is willing to interpret observations close to the 75^{th} percentile as indicative of higher 'ability' than at lower percentiles (on the grounds that such observations have atypically high wages, given their characteristics), the quantile regressions will thus be informative of the effect of education on earnings across individuals with varying ability[5].

[4] We are very limited in our ability to address the endogeneity problem by means of an instrumental variables approach, because few instruments are available in the data. We have information on parental education, but only for the sub-sample of individuals co-habiting with their parents at the time of the survey. We also have data on spouse's education, but only for the sub-sample of married individuals at the time of the survey. We have no data on the supply of education at a young age (Card, 1999). We have considered two-stage least squares results using parents' and spouse's education as instruments, but given the large (and potentially endogenous) gaps in the instruments data, and given that parental and spouse's education are dubious instruments (parents' education may not be a valid instrument since unobserved ability is probably inherited; spouse's education may not be a valid instrument since the unobserved ability of husband and wife is probably correlated), we have decided not to emphasize these results. We discuss them briefly in footnotes 10 and 16.

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

Part III
Results for 1998/99

In this part we undertake a detailed analysis of the PIHS 1998/99 data. We divide the analysis into the following sections. Section 1 provides details on the sample and shows summary statistics on key variables. Section 2 examines the effects of education and cognitive skills on occupational outcome. Section 3 analyses the effects of education and cognitive skills on earnings, conditional on occupational outcome. For facilitate comparison with the results for 2001/02, reported in Part IV below, all tables and figures associated with the present part of the analysis have a suffix "A".

1. Data and descriptive statistics[6]

The data come from the 1998-99 round of the Pakistan Integrated Household Survey (PIHS). Following a two-stage sampling strategy, the PIHS provides a nationally representative sample made up of around 16,000 households, which represent roughly 115,000 observations. The household questionnaire is composed of a number of detailed modules on such characteristics as income, education, health, maternity and family planning, consumption and expenses, housing conditions and available services. In addition, there are modules that concentrate on household enterprises and agricultural activities—including associated expenses and revenues. These modules enable us to define five categories of occupations: wage employment, non-farm self employment, agriculture, unemployment, and out of the labour force.

One important issue refers to the construction of the earnings variable. For individuals who are either unemployed or out of the labour force, we cannot construct a measure of earnings. For self-employed and agricultural workers we derive earnings from the specialized modules on household enterprises and agricultural activities respectively. A simple, yet comprehensive computation of recurring (non-durable) expenses and revenues—including produced or harvested goods consumed by the household—attributed to enterprise or agricultural endeavors is used to estimate earnings for these types of workers. The earnings of paid employees, however, are derived from the sum of reported income—cash, other occupations, in kind, pensions, etc.—from the income module.

[6] We are most grateful to Alonso Sánchez for providing substantial input to this sub-section.

Table 1A shows summary statistics for selected variables used in the analysis, for the full sample and for the five occupation categories identified. Our sample consists of individuals aged between 16 and 70 not currently enrolled in school. Unemployed individuals are those who seek employment and are available for it while out of labor force (OLF) individuals are those who do not seek employment such as housewives and the retired. The labor force participation rate is about 51% and unemployment rate is 6%.

Table 1A shows that average earnings in the full sample are 30,277 Pakistan rupees, which corresponds to approximately USD 600. There are significant differences in average earnings across the three job categories for which a measure of earnings can be constructed (this is not possible for non-workers). Self-employed and wage-employed earn on average about 70% more than individuals working in the agricultural sector, and this is mirrored by a similar differential in education: the average years of education in agriculture is 2.5 whereas for the self-employed and the wage-employed average education is between 4.5 and 5.4 years. It is worth noting that the average level of education amongst OLF persons is similar to that for agricultural workers. The pattern for literacy and numeracy skills is similar: 55 percent or more of the 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 OLF persons, less than 35 percent can read and write and less than 60% have basic math skills. Finally, it is worth noting that although the mean of earnings for the self-employed exceeds mean earnings for the wage employed, this is not true for earnings in natural logarithms (where the numbers imply a 17% premium of wage employment compared to self-employment) or for median earnings. The reason is that the distribution of earnings differs across the sectors, as can be seen the lower panel of Table 1A.

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 on the one hand, and agricultural workers and OLF persons on the other. This suggests that skills matter a lot for which of these two broadly defined occupation groups individuals end up in. While unemployed individuals possess the mean skill levels of wage and self-employed persons, they clearly queue for suitable job opportunities in the labor market. We now investigate the correlates of occupational outcome more in detail.

2. Education and occupational attainment

As shown very clearly in Table 1A, average earnings vary dramatically between individuals that are either self-employed or wage-employed, and individuals that work in the agricultural sector. This table also shows that the average levels of education and skills vary substantially between these two groups. It therefore seems very likely that one channel by which education raises incomes is by enabling individuals to get a job in a high-earnings sector. In this section we look at the effects of education and skills on occupational outcome. As discussed above, we define five 'occupations' in the data: self-employment, agriculture, wage employment, unemployment and individuals out of the labour force (OLF). From a policy point of view, the link between skills and labor market outcomes amongst the relatively young deserves special attention. Accordingly, in what follows we will analyze labor market outcomes for the young age group (16-30 year olds) separately from that for the old age group (31 to 70 year olds).

To understand the role of skills and family background factors in this context, we model occupational outcome 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. We therefore report marginal effects and conduct graphical analysis based on the results, and relegate all the underlying regression results to Appendix 1. Whenever education is included as an explanatory variable, we exclude the literacy and numeracy variables, and vice versa. This is because these dimensions of skills are highly correlated, and we have no interest in documenting the effects of education conditional on literacy and numeracy skills or the other way around. We run all regressions separately for men and women.

We begin by modeling occupational outcomes for men and women and by age group (young and old), and use years of education as our measure of skills. Table 2A shows marginal effects for number of children, number of elderly people in the household and marital status. While this is not of central interest to us, it is perhaps worth noting that the number of children significantly reduces the likelihood that an individual is in wage-employment (which is highly paid) for men but somewhat surprisingly not for women. One possible reason for this is that wage-employment is a less flexible occupation (in terms of

working hours for example) than the other job categories considered. For men, being married strongly increases the likelihood of working and reduces the likelihood of being unemployed and of being 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 1A illustrates the estimated association between years of education and the predicted likelihoods of 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. It is quite clear that for men the likelihood of 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 non-worker (both unemployment and OLF) is *increasing* with education. One possible reason for this is that individuals with a lot of education are willing to wait for a good job opportunity before taking paid employment. The likelihood of self-employment is inverse u-shaped in education, peaking at about 8 years of education.

For women the picture is very different indeed. Women with up to about 8 years of education are very unlikely to work. As education increases to secondary level and beyond, 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 wage job is approximately 0.50. Correspondingly, until about 10-12 years of schooling, education has no relationship with labor force participation but after that participation rises sharply with education (the OLF curve falls sharply). It is thus very clear that education matters much more for women than men in terms of determining what type of occupation the individual ends up with.

Figure 2A plots the estimated occupation probabilities as a function of age again for young persons (aged 16-30), holding all other explanatory variables fixed at the sample mean values. This is informative of the nature of the transition from education to work. Perhaps the most interesting result here is that women enter into gainful employment relatively late, only after about age 25 or so. By contrast, between the ages of 15 and 25, men enter the labour force at a rapid rate so that by about age 25, almost all men are labour force participants (the OLF curve falls sharply between ages 15 and 25). The relationship between age and participation in wage employment is a striking inverted-U shape: up to about age 25, likelihood of wage employment increases with age but then the relationship 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. It is possible this is because young people can only enter self-employment once they have accumulated some savings.

Figures 3A and 4A show repeat the type of calculations illustrated in Figures 1A and 2A for older individuals only (aged 31-70). In Figure 3A, a striking difference regarding the role of education is apparent for men: amongst 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 (unemployed/OLF). By contrast, older men's likelihood of being wage employed is strongly responsive to education. Amongst older women the basic patterns are similar to those for the young.

Table 3A presents the marginal effects of basic literacy and numeracy on the likelihood of being in different labor market states. The descriptive statistics discussed earlier made clear that wage and self-employment are the well-paying parts of the labor market in Pakistan and that agriculture is not. Overall, Table 3A 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 in agriculture, and the effect is significantly higher for men than for women in both age groups. However, somewhat surprisingly, being literate is associated with significantly *increased* chances of both being OLF and being 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 less extent unemployed), OLF perhaps due to cultural norms or their greater efficiency in the production of home goods. There is a weak suggestion that literacy reduces both young and old women's entry into self-employment but promotes young men's entry into self-employment.

Somewhat surprisingly, numeracy is not related to the chances of being in wage employment, suggesting that many waged jobs are unskilled, not requiring numerate individuals. But numeracy has a high association with the chances of being in self-employment, for men. This could be either because numeracy promotes entry into self-employment (causation from being numerate to entering the self-employment occupation) or because people in self-employment end up becoming numerate i.e., numeracy is learnt 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 not an escape route from the OLF state for women. This could be due to cultural norms or due to the earnings rewards of numeracy differing for men and women. We turn to these in the next section.

Before we do that, it is worthy of notice that 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 self-employment very substantially for both the young and the old sample, in the young sample the relationship is significantly smaller (-11.0 points compared with -16.7 points for the male sample). Similarly the relationship between numeracy and the likelihood of self-employment is less than half in size for young men as for older men. The reduction the size of the relationships – when moving from the old to the young sample – is generally smaller for women than men.

3. Education and Earnings

3.1 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 we see from Table 1A, wage employment absorbs only about half of the total labor force. Half 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 our knowledge, this question has not been addressed for Pakistan. While in common with the literature we use the term 'returns to education', strictly speaking the coefficient on the Mincerian earnings function is simply the gross earnings premium from an extra year of education and is not the 'return' to education since it does not take the cost of education into account.

Table 4A presents basic OLS estimates of the Mincerian returns to education in Pakistan, by occupation, gender and age group. It shows that the returns to education are very

precisely determined, even in cases where sample sizes are very small. As will be 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 returns to education are invariably statistically significantly greater for the older group than for the young. 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 when our 'old' age group got their jobs, primary completers were in more scarce supply and 5 to 8 years' education may have been sufficient to obtain a white-collar job. Those who obtained such jobs remain in them today. However, due to the rapid expansion of the supply of educated persons, grade 5 to 8 completers among those 'young' (aged 16-30) 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, there is a compression of wages 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 the tasks performed by older persons with the same education levels.

Table 4A also shows that returns to education are significantly and substantially greater for women than men in all occupations and in both age groups (except among the young in agriculture)[7]. The fact that returns to education in *wage* employment in Pakistan are about three-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 which require (or which are largely reserved for) educated women, such as nursing and primary school teaching, which are predominantly female jobs. However, the reasons for the higher earnings premium for women than men in self-employment are less clear, even though the female premium over the male is not so high in self-employment as in wage employment. Returns to education are particularly low for young men in agriculture and in wage employment.

Interestingly, in this data, returns to education in agriculture are similar to those in other occupations, at least among the older age group. This is similar to the findings of

[7] When we do not divide the sample into young and old age groups and estimate pooled equations (not shown), the return to each extra year of schooling in *wage* employment is 5.3% for men and three times higher i.e. 16.0% for women, similar to the estimates by gender in Pakistan (Aslam, 2006) using PIHS 2001-02 data.

Gallacher (2000) who finds that in Argentina, returns to education in agriculture for farms of average size was equal to the returns to education in wage employment[8].

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 – which is 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 with strong economic incentives to acquire schooling. Given that Pakistan has one of the world's largest (if not the largest) gender gaps in school enrolment and in literacy, these strong labor market incentives can help to redress those gaps providing supply of schooling is ensured and credit constraints that may impede girls' enrolment are removed through, for instance, attendance contingent cash subsidies, as in Bangladesh which has virtually eliminated gender gaps in its secondary school enrolments partly with the help of a female school stipend program.

However, even though returns to education may be high for women, they actually have much lower earnings 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 the lower levels of education. This is clear from the graphs of predicted earnings in Figures 5A to 7A where although the slope of the education-earnings relationship is steeper for women, the intercept is far lower for women than 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 is due to potential discrimination in the labor market. Education of women helps to reduce that earnings gap, i.e. there is less gender discrimination among the educated in the Pakistan labor market. Thus, if Pakistan wishes to reduce its gender gaps in education by improving women's incentives to acquire education, it needs to not only improve school supply and ease credit constraints but also to reform labor market policies in ways that reduce gender-differentiated treatment by employers.

[8] 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 is a modernizing agricultural environment. [cite more recent literature on returns to education in agriculture from Huffman and from Appleton et al].

As discussed above, we have also estimated the earnings equations for the self-employed and agricultural workers adding controls for productive assets. In the case of the self-employed, we add the log of the capital stock value (defined as the replacement value of buildings, plant and equipment) per self-employed individual in the household, while for agricultural workers we add the log of acres of land per individual engaged in agricultural production in the household. In doing so, we move from estimating reduced form earnings equations towards estimating profit functions with controls for fixed inputs, which changes the interpretation of the results somewhat.

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 where it is equal to 0.27. For agriculture, the coefficient on education falls by less than 0.01 for young men and women, and by about a 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 where it is equal to 0.10.

How to interpret these results depends on the causal relationship between education and the productive assets. If on the one hand assets depend on education (e.g. because education raises the marginal product of land, and so educated farmers choose more land), then our 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 our results with controls for land suggest our earlier results are overestimates of 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.

3.2 Extensions on the education-earnings relationship

Correcting returns estimates for endogeneity bias

As stated in Part II, OLS estimates of returns to education potentially suffer from sample selectivity bias and endogeneity bias. We attempt to address the former by employing the Heckman procedure, explained in Part II. The multinomial logit equations in the Appendix tables were used to calculate the selectivity terms. The results are presented in Table 5A. The selectivity term is statistically significant in 5 out of 12 earnings regressions. The introduction of the selection term generally reduces the return to education and in 3 cases (waged young women and waged old men and women), this reduction is statistically significant. Since selectivity correction makes a difference in some cases, we prefer the selectivity corrected equations to OLS.

The problem of endogenous sample selection is akin to the problem of endogeneity bias, as discussed in Part II. We approach the endogeneity issue by estimating a household fixed effects earnings function for waged work. We cannot estimate this for self- and agricultural-employment since there is no within-household variation in these cases. The results in Table 6A yield similar results to those in Table 5A: returns to education fall compared with OLS returns in Table 4A, though they generally fall more than when correcting for selectivity bias in Table 5A[9]. The household fixed effects approach is a powerful way to address endogeneity since the identification of the effect of education on earnings comes only from within-family variation among members in earnings and in education, and as such it nets out the effect of shared ability, akin to the twin-differencing approach. However, the reduction in estimated returns to education in Table 6A compared with the OLS Table 4A may represent not only a correction for endogeneity (or 'ability') bias. It could also arise from measurement error bias which is exacerbated in differenced models and which downward biases coefficients. For this reason and because the household fixed effects results can be estimated only for the sub-sample of wage employed persons, we present the selectivity corrected results as our preferred estimates.[10]

[9] Appendix Table A9 presents household fixed effects estimates of the earnings function for wage workers with education *level* rather than years of education.

[10] We have also estimated the linear model for the wage employees, using two-stage least squares. Results can be summarized as follows: i) Young men: using father's and mother's education as instruments, and losing about 50% of the observations in the process (see footnote 4), the coefficient on education rises from 0.033 (OLS, see Table 4A) to 0.064 (significant at the 1% level), and the validity of the overidentifying restrictions is rejected at the 5% level; adding spouse's education to the instrument is not feasible as we would lose too many observations; using spouse's education as the only instrument, we lose 60% of the

Shape of the education-earnings relationship

What is the shape of the education-earnings relationship in different occupations? So far we have imposed a linear relationship between 'years of education' and earnings in Table 5A. Table 7A, 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 Appendices A9A and A10A respectively. Table 7A 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 and 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 labour market is not generally characterized by the commonly assumed concave relationship which implies diminishing returns to extra years of schooling.

The non-linearities of the education-earnings relationship are explored further in Table 8A which includes a dummy variable for each education level. The selectivity correction estimator is preferred. OLS yields significantly higher coefficients compared with selectivity corrected estimates in several cases and is relegated to Appendix A11A. The household fixed effects results for the wage employed are included in Appendix A12A. The base education category is 'no education'. The marginal return to each year of primary education, to each year of middle education and so forth, calculated from Table 8A, are set out in Table 9A. It confirms some patterns noted earlier. For instance, it shows that marginal returns to education are generally substantially lower for men than women in both wage and self-employment, though 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 schooling levels. Among young men in waged employment, marginal

observations, and the coefficient rises to 0.068 (significant at the 1% level). ii) Young women: using father's and mother's education as instruments, we lose about 60% of the observations, the coefficient on education falls from 0.149 (OLS, see Table 4A) to 0.137 (significant at the 1% level), and the validity of the overidentifying restrictions is accepted at the 10% level; adding spouse's education to the instrument is not feasible as we would lose too many observations; using spouse's education as the only instrument, we lose 60% of the observations, and the coefficient rises to 0.18 (significant at the 1% level). iii) 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, we lose 10% of the observations, and the coefficient rises from 0.066 (OLS; Table 4A) to 0.102 (significant at the 1% level). iv) 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, we lose 30% of the observations, and the coefficient rises from 0.172 (OLS; Table 4A) to 0.184 (significant at the 1% level).

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, among young and old waged women since their marginal return to education at the secondary schooling level is substantially higher than at the primary level. For women, returns estimates beyond secondary are typically insignificant as they are based on very small samples (few women have more than lower secondary education). The evidence of Tables 7 to 9 taken together suggests that the education-earnings relationship in Pakistan is not concave in any of the occupations, i.e. there is no evidence of diminishing marginal returns to education in Pakistan. This is confirmed in Figures 5A, 6A and 7A which show the relationship between education and predicted earnings.

Earnings and Cognitive Skills

Table 10A shows OLS and Table 11A shows selectivity corrected earnings functions by occupation with cognitive skills measures on the right hand side. The first set of columns numbered '1' are individual level earnings functions for waged workers, estimated separately for men and women. The next two columns are an earnings function at the household level, taking only those household members into account who were employed in a household self-employment enterprise (column 2) and in agricultural self-employment (column 3). Years of schooling is not included in the earnings functions. This is because we wish to estimate the total return to cognitive skills irrespective of whether they were acquired through schooling or not.[11] 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, we discuss the selectivity corrected results. Household fixed effects results are reported in Appendix A13A and they show smaller sized effects than those in Table 11A, which could be either because the fixed effects method provides a tighter upper bound on the

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

skill effect than a selectivity correction approach, or because of attenuation bias in the fixed effects equation due to heightened measurement error.

Table 11A shows strong returns to literacy in all occupations. In most cases, the returns to literacy are dramatically larger for women than men. This is at least partly due to a scarcity premium since far fewer women than men are literate. Fewer women than men have the years of schooling required to develop literacy skills, and women are likely to have attended poorer quality schools than men in Pakistan[12].

While literacy returns in waged work are only about one fifth as large for men as for women, they are nevertheless still substantial and statistically significant. Literacy has payoffs for men in agriculture: literate men are significantly more productive than illiterate men (for women the point estimate is large but not statistically significant). Young men's literacy return in agriculture is double that in wage employment.

Significant positive returns to *numeracy* skills accrue to both old men and women in agriculture. While they also accrue to old men in waged work, the size of the return is only one third as large as 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) does raise agricultural earnings. The lack of returns to skills in agriculture for women could arise if household males make farming decisions due to the gender division of roles in this traditional occupation.

The educational decisions of today's children will depend much more on the observed pattern of returns to education and to skills among the young adults than among 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 notion. Young men's return to education in agriculture is statistically significant at the 5% level only from middle schooling onwards, suggesting that

[12] Aslam and Kingdon (2006) show that girls receive significantly lower educational expenditures within the household than boys in Pakistan. Aslam (2007) finds that girls also face poorer quality schooling than boys in Pakistan as they are very significantly less likely to be sent to private schools than their brothers, combined with the fact that 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 et al, 2001; Andrabi et al, 2002; Arif and Saqib, 2003).

it takes 8 years of schooling to acquire cognitive skills. This highlights the importance of 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 and this has implications for the inequality-reducing role of education. However, the distribution of returns to education across the earnings spectrum is not known for Pakistan, as for most other developing countries. We examine heterogeneity in returns to education to ask whether some individuals benefit more from education than others and why, and the inequality implications of that.

There is now a literature investigating the pattern of returns to an additional year of education along the earnings distribution using quantile regression (QR) analysis. An examination of the results suggests that in developed countries returns to education increase with quantiles (higher for higher earnings quantiles), in middle-income countries the evidence is mixed, and in the few developing countries for which evidence exists returns decrease with quantiles, i.e. returns to education are higher for lower earnings quantiles[13].

If returns increase as one goes from the lower to the higher end of the earnings distribution, this can be interpreted as indicating that ability and education complement each other, with more able workers benefiting more (in terms of higher earnings) from additional investment in education. On the other hand, a negative relationship between ability and returns to education (decreasing returns with earnings quantiles) suggests substitutability between education and ability. Finally, if there is no distinct pattern, then average returns (in the absence of biases in their estimation) capture the overall profitability of education.

We used the PIHS (1998-99) to estimate quantile regressions and our results are reported in Table 12A. The results show that in wage employment, for women, returns to education are highest in the lowest quantile of earnings (bottom quartile) and lowest in our highest earnings group (the top quartile). In other words, those with lower ability have higher

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

rates of return to education. This is true for both the young and the old age group of women. This suggests that for women waged workers, education is inequality reducing, since education reduces the wage differences between low and high ability individuals, rather than increasing them. There is no such pattern for males.

In self-employment, both among young women and among 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, though this difference is not statistically significant due to imprecision of estimates on account of the small sample size. For old men in self-employment, the return to education in the top quartile (7.2%) is 1.6 percentage points (or 28%) higher than that in the bottom quartile (5.6%), and this difference is statistically significant since both are very precisely determined. However, the size of the difference in returns is not economically large. Thus, we can say that in agriculture and self-employment, there is no strong pattern of returns to education being very different 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 women than among 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 from women's education and it further boosts the already strong efficiency case for public subsidization of girls' schooling in Pakistan.

Part IV
Results for 2001/02

In this part we analyse of the PIHS 2001/02 data. We use the same structure as in Part III, that is: Section 1 provides details on the sample and shows summary statistics on key variables; Section 2 examines the effects of education and cognitive skills on occupational outcome; and Section 3 analyses the effects of education and cognitive skills on earnings, conditional on occupational outcome. All tables and figures based on 2001/02 data have a suffix "B". The main purpose of this part of the analysis is to see if the key findings based on the earlier wave have changed, and if so how. In the interest of brevity, we concentrate mostly on highlighting the changes where there are any.

1. Data and descriptive statistics

The data used for this part of the analysis come from the 2001/02 round of the Pakistan Integrated Household Survey (PIHS). This sampling strategy for this survey is the same as for the 1998/99 survey, discussed in Section 1 of Part III above. The same procedures for defining occupations and for calculating earnings are used for both rounds of the survey, which allow for an adequate comparison.

Table 1B shows summary statistics for selected variables used in the analysis, for the full sample and for the five occupation categories identified, for 2001-02. Comparing these with the summary statistics for 1998-99 (in Table 1A, discussed earlier), it is clear that there is a good deal of similarity in the overall labor market picture in most respects, which is perhaps unsurprising given the short 3 year gap in between. The labor force participation rate remained the same at about 51% and the distribution of the adult population into the different labor market states did not change greatly over time except for the proportion of the labor force employed in agriculture which fell by 5 percentage points over these 3 years (from 30% to 25%). Correspondingly, the proportion of the labor force employed in wage employment rose by 2.5 points, in self-employment by 1.4 points and in unemployment by 1.2 points. Average earnings in the full sample did not change in nominal terms (suggesting a fall in real terms) though this masks modest changes in opposing directions in the mean earnings of self- and wage-employed groups. The very large difference in mean earnings between agricultural workers on the one hand and both self-employed and wage employed workers on the other,

remains in 2001-02. The hierarchy of average years of education by occupation also did not change over time though mean years of education among the self-employed increased conspicuously and mean education in wage employment fell somewhat. These changes in education by occupation explain, at least in part, the evident reduction in mean earnings in wage employment and rise in mean earnings in self-employment over the 3 years. The percentage of workers who are numerate increased appreciably over time in most occupations but not the percentage of workers who are literate.

In summary, descriptive statistics show that the overall labor market picture has not changed much. Relevant quantities have moved in the expected directions, e.g. there was a reduction in the proportion employed in agriculture, mean education and cognitive skills have risen and mean number of children has fallen over time. It remains the case in 2001-02 that the main difference with regard to skills and earnings is between self-employed and wage-employed on the one hand, and agricultural workers and OLF persons on the other. This suggests that skills continue to matter a lot for which of these two broadly defined occupation groups individuals end up in. We now investigate the correlates of occupational outcome more in detail.

2. Education and occupational attainment

We model occupational outcomes for men and women and by age group (young and old), and use years of education as our measure of skills. Table 2B shows marginal effects for number of children, number of elderly people in the household and marital status for 2001-02. While most of the results are remarkably similar to those in Table 2A for 1998-99, the differences suggest that the way in which household demographics impinge on occupational choice/outcomes seems to have become stronger over time. There are several examples to illustrate this[14].

Figure 1B shows the relationship between years of education and the predicted likelihoods of occupational outcomes, for young men (panel i) and young women (panel ii),

[14] 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 results for number of elderly people on chances of being in agriculture are mostly stronger in 2001-02. Again, the marginal effect of having elderly persons in the household on the chances of being wage employed are significantly stronger for men in 2001-02 than in 1998-99. Lastly, number of children in the household increases older men's chances of being OLF very statistically significantly in 2001-02 but not in 1998-99. Of course, this is not always the case and there are one or two counter-examples, e.g. the results for number of children on the chances of being in self-employment are less strong in 2001-02, but this is a less common occurrence.

evaluated at the sample mean values of the other explanatory variables in the model. Education is associated with a lower likelihood of being involved in agricultural production and a higher likelihood of being OLF and likelihood of self-employment has an inverse u-shaped in education.. In 2001-02 there is suggestion that education weakly reduces young men's chances of being in wage employment. The extent to which education reduced 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 1999. For women the picture for 2001-02 is remarkably similar to that for 1998-99 with women with up to about 8 years of education are very unlikely to work but education at secondary level and beyond strongly raising the likelihood of wage-employment. Education matters more for women than men in terms of determining the type of occupation.

Figure 2B plots the estimated occupation probabilities as a function of age for young persons (aged 16-30), 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% chance of being in the labor force. By contrast, age matters strongly to men's decision to enter gainful employment, i.e. the OLF curve falls sharply between ages 15 and 25. Age and men's waged work participation have a inverted-U shaped relationship. Comparing with Figure 2A for 1998-99 indicates that the relationship of age with self-employment chances was far steeper in 1999. Panel (ii) for young women shows that women's occupational outcomes have become less responsive to age over time. This is particularly conspicuous for the relationship between age and the chances of being in wage employment and, to a lesser extent, being OLF.

Figures 3B and 4B repeat Figures 1B and 2B for older individuals only (aged 31-70) in 2001-02. In Figure 3B, older men's likelihood of being wage employed rises strongly with education beyond 5 years of education but for older women's chances of wage employment rise with education only beyond 10 years of education. Education also deters entry into agriculture. Amongst older women, very high levels of education make it pretty certain that they will be in wage work and the relationship is steeper than for young women. Education beyond secondary level also spurs older women to become labor force participants. The basic patterns for older women are similar to 1999. Figure 4B shows the relationship between *age* and occupation outcome for older men and women. For men, age raises chances of wage employment but has relatively little effect on the likelihood of entering the other occupations

though In 1998-99 older men's chances of being OLF had fallen strongly with age. Among older women (panel ii), age decreases the likelihood of being OLF (i.e. increases chances of workforce participation) and increases the likelihood of both waged and agricultural work, though the latter relationship is flatter than in 1999.

Table 3B sets out the marginal effects of literacy and numeracy on occupational outcome for 2001-02. Possession of 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 very greatly reduce the prospect of being in the lowest paying part of the labor market, namely in agriculture. Numeracy skills also strongly increase the probability of being in well paid work, both wage and self-employment. Comparison with Table 3A (fro 1998-99) shows that in general, the relationship of *literacy* with occupational outcome fell in some cases (e.g. for most groups' chances of entry into 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 is most conspicuous in wage employment and OLF but also 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 for well paid jobs and that skills increasingly play a bigger role in sorting people into different occupations or rationing the better paid jobs.

3. Education and Earnings

3.1 The basic relationship

Table 4B presents basic OLS estimates of the Mincerian returns to education in Pakistan, by occupation, gender and age group, for 2001/02. The results for the wage-employed are strikingly similar to the estimates based on the earlier wave of the data. 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, for old men the education coefficient is 0.066 (0.070 for 1998/99) and for old women it is 0.183 (0.172 for 1998/99). All estimates are

strongly significantly different from zero. It is thus clear that returns to education remain statistically significantly greater for the older group than for the young, amongst the wage-employed.

Amongst the self-employed and agricultural workers the changes in the education coefficient are somewhat larger, but in most cases we can accept the hypothesis that the education coefficient is constant across the two time periods. However there are two conspicuous differences, both occurring 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, and again this difference is statistically significant. Exogenous events like rainfall probably have a larger impact on earnings amongst agricultural workers and the self-employed, than on the earnings of the wage employed, which could be part of the explanation as to why we see some changes in the estimated returns for the former two occupations. However, it remains unclear why the returns to education amongst 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.[15]

Taking stock of these findings, we thus note 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. This is clear from the graphs of predicted earnings in Figures 5B to 7B where although the slope of the education-earnings relationship is steeper (at least on balance) for women, the intercept is far lower for women than men (these graphs are based on regressions yet to be discussed).

3.2 Extensions on the education-earnings relationship

[15] We have results (not reported) indicating 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. 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 where it is equal to 0.18.

Correcting returns estimates for endogeneity bias

As stated in Part II, OLS estimates of returns to education potentially suffer from sample selectivity bias and endogeneity bias. We attempt to address the former by employing the Heckman procedure, explained in Part II. The multinomial logit equations in Appendix 2 were used to calculate the selectivity terms. The results are presented in Table 5B. The selectivity term is statistically significant in 5 out of 12 earnings regressions, and the introduction of the selection term generally reduces the return to education (although one exception is for 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 are very similar for this dataset as for the 1998/99 data. Since selectivity correction makes a difference in some cases, we prefer the selectivity corrected equations to OLS.

The problem of endogenous sample selection is akin to the problem of endogeneity bias, as discussed in Part II. We now allow for household fixed effects in estimating the earnings function for the wage-employed, on the grounds that this can be viewed as an alternative way of addressing the endogeneity problem.[16] The fixed effects results, shown in Table 6B, indicate lower returns to education than the OLS estimates. This is consistent with the selectivity corrected estimates, and also exactly what we found based on the 1998/99 data. In fact, as can be seen in Tables 6A and 6B, the fixed effects results for 2001/02 are very similar to the fixed effects results for 1998/99. It is possible, of course, that the reduction in estimated returns to education in Table 6B compared with the OLS results may be driven

[16] Two stage least squares results for wage employees can be summarized as follows: i) Young men: using father's and mother's education as instruments, and losing about 40% of the observations in the process (see footnote 4), the coefficient on education rises from 0.035 (OLS, see Table 4B) to 0.059 (significant at the 1% level), and the validity of the overidentifying restrictions is rejected at the 5% level; adding spouse's education to the instrument is not feasible as we would lose 80% of the observations; using spouse's education as the only instrument, we lose 60% of the observations, and the coefficient rises to 0.075 (significant at the 1% level). ii) Young women: using father's and mother's education as instruments, we lose about 60% of the observations, the coefficient on education falls from 0.144 (OLS, see Table 4B) to 0.129 (significant at the 1% level), and the validity of the overidentifying restrictions is accepted at the 10% level; adding spouse's education to the instrument is not feasible as we would lose 99% of the observations; using spouse's education as the only instrument, we lose 55% of the observations, and the coefficient rises to 0.18 (significant at the 1% level). iii) 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, we lose 10% of the observations, and the coefficient rises from 0.070 (OLS; Table 4B) to 0.105 (significant at the 1% level). iv) 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, we lose 25% of the observations, and the coefficient rises from 0.183 (OLS; Table 4B) to 0.192 (significant at the 1% level).

by measurement errors bias (see section 3.2 in Part III). For this reason, and because the household fixed effects results can be estimated only for the sub-sample of wage employed persons, we continue to take the selectivity corrected results as our preferred estimates.

Shape of the education-earnings relationship

We now relax linearity on the association between education and earnings. Table 7B, estimated using the preferred sample selectivity corrected estimator, introduces quadratic terms in education. Its OLS and household-fixed-effects counterparts are included in Appendices A9B and A10B respectively. Table 7B shows that the education-earnings relationship is convex for both old and young men in wage-employment, and 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, we cannot reject a linear relationship.

The non-linearities of the education-earnings relationship are explored further in Table 8B which includes a dummy variable for each education level, and which includes a selectivity correction term.[17] The base education category is 'no education'. The marginal return to each year of primary education, to each year of middle education and so forth, calculated from Table 8B, are set out in Table 9B. It confirms that marginal returns to education are generally substantially lower for men than women in both wage and self-employment, though not in agriculture. It should be noted however that the marginal returns to education in agriculture for women are very imprecisely estimated. It also confirms that marginal returns are generally higher for the older age group than for the younger one. The evidence of Tables 7B to 9B taken together suggests that the Pakistan labour market is not generally characterized by the commonly assumed concave relationship which implies diminishing returns to extra years of schooling.

Earnings and Cognitive Skills

We now turn to the specifications where education is replaced by our measures of cognitive skills. Table 10B shows OLS estimates and Table 11B shows results corrected for selectivity. We focus on the latter.[18] The results indicate positive, and often high, returns to literacy in all

[17] OLS results are shown Appendix A11B, and household fixed effects results for the wage employed are included in Appendix A12B.
[18] Household fixed effects results are reported in Appendix A13B.

cases except for women in agriculture (where returns insignificant). In wage employment, and amongst young self-employed individuals, the returns to literacy are much larger for women than men. This is similar to what we found based on the 1998/99 data, and could be due to a scarcity premium since far fewer women than men are literate.

The results on numeracy skills are quite mixed. In four cases, the estimated coefficient on numeracy skills is actually negative and significant, suggesting these skills *reduce* earnings. This counter-intuitive result is due to the fact that numeracy and literacy skills are highly correlated. If we exclude the literacy variable, the coefficient on numeracy skills tends to rise substantially. In other words, while the relationship between numeracy skills and earnings *conditional on literacy skills* is sometimes negative and significant, the unconditional relationship is usually positive and often large. This is probably how the effects of numeracy should be viewed.

Heterogeneity in returns to education

Finally, we return to quantile regressions as a basis for examining heterogeneity in returns to education. Results, shown in Table 12B, indicate that in wage employment, for women, returns to education are highest in the lowest quantile of earnings (bottom quartile) and lowest in our highest earnings group (the top quartile). In other words, those with lower ability have higher rates of return to education. This is true for both the young and the old age group of women. This suggests that for women waged workers, education is inequality reducing, since education reduces the wage differences between low and high ability individuals. Recall that a very similar pattern was found for the 1998/99 data. For the other occupation categories results are less clear.

Part V
Conclusions

This paper has focused on varied aspects of the education-earnings relationship in Pakistan at two points in time: 1998-99 and 2001-02. It has sought to examine (i) the role of education in occupational attainment; (ii) the role of education in raising earnings, conditional on occupation; (iii) the role of cognitive skills in both occupational attainment and earnings determination; and (iv) the role of education in earnings at different points of the earnings distribution. The labour market benefits of education accrue both from education promoting a person's entry into the lucrative occupations and, conditional on occupation, by raising earnings. The findings from the two rounds of the Pakistan data are remarkably similar.

We find that education plays a very important role in occupational outcomes but that this role differs greatly between the genders. Both young and older women begin to take advantage of the benefits of education in earnest only after about 10 years of schooling when they begin to join the labor force and to enter wage employment. Among young men, the likelihood of wage employment is unresponsive to education level and young men also increasingly quit the labor force or queue unemployed as education level increases.

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

The paper investigated the shape of the education-earnings relationship. It finds that the shape is not concave, with diminishing returns to education, as conventional wisdom suggests. In wage employment for men and for some worker groups in other occupations, the relationship is convex. The implications of this are considerable. Extant education and labor market policy is predicated on the assumption that returns to education are the greatest at the

primary level and are progressively lower at secondary and tertiary levels. The Millennium Development Goals also presume that the completion of basic education will help towards realization of the goal to halve world poverty by 2015. However, if 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 will not prove an effective means of helping people to climb out of poverty.

We estimated returns to education along the earnings distribution, separately for young and old men and women. While the findings are mixed, one clear pattern is discernible. Among both young and old women in wage employment, education is inequality-reducing: lower ability women have higher returns to education than higher ability ones. Given that education is associated mainly with *wage* employment for Pakistani women (Figure 3A, ii), the fact that it plays an inequality reducing role is welcome news and it can be viewed as a non-market 'externality' effect of women's education, which 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. We find that cognitive skills have big pay-offs for both women and men in Pakistan. In particular literacy promotes entry into the lucrative parts of the labor market for both men and women, though the size of the relationship is bigger for men. Conditional on occupation, literacy is associated with very substantially higher earnings within wage and self-employment again for both men and women, though in this case the size of the relationship is very significantly bigger for women than men.

References

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

Andrabi, T.,Das, J. and Khwaja, A. (2002), 'The Rise of Private Schooling in Pakistan: Catering to the Urban Elite or Educating the Rural Poor?' World Bank and Harvard University.

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

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

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

Aslam, M. (2007), 'Rates of Return to Education by Gender in Pakistan', GPRG-WPS-064.

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

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

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

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

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

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

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

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

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

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

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

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

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

Main Results, 1998/99

Table 1A
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

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 2A
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 1. Robust t-statistics in parentheses. $^{+}$ significant at 10% level; * significant at 5% level; ** significant at 1% level.

Table 3A
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 1.

Table 4A
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 5A
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 6A
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 7A
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 8A
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 9A
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 8A.
* 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 10A
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 11A
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 maths problem	0.046	0.195	-0.025	-0.606	0.332	0.252
	(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 maths problem	0.106	0.084	0.228	0.264	0.335	0.354
	(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.

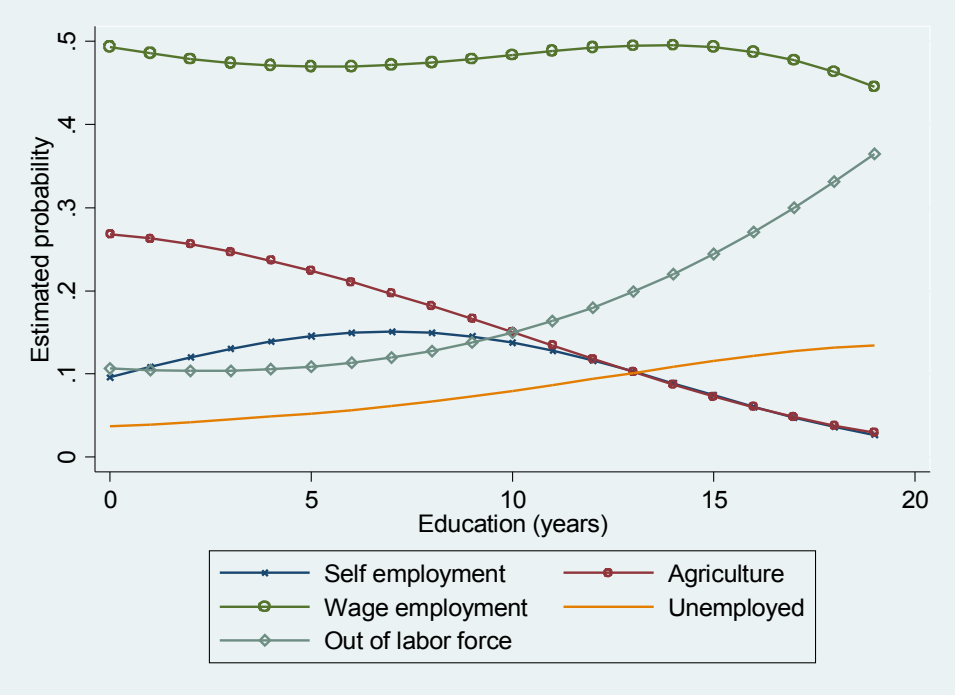
Table 12A
Earnings and years of schooling: Quantile regressions

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

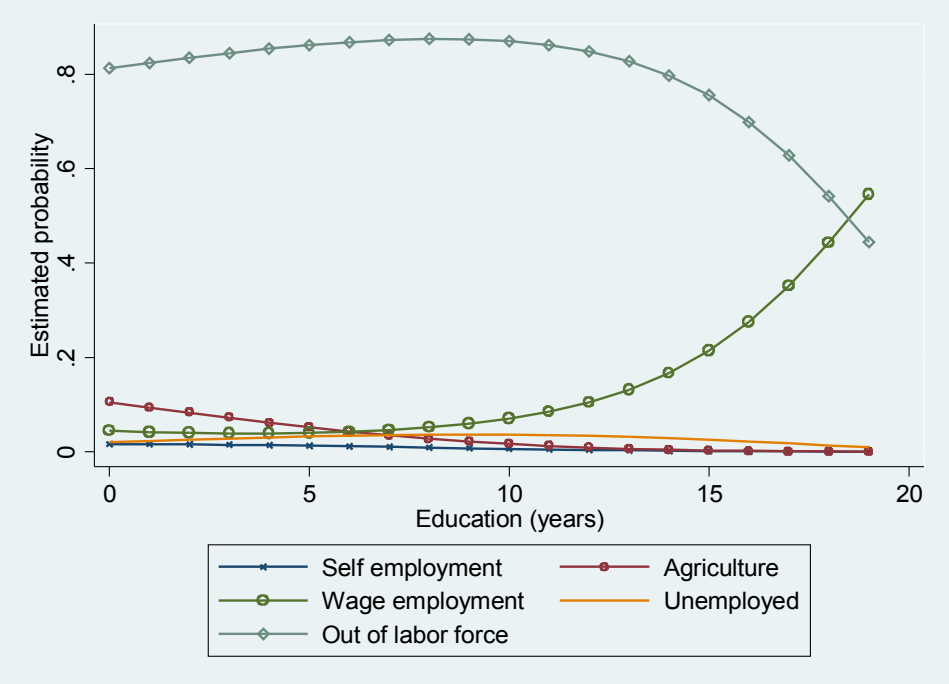
Note: Age, age squared, and province dummy variables are included in all regressions. Standard errors in parentheses.

Figure 1A
Young individuals: Estimated probability of occupation and education

i) Men



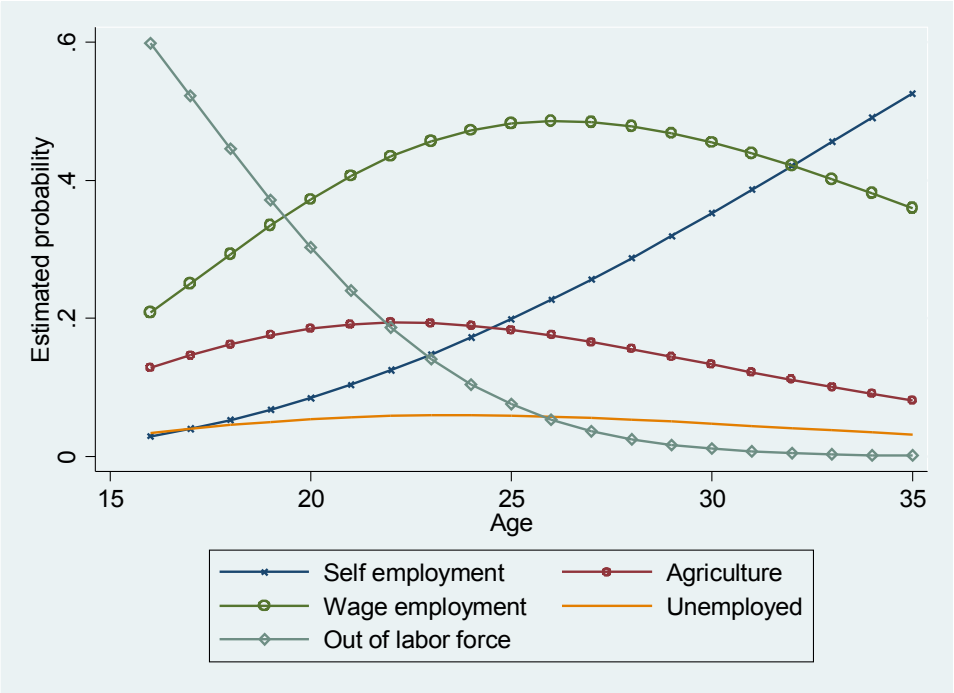
ii) Women



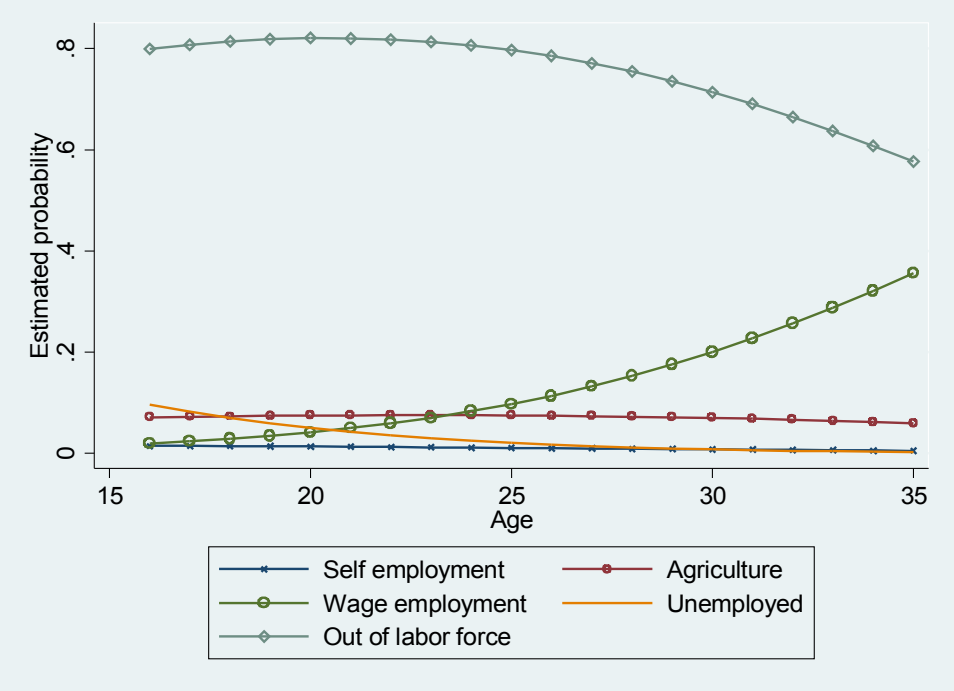
Note: These predictions are based on the multinomial logits reported in Appendix 1.

Figure 2A
Young individuals: Estimated probability of occupation and age

i) Men



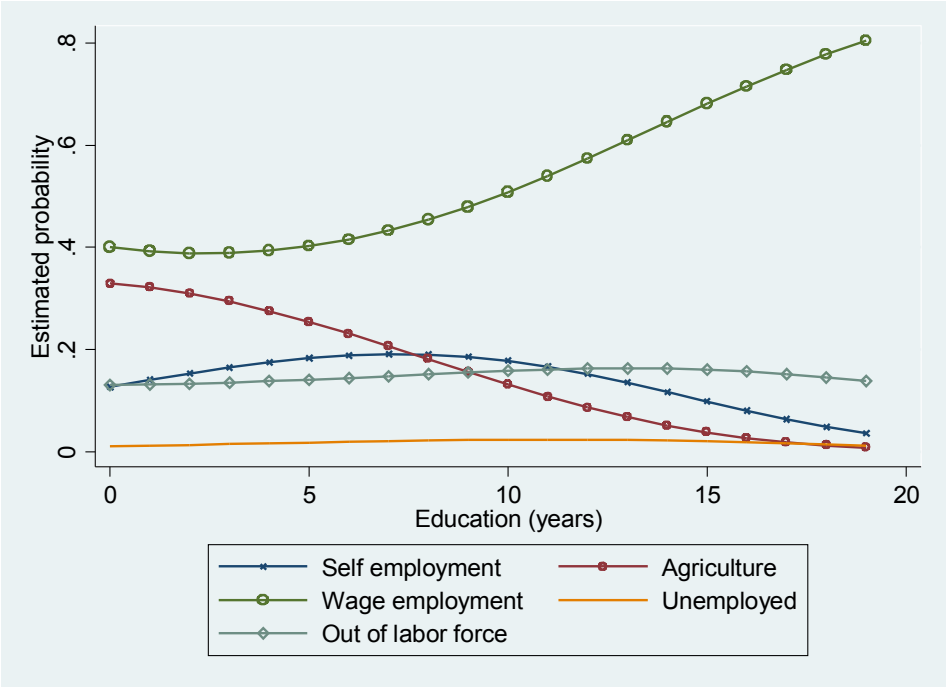
ii) Women



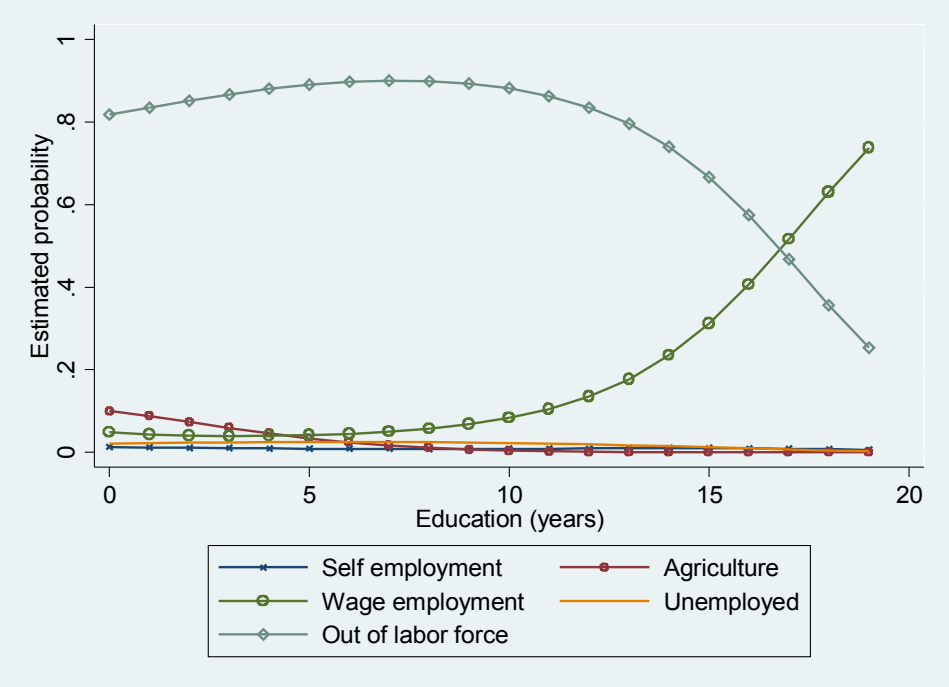
Note: These predictions are based on the multinomial logits reported in Appendix 1.

Figure 3A
Old individuals: Estimated probability of occupation and education

i) Men



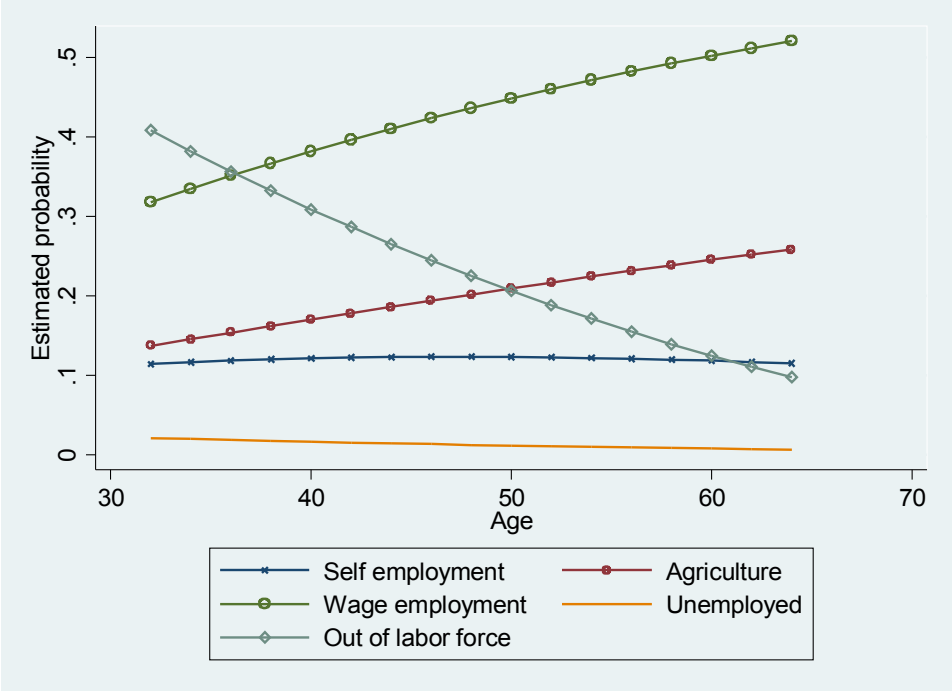
ii) Women



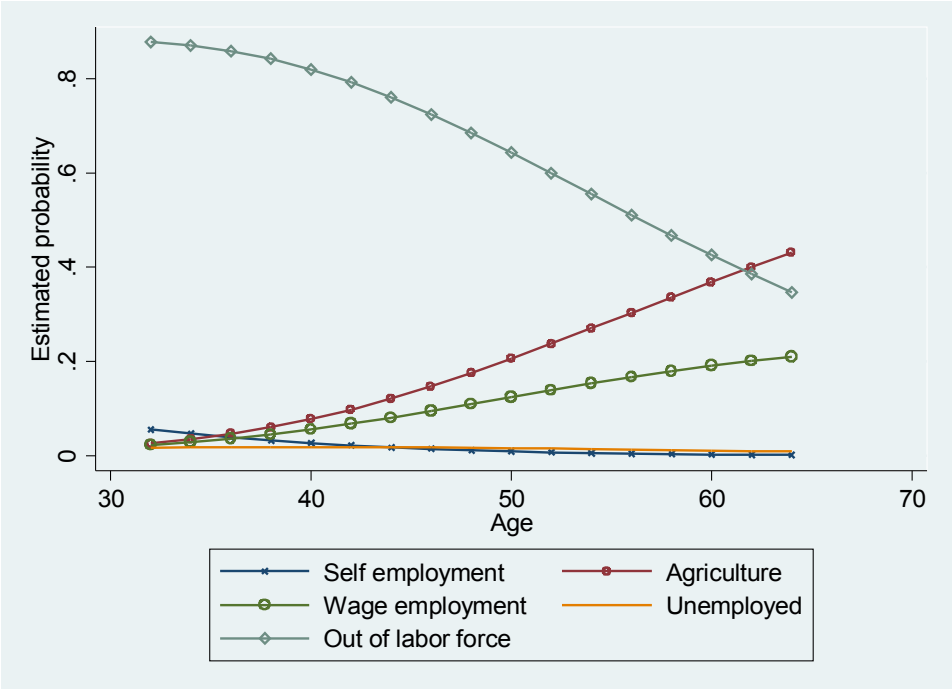
Note: These predictions are based on the multinomial logits reported in Appendix 1.

Figure 4A
Old individuals: Estimated probability of occupation and age

i) Men

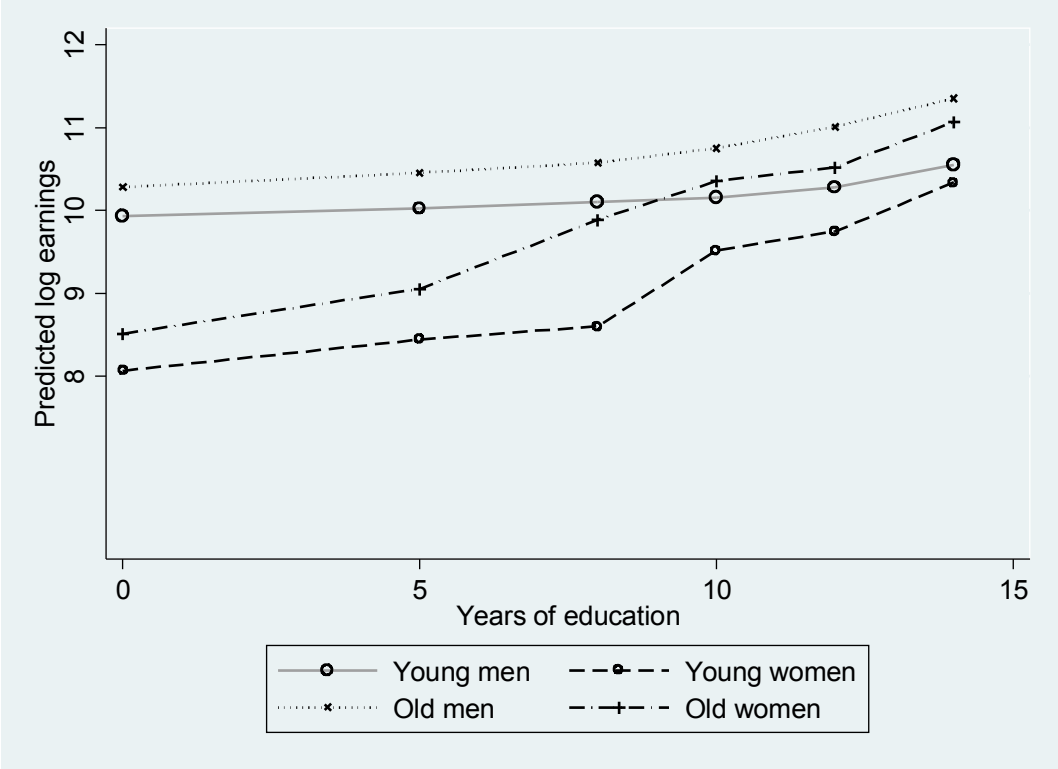


ii) Women



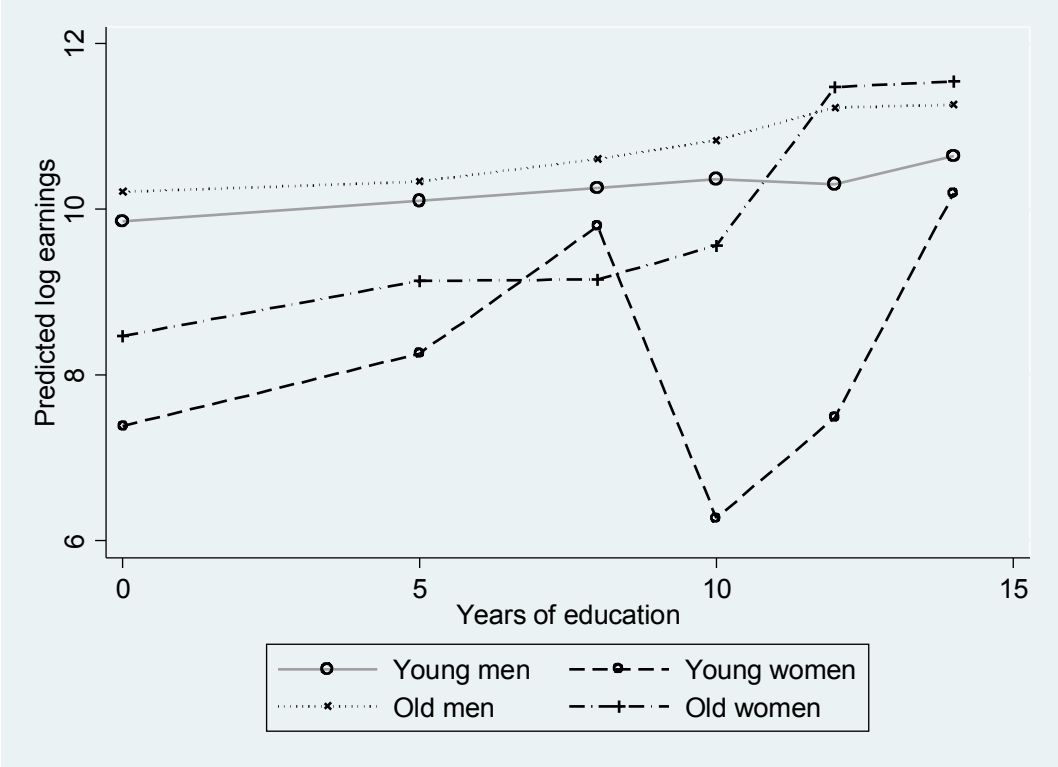
Note: These predictions are based on the multinomial logits reported in Appendix 1.

Figure 5A
Predicted earnings and level of education: Wage employed



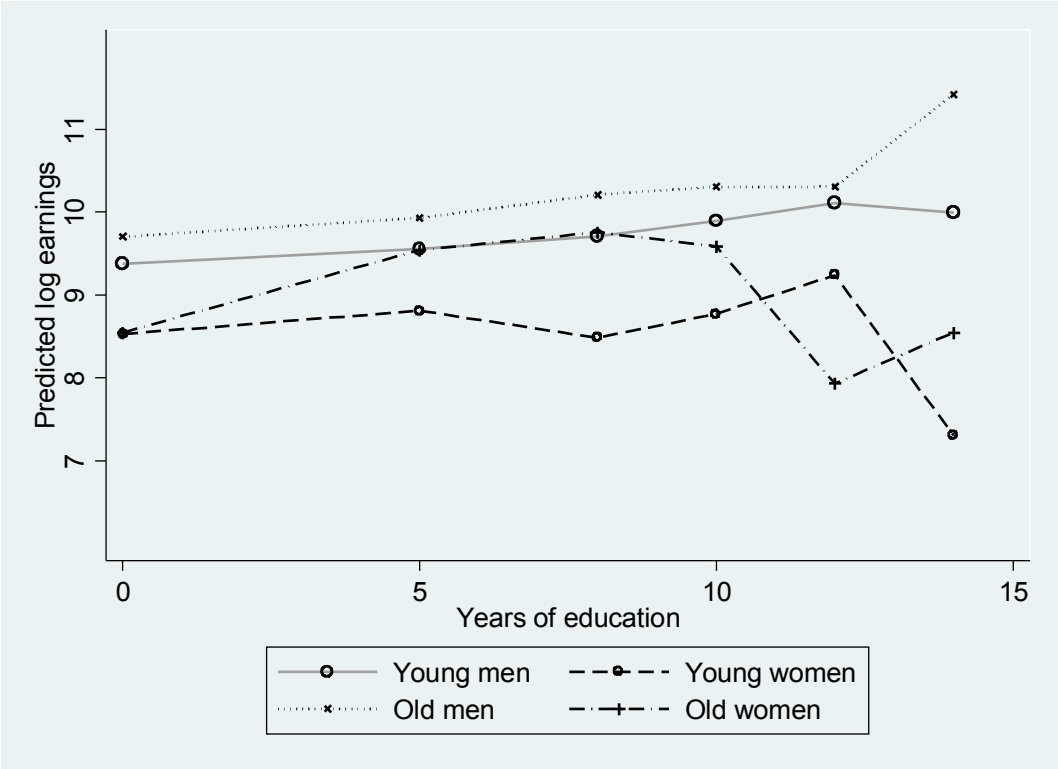
Note: These predictions are based on the results reported in Table 10A.

Figure 6A
Predicted earnings and level of education: Self employed



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

Figure 7A
Predicted earnings and level of education: Agriculture



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

Main Results, 2001/02

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

	All	Self-employed	Agricult. employed	Wage employed	Unemp-loyed	Out of labor force
Annual earnings	30148	39246	20754	33219	--	--
[*median*]	[*24000*]	[*30000*]	[*11113*]	[*30000*]		
Log earnings	9.75	10.05	9.22	9.98	--	--
[*median*]	[*10.09*]	[*10.31*]	[*9.32*]	[*10.31*]		
Years of education	3.51	5.14	2.51	5.01	4.39	2.67
Age	35.3	35.6	36.8	33.6	30.0	36.0
Proportion men	0.47	0.90	0.67	0.83	0.44	0.14
Math skills	0.74	0.89	0.74	0.84	0.69	0.67
Read & write skills	0.41	0.61	0.31	0.55	0.49	0.32
# children aged < 12 in household	2.58	2.62	2.72	2.48	2.23	2.61
# individuals aged > 65 in household	0.26	0.25	0.30	0.20	0.18	0.28
Proportion married	0.68	0.70	0.72	0.67	0.50	0.68
Observations	47741	3798	6164	12798	1774	23207
Earnings obs	22760	3798	6164	12798	0	0

Earnings distributions:

Note: Earnings are measured in 2001/02 Pakistan rupees. The USD exchange rate over the sampling period is approximately 64. 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 2B
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	0.003	0.000	0.001	-0.001
	(2.43)*	(0.07)	(0.99)	(1.44)
# individuals aged > 65 in household	0.010	-0.004	0.018	0.002
	(1.40)	(1.55)	(2.71)**	(1.05)
Individual is married	0.046	-0.015	0.028	-0.009
	(4.67)**	(9.78)**	(2.13)*	(3.92)**
2. Agriculture				
# children aged < 12 in household	0.009	0.002	0.007	0.001
	(6.24)**	(1.77)$^{+}$	(4.68)**	(1.05)
# individuals aged > 65 in household	0.046	0.007	0.054	0.011
	(6.40)**	(1.64)	(7.72)**	(2.43)*
Individual is married	0.036	0.000	0.039	0.034
	(3.54)**	(0.06)	(2.81)**	(4.00)**
3. Wage employment				
# children aged < 12 in household	-0.014	-0.001	-0.013	-0.002
	(6.66)**	(0.68)	(6.71)**	(1.38)
# individuals aged > 65 in household	-0.044	-0.004	-0.057	-0.003
	(4.24)**	(0.79)	(6.42)**	(0.52)
Individual is married	0.050	-0.048	0.088	-0.039
	(4.08)**	(10.64)**	(5.36)**	(6.48)**
4. Unemployed				
# children aged < 12 in household	-0.001	-0.002	-0.002	-0.001
	(1.21)	(2.93)**	(2.70)**	(1.94)$^{+}$
# individuals aged > 65 in household	0.000	-0.013	-0.003	-0.011
	(0.08)	(2.84)**	(1.08)	(2.55)*
Individual is married	-0.041	0.002	-0.022	0.002
	(11.15)**	(0.52)	(7.07)**	(0.39)
5. Out of labor force				
# children aged < 12 in household	0.003	0.002	0.006	0.003
	(1.96)$^{+}$	(1.05)	(4.86)**	(1.71)$^{+}$
# individuals aged > 65 in household	-0.012	0.013	-0.012	0.000
	(1.59)	(1.75)$^{+}$	(1.92)$^{+}$	(0.05)
Individual is married	-0.091	0.062	-0.134	0.012
	(15.84)**	(7.46)**	(16.04)**	(1.09)

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 3B
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	0.034	0.009	0.042	0.009
	(2.31)*	(2.53)*	(3.67)**	(2.97)**
Can read & write	0.052	-0.004	0.038	-0.001
	(5.27)**	(1.93)$^{+}$	(4.54)**	(0.54)
2. Agriculture				
Can solve simple maths problem	0.014	0.025	0.024	0.018
	(1.23)	(4.32)**	(2.36)*	(3.20)**
Can read & write	-0.093	-0.098	-0.136	-0.092
	(13.15)**	(30.13)**	(21.82)**	(24.89)**
3. Wage employment				
Can solve simple maths problem	-0.003	0.022	-0.054	0.029
	(0.14)	(2.99)**	(4.13)**	(4.39)**
Can read & write	-0.062	0.009	0.070	0.015
	(5.05)**	(1.44)	(6.96)**	(2.19)*
4. Unemployed				
Can solve simple maths problem	0.012	-0.022	-0.007	-0.025
	(1.04)	(5.45)**	(1.82)$^{+}$	(9.77)**
Can read & write	0.036	0.014	0.008	0.001
	(4.47)**	(2.38)*	(2.14)*	(0.12)
5. Out of labor force				
Can solve simple maths problem	-0.058	-0.034	-0.005	-0.032
	(4.90)**	(3.42)**	(0.64)	(3.62)**
Can read & write	0.067	0.079	0.020	0.078
	(6.50)**	(9.08)**	(2.81)**	(8.58)**

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

Table 4B
Earnings and years of schooling

	1. Wage employed		2. Self employed		3. Agriculture	
	Men	Women	Men	Women	Men	Women
A. Young						
Education	0.035	0.144	0.049	0.060	0.067	0.102
	(18.85)**	(20.78)**	(6.90)**	(2.45)*	(5.56)**	(2.02)*
Age	0.153	0.026	0.085	0.418	0.152	0.240
	(6.04)**	(0.27)	(0.87)	(1.39)	(1.03)	(1.01)
Age squared	-0.002	0.000	-0.001	-0.008	-0.002	-0.003
	(3.70)**	(0.21)	(0.31)	(1.25)	(0.58)	(0.66)
# Individuals	5246	963	1543	162	1782	887
B. Old						
Education	0.070	0.183	0.056	0.056	0.091	-0.009
	(52.86)**	(37.36)**	(11.23)**	(2.21)*	(11.11)**	(0.12)
Age	0.070	0.159	0.024	-0.057	-0.002	-0.141
	(9.08)**	(4.42)**	(1.05)	(0.71)	(0.08)	(1.97)*
Age squared	-0.001	-0.002	-0.000	0.001	-0.000	0.001
	(8.75)**	(4.11)**	(1.52)	(1.05)	(0.08)	(1.84)
# Individuals	5593	996	1927	166	2516	979

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 5B
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	0.040	0.117	0.043	0.067	0.086	-0.247
	(18.99)**	(11.29)**	(4.98)**	(2.67)**	(4.98)**	(0.80)
Selection term	-0.543	-0.697	-0.353	0.721	-0.618	3.271
	(4.42)**	(3.79)**	(1.18)	(1.25)	(1.57)	(1.15)
# Individuals	5246	963	1543	162	1782	887
B. Old						
Education	0.056	0.165	0.056	0.037	0.123	-0.087
	(32.50)**	(14.95)**	(10.46)**	(1.39)	(8.32)**	(0.64)
Selection term	-0.686	-0.361	0.001	-2.245	-0.735	0.521
	(10.73)**	(2.27)*	(0.01)	(2.88)**	(2.64)**	(0.69)
# Individuals	5593	996	1927	166	2516	979

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

	Young Men	Young Women	Old Men	Old Women
Education	0.011	0.103	0.040	0.133
	(3.08)**	(17.06)**	(10.12)**	(21.05)**
# Individuals	5246	963	5593	996

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 7B
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	-0.014	0.096	-0.041	0.058	0.123	0.038
	(2.19)*	(3.25)**	(1.21)	(0.60)	(3.40)**	(0.11)
Education squared	0.004	0.002	0.006	0.001	-0.004	-0.033
	(8.73)**	(0.83)	(2.57)*	(0.08)	(1.19)	(2.04)*
Selection term	-0.179	-0.507	-1.067	0.696	-0.443	2.855
	(1.38)	(1.80)	(2.63)**	(1.00)	(1.07)	(1.00)
# Individuals	5246	963	1543	162	1782	887
B. Old						
Education	0.031	0.250	0.068	-0.207	0.097	0.508
	(6.67)**	(4.81)**	(1.88)	(2.23)*	(4.03)**	(2.05)*
Education squared	0.002	-0.008	-0.001	0.023	0.004	-0.110
	(5.47)**	(1.65)	(0.35)	(2.74)**	(1.40)	(2.87)**
Selection term	-0.389	-0.998	0.129	-3.385	-0.958	0.684
	(4.41)**	(2.29)*	(0.30)	(3.85)**	(3.01)**	(0.90)
# Individuals	5593	996	1927	166	2516	979

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 8B
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	0.071	0.459	-0.186	0.194	0.541	1.189
	(2.54)*	(3.35)**	(1.51)	(0.56)	(3.88)**	(1.42)
Middle school	0.121	0.912	0.088	0.571	0.568	0.700
	(3.80)**	(5.11)**	(0.62)	(1.64)	(3.24)**	(0.47)
Secondary	0.254	1.158	0.121	0.067	0.869	2.234
	(8.85)**	(9.25)**	(0.90)	(0.19)	(4.70)**	(0.99)
Higher secondary	0.372	1.245	0.216	1.448	0.589	-0.618
	(9.71)**	(6.88)**	(1.43)	(2.54)*	(1.97)*	(0.23)
Tertiary	0.725	1.642	0.738	1.265	1.138	0.394
	(19.51)**	(7.75)**	(4.50)**	(2.07)*	(2.73)**	(0.10)
Selection term	-0.177	-0.836	-1.112	0.465	-0.306	-0.509
	(1.37)	(3.37)**	(2.87)**	(0.68)	(0.75)	(0.28)
# Individuals	5246	963	1543	162	1782	887
B. Old						
Primary	0.234	0.766	0.111	0.299	0.572	0.355
	(11.58)**	(4.49)**	(1.12)	(0.87)	(6.04)**	(0.75)
Middle school	0.369	1.230	0.213	-0.820	0.898	-1.065
	(14.83)**	(4.74)**	(1.74)	(2.20)*	(6.74)**	(0.80)
Secondary	0.521	1.784	0.431	0.047	1.144	-5.170
	(24.19)**	(10.21)**	(3.86)**	(0.12)	(6.92)**	(2.07)*
Higher secondary	0.635	1.955	0.724	0.796	1.355	--
	(20.63)**	(14.59)**	(5.50)**	(1.25)	(4.61)**	
Tertiary	0.923	1.693	0.847	2.829	2.498	--
	(24.43)**	(6.83)**	(8.35)**	(3.56)**	(6.91)**	
Selection term	-0.588	-1.109	-0.477	-2.752	-0.841	0.287
	(6.77)**	(4.62)**	(1.31)	(3.03)**	(2.75)**	(0.45)
# Individuals	5593	996	1927	166	2516	979

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

	1. Wage employed		2. Self employed		3. Agriculture	
	Men	Women	Men	Women	Men	Women
A. Young						
Primary	1.4 *	9.2 **	-3.7	3.9	10.8 **	23.8
Middle school	1.7	15.1	9.1 **	12.6	0.9	-16.3
Secondary	6.6 *	12.3	1.7	-25.2	15.0	76.7
Higher secondary	5.9	4.3	4.7	69.1 *	-14.0	-142.6
Tertiary	17.7 **	19.9	26.1	-9.1	27.5	50.6
B. Old						
Primary	4.7 **	15.3 **	2.2	6.0	11.4 **	7.1
Middle school	4.5	15.5	3.4	-37.3 *	10.9	-47.3
Secondary	7.6	27.7	10.9	43.3 *	12.3	-205.3
Higher secondary	5.7	8.6	14.6	37.5	10.5	--
Tertiary	14.4 **	-13.1	6.1	101.6	57.2 +	--

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 8B.
If the marginal return to education at a given *level* of education is statistically significantly different from the marginal return at the education level immediately below it, this is indicated by * (5% level) and ** (1% level).

Table 10B
Earnings, literacy and numeracy

	1. Wage employed		2. Self employed		3. Agriculture	
	Men	Women	Men	Women	Men	Women
A. Young						
Can solve simple maths problem	-0.063	-0.013	0.241	-1.001	0.212	0.479
	(2.08)*	(0.10)	(1.76)	(3.47)**	(1.32)	(2.40)*
Can read & write	0.239	1.302	0.285	0.834	0.405	0.364
	(11.18)**	(13.19)**	(3.37)**	(3.40)**	(3.39)**	(1.08)
Age	0.179	0.153	0.127	0.376	0.188	0.201
	(6.95)**	(1.45)	(1.30)	(1.24)	(1.27)	(0.83)
Age squared	-0.002	-0.002	-0.001	-0.007	-0.003	-0.003
	(4.49)**	(0.81)	(0.72)	(1.13)	(0.80)	(0.51)
# Individuals	5246	963	1543	162	1782	887
B. Old						
Can solve simple maths problem	-0.101	-0.092	0.171	-0.246	0.011	0.121
	(4.55)**	(0.90)	(2.05)*	(1.06)	(0.13)	(0.82)
Can read & write	0.630	2.046	0.397	0.288	0.670	-0.170
	(36.72)**	(21.47)**	(7.09)**	(1.25)	(8.72)**	(0.39)
Age	0.074	0.155	0.019	-0.065	-0.009	-0.144
	(8.87)**	(4.03)**	(0.81)	(0.80)	(0.30)	(1.99)*
Age squared	-0.001	-0.002	-0.000	0.001	0.000	0.001
	(8.84)**	(3.72)**	(1.29)	(1.11)	(0.10)	(1.86)
# Individuals	5593	996	1927	166	2516	979

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 11B
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 maths problem	-0.058	-0.156	0.131	-0.957	0.181	0.985
	(1.92)	(1.17)	(0.88)	(2.47)*	(1.12)	(1.92)
Can read & write	0.267	1.063	0.144	0.798	0.516	-2.146
	(10.94)**	(9.83)**	(1.27)	(3.21)**	(3.19)**	(0.91)
Selection term	-0.301	-1.323	-0.756	0.182	-0.395	2.935
	(2.12)*	(5.02)**	(1.84)	(0.25)	(1.00)	(1.07)
# Individuals	5246	963	1543	162	1782	887
B. Old						
Can solve simple maths problem	-0.056	-0.271	0.218	-1.170	-0.052	0.208
	(2.32)*	(2.24)*	(2.03)*	(3.76)**	(0.57)	(1.20)
Can read & write	0.574	1.916	0.443	0.310	1.038	-1.004
	(27.89)**	(16.99)**	(5.31)**	(1.35)	(7.30)**	(1.01)
Selection term	-0.420	-1.051	0.340	-4.042	-0.995	0.729
	(4.53)**	(2.54)*	(0.73)	(4.45)**	(3.08)**	(0.94)
# Individuals	5593	996	1927	166	2516	979

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.

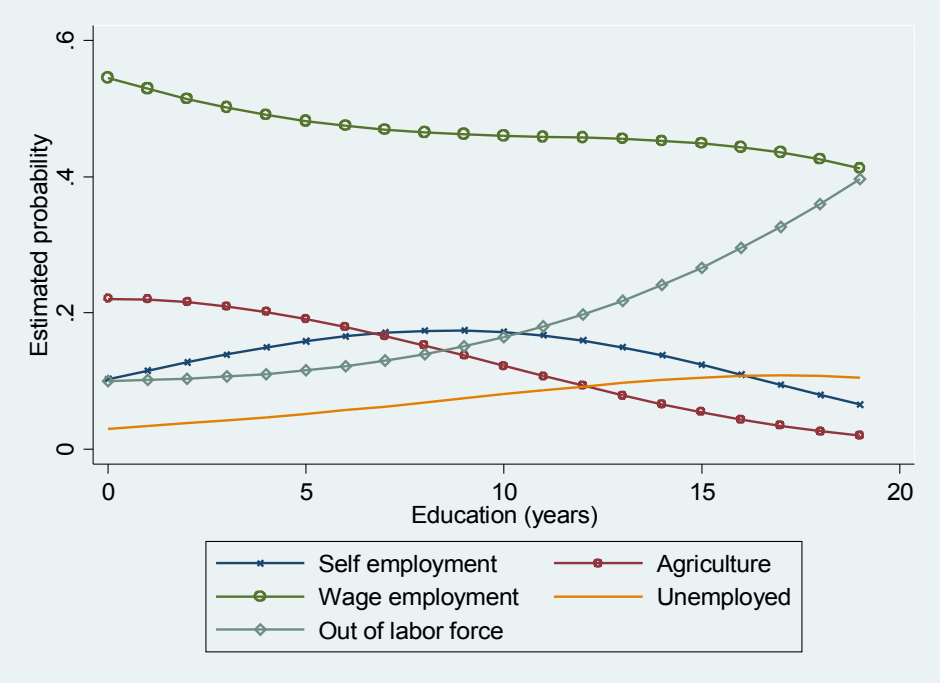
Table 12B
Earnings and years of schooling: Quantile regressions

	1. Wage employed		2. Self employed		3. Agriculture	
	Men	Women	Men	Women	Men	Women
A. Young						
Education, P25 (low)	0.036 (21.08)**	0.164 (45.79)**	0.043 (5.48)**	0.061 (2.37)*	0.067 (3.43)**	0.102 (1.12)
Education, P50 (median)	0.036 (24.49)**	0.153 (50.28)**	0.048 (5.96)**	0.090 (3.42)**	0.077 (6.41)**	0.083 (1.68)+
Education, P75 (high)	0.034 (18.88)**	0.138 (38.34)**	0.049 (7.12)**	0.064 (2.84)**	0.055 (3.88)**	0.013 (0.24)
# Individuals	5246	963	1543	162	1782	887
B. Old						
Education, P25 (low)	0.068 (47.61)**	0.226 (65.39)**	0.051 (8.99)**	0.059 (2.10)*	0.094 (6.97)**	-0.037 (0.31)
Education, P50 (median)	0.061 (49.21)**	0.190 (63.70)**	0.055 (9.64)**	0.042 (1.49)	0.082 (10.08)**	0.023 (0.34)
Education, P75 (high)	0.063 (40.69)**	0.149 (42.26)**	0.065 (13.44)**	-0.166 (1.94)+	0.077 (8.21)**	0.001 (0.02)
# Individuals	5593	996	1927	166	2516	979

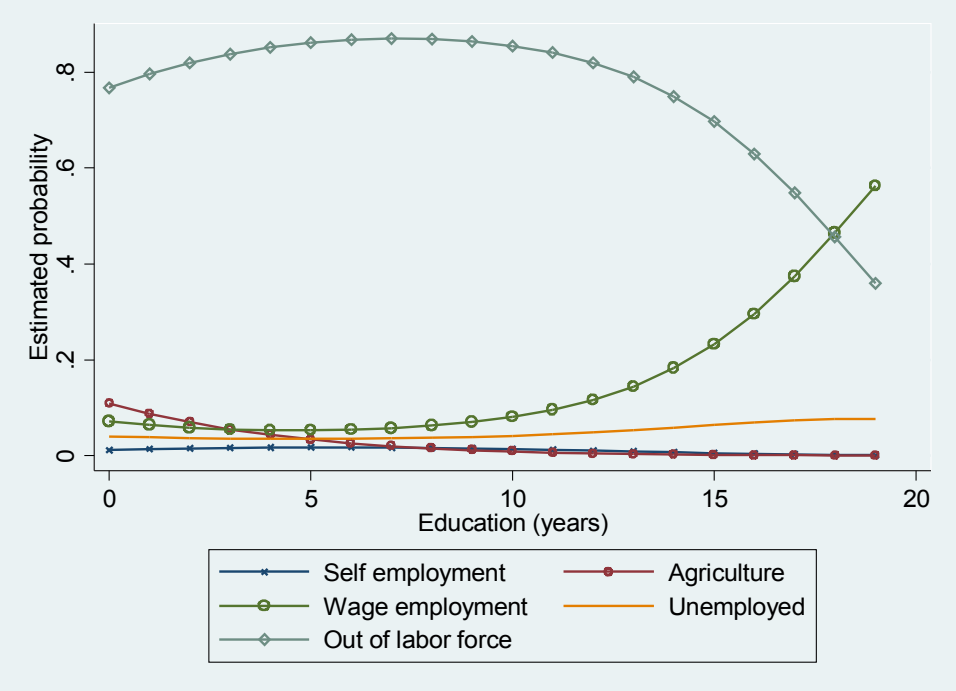
Note: Age, age squared, and province dummy variables are included in all regressions. Standard errors in parentheses.

Figure 1B
Young individuals: Estimated probability of occupation and education

i) Men



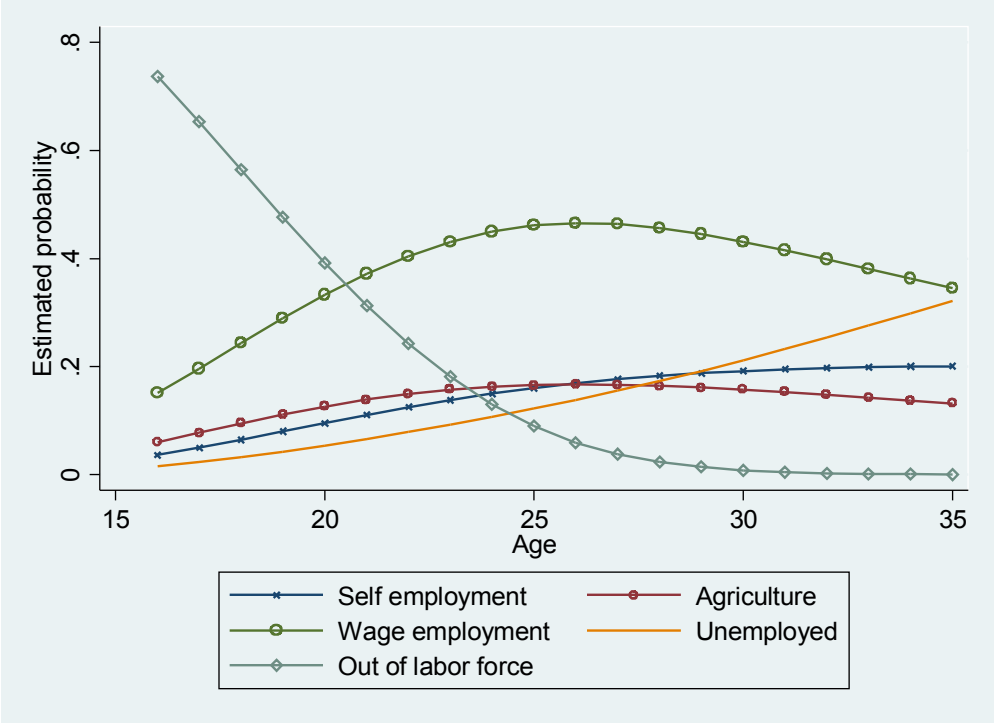
ii) Women



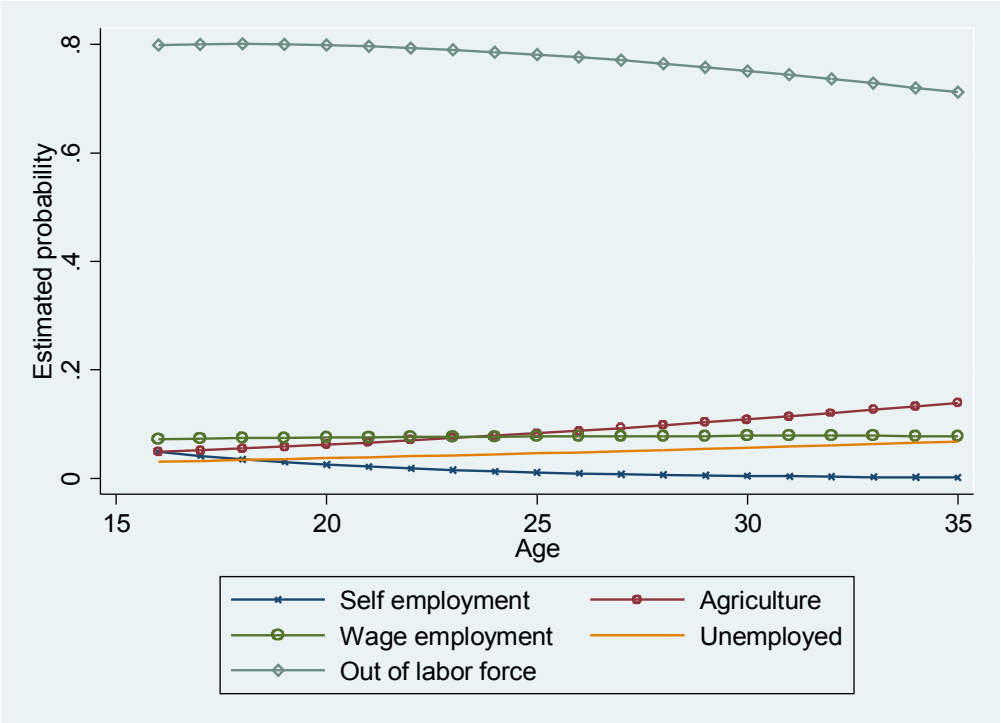
Note: These predictions are based on the multinomial logits reported in Appendix 2.

Figure 2B
Young individuals: Estimated probability of occupation and age

i) Men



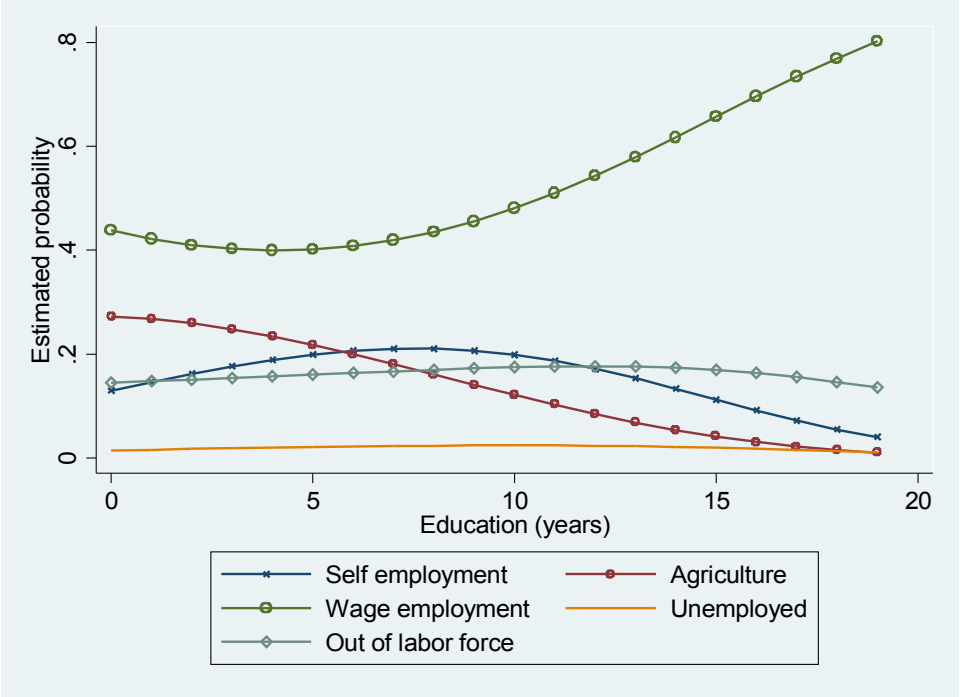
ii) Women



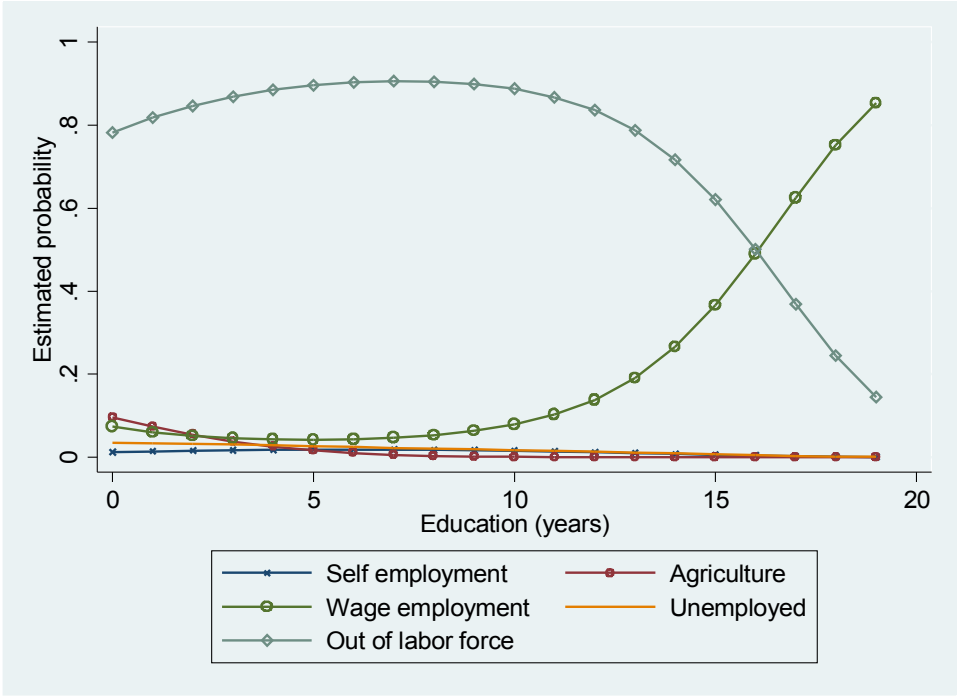
Note: These predictions are based on the multinomial logits reported in Appendix 2.

Figure 3B
Old individuals: Estimated probability of occupation and education

i) Men



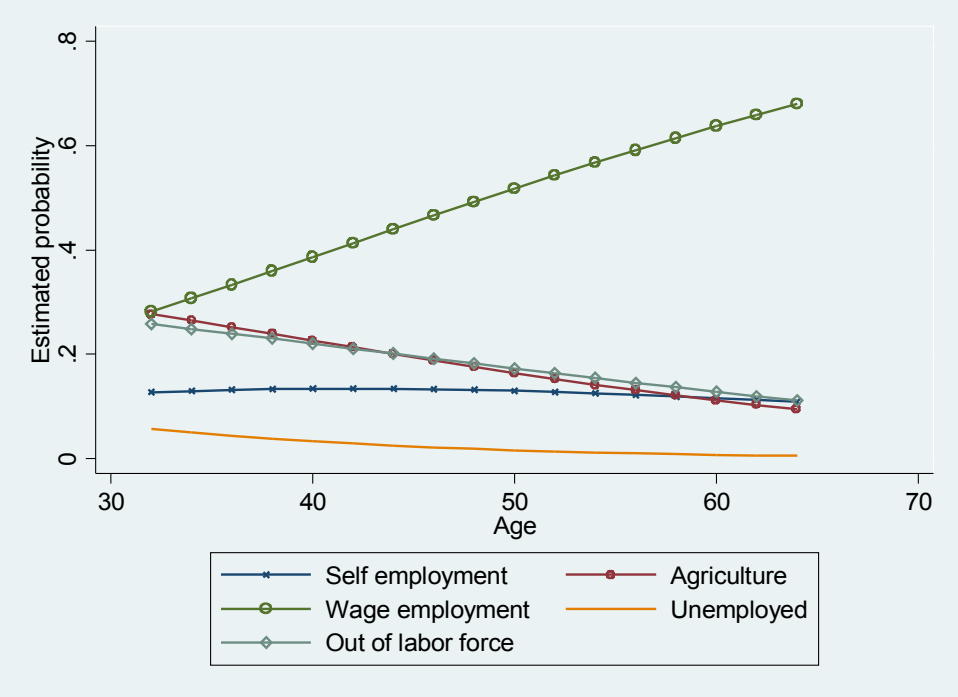
ii) Women



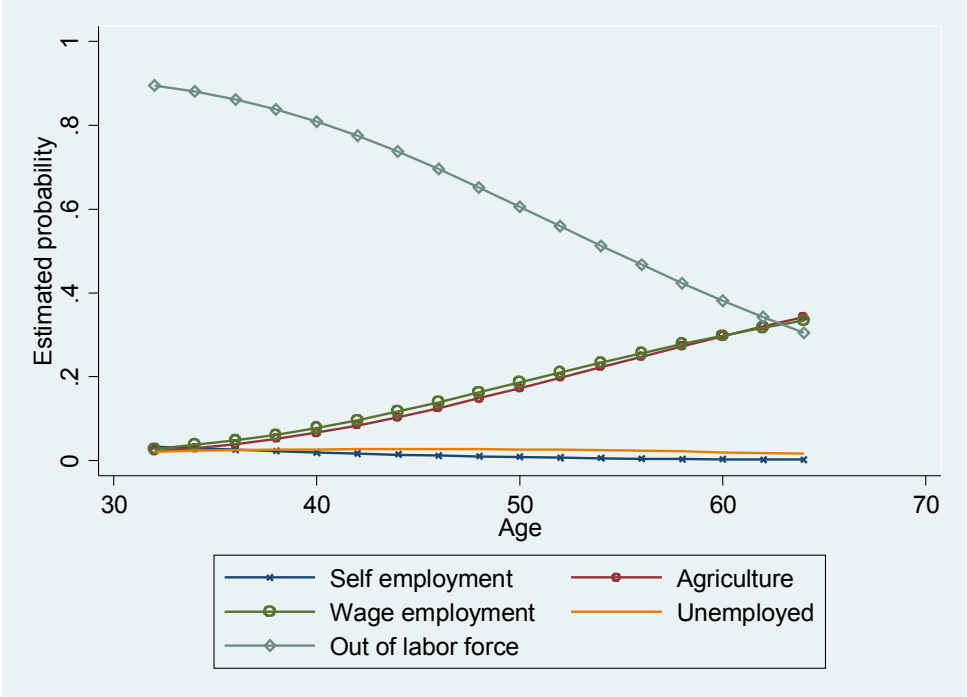
Note: These predictions are based on the multinomial logits reported in Appendix 2.

Figure 4B
Old individuals: Estimated probability of occupation and age

i) Men

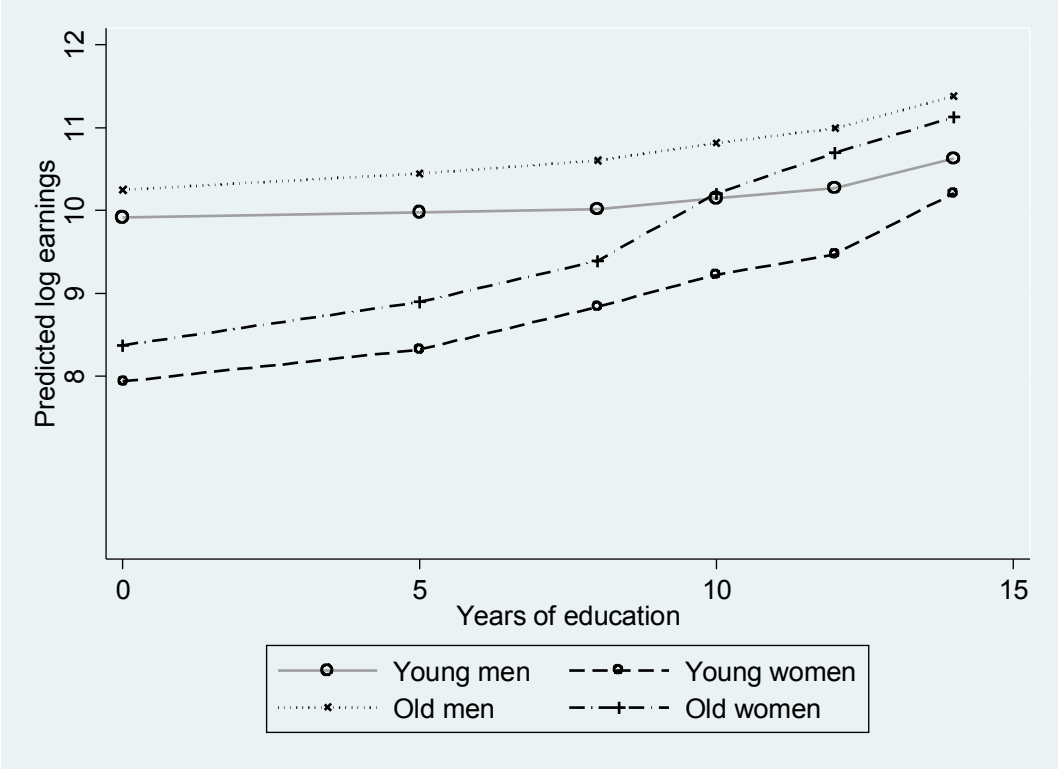


ii) Women



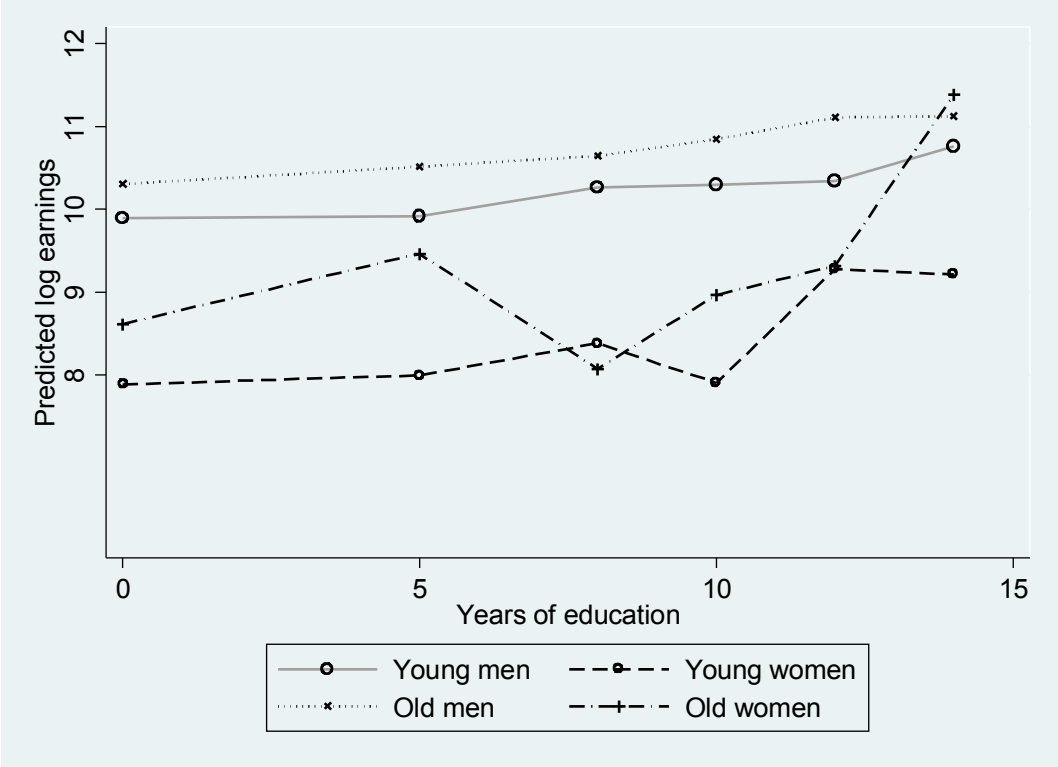
Note: These predictions are based on the multinomial logits reported in Appendix 2.

Figure 5B
Predicted earnings and level of education: Wage employed



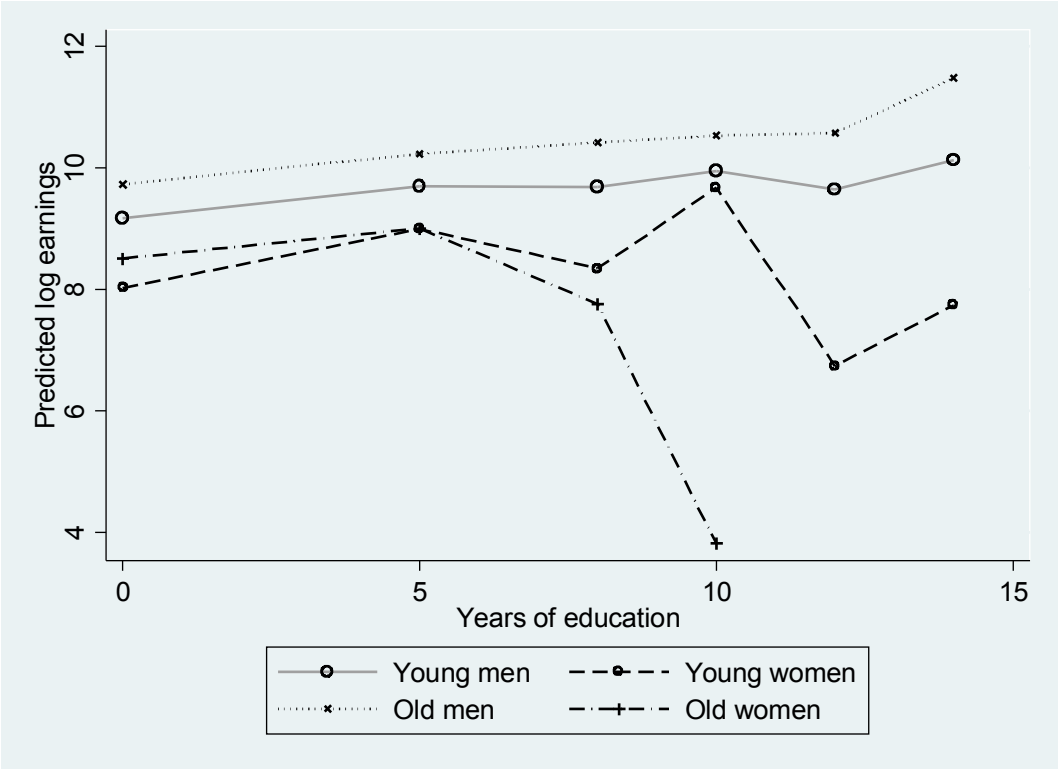
Note: These predictions are based on the results reported in Table 10B.

Figure 6B
Predicted earnings and level of education: Self employed



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

Figure 7B
Predicted earnings and level of education: Agriculture



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

Appendix 1
Additional Results, 1998/99

Table A1A
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 A2A
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 A3A
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 A4A
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 A5A
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 A6A
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 A7A
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 A8A
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 A9A
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 A10A
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 A11A
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 A12A
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 5 for information on how the education categories are defined.

Table A13A
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.

Appendix 2
Additional Results, 2001/02

Table A1B
Multinomial logit estimates. Omitted category: Wage employment. Young men.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	0.156	0.037	0.154	0.042
	(7.68)**	(1.86)	(4.99)**	(2.01)*
Education squared	-0.009	-0.008	-0.003	0.003
	(5.57)**	(4.50)**	(1.52)	(2.02)*
Age	0.066	0.014	0.167	-0.540
	(0.78)	(0.18)	(1.26)	(5.89)**
Age squared	-0.002	-0.001	-0.006	0.008
	(1.08)	(0.80)	(1.95)	(4.11)**
# of children in hh	0.054	0.085	0.009	0.052
under 12 years of age	(4.26)**	(7.31)**	(0.44)	(3.69)**
# of elderly in hh over	0.164	0.375	0.079	-0.002
65 years of age	(2.61)**	(6.56)**	(0.83)	(0.03)
Married	0.192	0.109	-0.984	-1.009
	(2.64)**	(1.56)	(7.39)**	(9.91)**
Observations	10653	10653	10653	10653

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

Table A2B
Multinomial logit estimates. Omitted category: Wage employment. Young women.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	(4.23)**	(1.60)	(1.80)	(7.11)**
	-0.027	-0.018	-0.008	-0.017
Education squared	(4.84)**	(3.16)**	(2.57)*	(9.33)**
	-0.178	0.055	0.039	-0.010
Age	(0.78)	(0.42)	(0.25)	(0.10)
	0.004	-0.002	-0.002	-0.001
Age squared	(0.89)	(0.78)	(0.69)	(0.30)
	0.012	0.033	-0.052	0.012
# of children in hh	(0.32)	(1.62)	(1.96)	(0.77)
under 12 years of age	-0.239	0.158	-0.256	0.072
# of elderly in hh over	(1.21)	(1.64)	(1.98)*	(0.97)
65 years of age	-0.504	0.617	0.680	0.701
Married	(2.29)*	(5.10)**	(4.93)**	(8.12)**
Observations	12423	12423	12423	12423

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

Table A3B
Multinomial logit estimates. Omitted category: Wage employment. Old men.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	0.177	0.042	0.146	0.080
	(10.49)**	(2.31)*	(3.48)**	(4.08)**
Education squared	-0.015	-0.013	-0.010	-0.007
	(11.22)**	(8.36)**	(3.35)**	(4.64)**
Age	-0.046	-0.093	-0.143	-0.109
	(1.71)	(3.83)**	(2.26)*	(3.35)**
Age squared	0.001	0.001	0.002	0.002
	(2.68)**	(5.56)**	(2.73)**	(7.62)**
# of children in hh	0.041	0.071	-0.055	0.088
under 12 years of age	(3.43)**	(6.54)**	(1.65)	(6.65)**
# of elderly in hh over	0.255	0.418	-0.028	0.065
65 years of age	(4.59)**	(8.13)**	(0.18)	(0.93)
Married	-0.056	-0.087	-1.103	-1.172
	(0.51)	(0.87)	(5.80)**	(11.75)**
Observations	12124	12124	12124	12124

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

Table A4B
Multinomial logit estimates. Omitted category: Wage employment. Old women.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Years of education	0.360	-0.019	0.201	0.280
	(4.81)**	(0.16)	(2.98)**	(9.25)**
Education squared	-0.035	-0.045	-0.028	-0.027
	(5.19)**	(2.31)*	(4.44)**	(12.13)**
Age	-0.228	0.006	-0.118	-0.169
	(2.71)**	(0.13)	(1.80)	(4.50)**
Age squared	0.002	0.000	0.001	0.002
	(2.64)**	(0.38)	(1.97)*	(5.66)**
# of children in hh	-0.032	0.034	-0.024	0.025
under 12 years of age	(0.80)	(1.65)	(0.85)	(1.48)
# of elderly in hh over	0.191	0.185	-0.291	0.034
65 years of age	(1.16)	(1.94)	(2.01)*	(0.46)
Married	-0.128	1.001	0.522	0.470
	(0.58)	(6.85)**	(2.86)**	(4.97)**
Observations	12541	12541	12541	12541

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

Table A5B
Multinomial logit estimates. Omitted category: Wage employment. Young men.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	0.271	0.095	0.202	-0.400
maths problems	(2.20)*	(1.06)	(0.98)	(3.42)**
Can read & write	0.513	-0.394	0.882	0.724
	(6.60)**	(6.05)**	(6.67)**	(8.20)**
Age	0.070	-0.007	0.230	-0.482
	(0.83)	(0.09)	(1.74)	(5.31)**
Age squared	-0.002	-0.001	-0.007	0.008
	(1.17)	(0.67)	(2.29)*	(3.77)**
# of children in hh	0.053	0.088	0.002	0.047
under 12 years of age	(4.21)**	(7.58)**	(0.09)	(3.38)**
# of elderly in hh over	0.161	0.370	0.089	0.015
65 years of age	(2.55)*	(6.49)**	(0.94)	(0.20)
Married	0.192	0.139	-1.053	-1.066
	(2.65)**	(1.99)*	(7.94)**	(10.50)**
Observations	10653	10653	10653	10653

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

Table A6B
Multinomial logit estimates. Omitted category: Wage employment. Young women.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	0.469	0.058	-0.828	-0.367
maths problems	(1.89)	(0.46)	(5.24)**	(3.58)**
Can read & write	-0.431	-2.086	0.232	-0.003
	(2.13)*	(14.58)**	(1.56)	(0.03)
Age	-0.348	-0.096	-0.035	-0.130
	(1.54)	(0.74)	(0.23)	(1.39)
Age squared	0.007	0.001	-0.001	0.001
	(1.51)	(0.21)	(0.32)	(0.74)
# of children in hh	0.035	0.055	-0.034	0.032
under 12 years of age	(0.93)	(2.67)**	(1.30)	(2.02)*
# of elderly in hh over	-0.235	0.170	-0.231	0.089
65 years of age	(1.19)	(1.76)	(1.78)	(1.22)
Married	-0.416	0.715	0.756	0.786
	(1.91)	(5.97)**	(5.49)**	(9.22)**
Observations	12423	12423	12423	12423

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

Table A7B
Multinomial logit estimates. Omitted category: Wage employment. Old men.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	0.419	0.251	-0.201	0.099
maths problems	(4.61)**	(3.56)**	(0.93)	(1.11)
Can read & write	0.070	-0.850	0.292	-0.049
	(1.10)	(14.70)**	(1.63)	(0.68)
Age	-0.040	-0.085	-0.135	-0.103
	(1.49)	(3.52)**	(2.14)*	(3.18)**
Age squared	0.001	0.001	0.002	0.002
	(2.52)*	(5.40)**	(2.59)**	(7.46)**
# of children in hh	0.049	0.082	-0.052	0.093
under 12 years of age	(4.09)**	(7.63)**	(1.58)	(7.00)**
# of elderly in hh over	0.236	0.384	-0.025	0.059
65 years of age	(4.29)**	(7.58)**	(0.16)	(0.84)
Married	-0.099	-0.131	-1.104	-1.190
	(0.91)	(1.32)	(5.80)**	(11.96)**
Observations	12124	12124	12124	12124

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

Table A8B
Multinomial logit estimates. Omitted category: Wage employment. Old women.

	1. Self employment	2. Agriculture	3. Unemployed	4. Out of labor force
Can solve simple	0.372	-0.174	-1.166	-0.451
maths problems	(1.76)	(1.61)	(8.01)**	(5.22)**
Can read & write	-0.293	-3.022	-0.155	-0.071
	(1.38)	(11.32)**	(0.82)	(0.84)
Age	-0.226	0.012	-0.093	-0.159
	(2.70)**	(0.24)	(1.42)	(4.27)**
Age squared	0.002	0.000	0.001	0.002
	(2.64)**	(0.27)	(1.59)	(5.42)**
# of children in hh	-0.016	0.050	0.003	0.043
under 12 years of age	(0.39)	(2.44)*	(0.09)	(2.58)**
# of elderly in hh over	0.162	0.164	-0.282	0.023
65 years of age	(0.99)	(1.74)	(1.95)	(0.33)
Married	-0.112	1.008	0.527	0.475
	(0.51)	(6.94)**	(2.89)**	(5.12)**
Observations	12541	12541	12541	12541

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

Table A9B
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	-0.019	0.068	0.026	0.013	0.121	0.356
	(3.32)**	(2.72)**	(1.19)	(0.16)	(3.33)**	(2.77)**
Education squared	0.004	0.006	0.002	0.004	-0.005	-0.035
	(9.92)**	(3.37)**	(1.09)	(0.63)	(1.58)	(2.16)*
Age	0.161	0.002	0.084	0.382	0.149	0.245
	(6.40)**	(0.02)	(0.87)	(1.26)	(1.01)	(1.03)
Age squared	-0.002	0.001	-0.001	-0.007	-0.002	-0.003
	(4.22)**	(0.40)	(0.32)	(1.13)	(0.54)	(0.68)
# Individuals	5246	963	1543	162	1782	887
B. Old						
Education	0.021	0.146	0.058	-0.024	0.093	0.593
	(5.13)**	(5.17)**	(3.93)**	(0.30)	(3.86)**	(2.61)**
Education squared	0.004	0.003	-0.000	0.008	-0.000	-0.107
	(12.15)**	(1.47)	(0.18)	(1.06)	(0.09)	(2.80)**
Age	0.070	0.158	0.024	-0.054	-0.002	-0.157
	(9.06)**	(4.39)**	(1.05)	(0.68)	(0.07)	(2.19)*
Age squared	-0.001	-0.002	-0.000	0.001	-0.000	0.002
	(8.76)**	(4.10)**	(1.52)	(1.02)	(0.09)	(2.05)*
# Individuals	5593	996	1927	166	2516	979

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 A10B
Earnings and years of schooling among the wage employed:
Quadratic specification, with household fixed effects

	Young Men	Young Women	Old Men	Old Women
Education	-0.042	0.010	0.009	0.059
	(4.16)**	(0.49)	(0.87)	(2.33)*
Education squared	0.004	0.007	0.003	0.006
	(5.68)**	(4.91)**	(3.34)**	(3.25)**
# Individuals	5246	963	5593	996

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 A11B
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	0.056	0.385	0.030	0.108	0.530	0.971
	(2.15)*	(2.86)**	(0.31)	(0.35)	(3.83)**	(3.06)**
Middle school	0.100	0.900	0.375	0.496	0.514	0.323
	(3.51)**	(5.05)**	(3.68)**	(1.59)	(3.22)**	(0.46)
Secondary	0.231	1.288	0.406	0.018	0.785	1.653
	(9.46)**	(10.90)**	(4.37)**	(0.05)	(5.37)**	(1.26)
Higher secondary	0.351	1.535	0.451	1.390	0.480	-1.281
	(9.81)**	(10.08)**	(3.53)**	(2.43)*	(1.89)	(0.84)
Tertiary	0.708	2.273	0.865	1.328	0.962	-0.282
	(19.43)**	(20.24)**	(5.45)**	(2.21)*	(2.80)**	(0.08)
Age	0.163	-0.007	0.087	0.396	0.153	0.264
	(6.45)**	(0.07)	(0.90)	(1.29)	(1.04)	(1.11)
Age squared	-0.002	0.001	-0.001	-0.008	-0.002	-0.004
	(4.25)**	(0.51)	(0.34)	(1.15)	(0.57)	(0.76)
# Individuals	5246	963	1543	162	1782	887
B. Old						
Primary	0.197	0.522	0.210	0.850	0.502	0.493
	(10.10)**	(3.18)**	(3.20)**	(2.90)**	(5.50)**	(1.48)
Middle school	0.355	1.021	0.339	-0.541	0.690	-0.754
	(14.26)**	(3.93)**	(4.55)**	(1.49)	(6.27)**	(0.67)
Secondary	0.564	1.832	0.541	0.349	0.801	-4.691
	(27.87)**	(10.56)**	(7.48)**	(0.90)	(7.34)**	(2.07)*
Higher secondary	0.743	2.323	0.804	0.703	0.845	--
	(27.27)**	(23.32)**	(6.80)**	(1.10)	(3.70)**	
Tertiary	1.132	2.753	0.821	2.765	1.754	--
	(47.90)**	(38.76)**	(8.25)**	(3.47)**	(7.29)**	
Age	0.072	0.161	0.025	-0.058	-0.002	-0.159
	(9.18)**	(4.44)**	(1.10)	(0.73)	(0.07)	(2.22)*
Age squared	-0.001	-0.002	-0.000	0.001	-0.000	0.002
	(8.88)**	(4.14)**	(1.58)	(1.09)	(0.10)	(2.09)*
# Individuals	5593	996	1927	166	2516	979

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 A12B
Earnings and the level of schooling among the wage employed: Controlling for household fixed effects

	Young Men	Young Women	Old Men	Old Women
Primary	-0.079	0.025	0.217	0.259
	(1.79)	(0.25)	(4.39)**	(2.01)*
Middle school	-0.069	0.475	0.152	0.300
	(1.47)	(3.86)**	(2.34)*	(1.10)
Secondary	0.015	0.872	0.322	1.236
	(0.34)	(8.46)**	(5.44)**	(9.03)**
Higher secondary	0.092	1.229	0.433	1.684
	(1.44)	(10.03)**	(5.07)**	(10.49)**
Tertiary	0.328	1.692	0.743	2.145
	(4.63)**	(15.50)**	(10.53)**	(19.92)**
# Individuals	5246	963	5593	996

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 5 for information on how the education categories are defined.

Table A13B
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	0.073	-0.038	0.058	-0.164
	(1.14)	(0.40)	(0.97)	(2.04)*
Can read & write	-0.060	0.728	0.199	1.242
	(1.40)	(10.12)**	(4.42)**	(15.76)**
# Individuals	5246	963	5593	996

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.