Journal of Quantitative Economics, Vol. 10 No.2, July 2012

RETURNS TO EDUCATION IN INDIA: SOME RECENT EVIDENCE

TUSHAR AGRAWAL[1]

Abstract

This paper estimates returns to education in India using a nationally representative survey. We estimate the standard Mincerian wage equation separately for the rural and urban sectors. To account for the possibility of sample selection bias, Heckman two-step procedure is used. The findings indicate that returns to education increase with the level of education and differ for rural and urban residents. Private rates of return are higher for graduation level in both the sectors. In general, the disadvantaged social groups of the society tend to earn lower wages. We find family background is an important determinant affecting the earnings of individuals. Using quantile regression method, we show that the effect of education is not the same across the wage distribution. Returns differ considerably within education groups across different points of the wage distribution. Returns to education are positive at all quantiles. The results show that the returns are lower at the bottom quantiles and are higher at the upper quantiles.

Keywords: Returns to Education; Wage Differential; Quantile Regression; India

JEL Classification: C13, I20, I21, J24, J31

1. Introduction

Whether to continue education beyond a certain level or to enter the labour market is an important investment decision. According to the human capital investment theory, an individual would prefer to attend school only if the present value of the expected benefits from schooling exceeds that of the expected costs (Becker 1964). Thus, an important determinant of the demand for schooling or training is its expected benefits. Since the benefits depend upon the quantity and quality of an individual's labour input, which in turn depends upon the human capital acquired during schooling, the education-wage relationship can be used to measure the returns to schooling.

[1] Indira Gandhi Institute of Development Research (IGIDR), Goregaon (E), Mumbai – 400065, India, Email: tushar@igidr.ac.in

I thank Sripad Motiram and S Chandrasekhar for comments and suggestions. I also thank P. Duraisamy for his useful comments on the introductory version of this paper. Comments on various aspects from an anonymous referee of this journal are gratefully acknowledged. An earlier version of this paper was presented at Institute of Economic Growth, Delhi. I thank seminar participants for useful discussions. Institutional support provided by IGIDR, Mumbai is gratefully acknowledged.

Investments in human capital (education) can be evaluated in terms of their rates of return. The estimation of rates of return to education is important for setting policy guidelines and evaluating specific programs. The estimates act as a useful indicator of the productivity of education and provide incentive for individuals to invest in their own human capital. While private rates of return are useful in explaining individuals' behavior in seeking education of different levels and types, social rates of return help in setting priorities for future educational investments. For example, what priority should be given to primary versus university or other levels of education? The comparison of profitability of human capital vis-à-vis physical capital can serve as a signal in guiding resource allocation between two forms of capital in developmental planning (Psacharopoulos 1985, 1994; Psacharopoulos and Patrinos 2004).

The purpose of this paper is to estimate the private returns to education in India using an earnings function approach. The paper provides recent evidence on these returns. We also examine the hypothesis of diminishing returns to education. The empirical analysis is based on a nationally representative household survey- India Human Development Survey (IHDS), which was conducted in 2004-05. We use the ordinary least squares (OLS) and quantile regression methods for the estimation purpose. The latter method provides a more comprehensive picture of the conditional wage distribution and allows for investigating the effect of education at different quantiles of the wage distribution. Since labour market conditions differ very much across the rural and urban sectors, the returns are estimated separately for the two sectors.

2. Literature Review

There is extensive literature on returns to education or schooling for both developed and developing countries. In the context of India, there are some studies based on nationally representative surveys (Duraisamy 2002; Dutta 2006; Kingdon and Theopold 2008; Madheswaran and Attewell 2007). Some other studies (Tilak 1987; Divakaran, 1996; Kingdon 1997, 1998) use small sample surveys and are confined to a particular district or state of the country. Quantile regression methods have been used widely in the developed nations primarily to examine the evolution of wage inequality. In India, these methods have been sparsely used, with two recent studies Chamarbagwala (2010) and Azam (2012) being exceptions. These two studies examine rural-urban inequality (in monthly per capita expenditure) and wage structure (in rural India), respectively.

In general, returns to education are higher for lower levels of education (e.g., primary) and decline with the level of education. This is due to the low cost of primary education relative to other levels of education and considerable productivity differentials between primary graduates and illiterate persons. Also, primary education provides the basis for further education. Social returns to education are lower than private returns because education is publicly subsidised in most countries and also due to the fact that estimates of social returns are not able to include social benefits of education. The rates of return to education vary significantly from country to country and also within a country over time. The returns are higher in the low-income (sub-Saharan African) and middle-income (Latin American/Caribbean) countries and are lower in the high-income (OECD) countries. This phenomenon could be due to differences in the relative scarcity of human to physical capital within each group of countries (Psacharopoulos 1985, 1994; Psacharopoulos and Patrinos 2004). Furthermore, returns differ across the wage distribution. The returns are higher for those who are in the top decile of the income distribution compared to those in the bottom decile. This may be due to ‘complementarity’ between ability and education; if persons with higher ability earn more the returns to those in the top deciles of the wage distribution would be higher (Harmon *et al*. 2003).

For India, Duraisamy (2002) estimates the returns to education by age-cohort, gender and location using the data from the National Sample Survey Organisation (NSSO) surveys. The study finds that private rates of return to education in India increase up to the secondary level and diminish afterwards. The rates of return per year of schooling in 1993-94 for the primary, middle, secondary, higher secondary and graduate levels of education are 7.9, 7.4, 17.3, 9.3 and 11.7%, respectively.[2] There are considerable gender and rural–urban differences in the returns. The returns at primary and secondary levels and for technical diploma are higher in rural areas than in urban areas. The returns at the middle, secondary and higher secondary levels are higher for women than that for men. The returns to women’s education are twice than that for men at the secondary level and are highest across all the educational levels. Further, the returns are higher for technical diploma as compared to college education particularly for men. An increase in the demand for highly qualified and technical persons, possibly because of the rapid industrialisation in the past decade, could explain the higher returns for higher secondary, technical diploma and other higher levels.

The returns to education also vary by the nature of employment or work contract. Dutta (2006) finds significant difference in the returns between casual and regular male workers using three rounds of the NSSO survey. While those in the former category face ‘flat’ returns, those in the latter category have positive and ‘U-shaped’ returns with respect to levels of education.[3] These patterns indicate that there is no incentive for casual workers to gain higher education (beyond primary schooling) whereas there is an incentive for regular workers to acquire higher levels of education. Dutta (2006) also finds evidence of changes in the returns to education over time (1983-1999) for regular workers and widening of the wage gap between graduation and primary education.[4] This has been attributed to trade liberalisation and other reforms that had taken place in India during the 1990s.

3. Estimating the Returns to Education: Some Empirical Issues

Private returns to schooling are usually estimated using the standard Mincer’s semi-logarithmic specification (Mincer 1974). The OLS estimation of the standard wage equation leads to biased estimates because of the unobserved ability and family background of an individual.[5] Ability of an individual may have an independent positive effect on earnings apart from the human capital variables usually accounted for by the amount of schooling accumulated and experience.

2 These results are based on OLS estimation. The joint maximum likelihood (JML) estimates are slightly higher for secondary and above levels of education.

3 The ‘U-shaped’ pattern means returns to primary level are low with regard to secondary and other higher levels, but higher than middle level of schooling.

4 For regular workers, the average returns to primary, middle and secondary schooling fell during 1983-1993 and the returns to graduate education increased during 1983-1993 and 1993-1999.

5 Another source of bias could be the presence of measurement error in either the earnings or education variable.

If an individual's ability and educational attainment are correlated, estimation of the wage equation would give biased results.[6] Ability may have contrasting effects on the returns. Individuals with higher ability are likely to convert schooling into human capital more effectively compared to the less able ones, and this in turn raises the returns for individuals with higher ability. On the other side if ability to progress in school is positively correlated with ability to earn this may reduce the returns; higher able persons may have been able to earn more in the labour market and due to higher opportunity cost in attending school, they may end up leaving the school earlier (Harmon *et al*. 2003).

Another problem could be due to omitting the individual's family background or social status. Parental education determines educational attainment of children and is highly correlated with children's schooling outcomes (Haveman *et al*. 1991; Card 1999). An individual's family background works in two ways: (i) by providing a better learning environment, and (ii) through better contacts or connections. Individuals belonging to more educated parents are more likely to get better information about employment and therefore obtain better paying or more secure jobs in the formal sector (Krishnan 1996; Siphambe 2000). In the literature, parents' education, father's occupation, household head's education and household income have been used to control for family background characteristics.

Ethnicity too has an external effect on human capital accumulation process (Borjas 1995). The earnings of children are not only affected by parental earnings but also by the mean earnings of the ethnic group in the parents' generation (ethnic capital). In India, caste is an important socioeconomic variable which also plays crucial role in determining earnings and occupation. The Indian society is divided into various caste divisions which represent a system of social stratification (Deshpande 2011).[7] This variable could also serve as an observable family characteristic.

The estimation of the above wage equation could also suffer from the problem of 'sample selection bias' if the wage functions are estimated using only the individuals who work and who therefore earn a wage. This might be a selective group and therefore not be a representative sample. A typical example is the women component in the labour supply. The OLS estimates in such a situation will be biased and inconsistent. To address this problem, estimation based on the method of maximum likelihood suggested by Heckman (1974) is usually applied.

One of the properties of OLS method is that the regression line passes through the mean of the sample. This method assumes that the regression coefficients are constant across the whole wage distribution and therefore the method can omit important features of the wage structure. Quantile regression methods allow us to examine the effect of each of the covariates along the entire wage distribution, thus give different parameter estimates at different points of the distribution. Quantile regression reduces sensitivity to outliers and enables us to examine how returns vary across different quantiles. In quantile regression not only the location but the shape of the wage distribution also changes (Buchinsky 1998).

---

6 Griliches (1977) explains the effect of omitting ability in the earnings function and argues that there is no good a priori reason to expect ability bias to be positive, it may turn out to be small or negative.

7 In India, the affirmative action programme is primarily caste-based. There are four broad groups (social groups): Scheduled Castes (SCs), Scheduled Tribes (STs), Other Backward Classes (OBCs) and Others. Among them the SCs and the STs are two socially disadvantaged social groups. SCs and STs have 15 and 7.5% reservation of seats respectively in government jobs and educational institutions.

4. Data

In this paper, we use the data from the India Human Development Survey (IHDS) 2005. The dataset is made available by the National Council of Applied Economic Research (NCAER), New Delhi, and the University of Maryland with particular focus on the issues related to human development. The IHDS is a nationally representative survey of 41,554 households in 1503 villages and 971 urban neighborhoods across India. These households include 215,754 individuals. The IHDS was conducted in all states and union territories of India except Andaman and Nicobar Islands, and Lakshadweep. These states include 384 districts, 1503 villages and 971 urban blocks located in 276 towns and cities. Villages and urban blocks form the primary sampling unit (PSU) from which the households are selected. Urban and rural PSUs are selected using a different design (Desai *et al.* 2010).

The survey has information on household characteristics: household residence (rural or urban), household size, membership of a social group, and religion; individual characteristics: age, education (number of standard years completed), gender, marital status and relation to the household head. A household belongs to one of the following social groups: Scheduled Castes (SCs), Scheduled Tribes (STs), Other Backward Classes (OBCs) and Others.

The survey also has information on occupation, industry, number of hours work in a usual day and wages and salaries of individuals, and the principal source of income for the household. The components of household income include farm income, income from interests (or dividend or capital gains), property, pension, income from other sources etc.

5. Methodology

Most studies on returns to education are based on the earnings function method, also known as human capital earnings function or 'Mincerian' method. An interesting aspect of Mincer's model is that the time spent during schooling is a key determinant of the earnings. The basic 'Mincerian' earnings function (Mincer 1974) is given as:

$$\ln w_i = \alpha + \beta s_i + \gamma_1 \exp_i + \gamma_2 \exp_i^2 + \varepsilon_i \qquad \text{... (1)}$$

where, *w* represents wage rate, *s* is the number of years of schooling completed, *exp* is years of labour market experience, exp^2 is experience squared, and ε is a random disturbance term capturing unobserved characteristics. In this function, the β coefficient on years of schooling can be interpreted as the average rate of return (or the percentage change in wages) to an additional year of schooling. The function assumes the rate of return to be the same for all levels of schooling. The experience variable is incorporated in the equation since an individual with higher experience in a job is likely to earn more. The experience squared term captures the possibility of a non-linear relationship between earnings and experience.

5.1 Econometric Specification

To take into account the sample selection bias, we use Heckman two-step procedure. The procedure involves two stages: in the first stage, a participation (selection) equation estimates the probability of having worked, and the second stage involves estimation of the wage (outcome) equation. It is necessary to find identifying variables (exclusion restrictions) that affect the selection equation but can be excluded from the wage equation.[8] The excluded variable should have a substantial impact on the probability of selection and should not be a determinant of the individual's earnings.

5.1.1 First Stage Probit Model

The first stage estimation, participation equation is given as:

$$y_i = z_i'\varphi + u_i \qquad \text{... (2)}$$

where, the dependent variable (y) takes a value of 1 if an individual participates in work and a value of 0 if not, z is a set of human capital, demographic and identifying variables, and u ~ N (0, σ^2_u). From the estimation of the participation equation, a selection variable (λ), known as the inverse Mills ratio, is created. The inverse Mills ratio is defined as the ratio of the probability density function to the cumulative distribution function of a distribution ($\widehat{\lambda_i} = \frac{\phi(z_i\widehat{\varphi})}{\Phi(z_i\widehat{\varphi})}$). This estimate is then used as an additional independent variable in the wage equation in the second stage.

Variables like non-labour (unearned) income of individuals or households, land ownership, number of dependent children, number of elderly persons and household size are used as identifying variables in the literature. In households with a large number of dependants (children), working age individuals especially women are more likely to accept flexible forms of work such as self-employment, informal or casual employment rather than wage work (Kingdon and Theopold 2008). Similarly individuals with land ownership and non-labour income are also less likely to attach with wage employment. However, land ownership could potentially be endogenous and correlated both with employment status and wages (Dutta 2006). In addition, it is not a good measure in the urban context. We use household size, number of children in a household and non-labour income of the individual or household as the exclusion restrictions. We expect negative signs on the variables household size and non-labour income whereas a positive sign on the variable number of children in a household.

5.1.2 Second Stage Wage Equation

The second stage involves estimating the wage function by ordinary least squares. Equation 1 can be extended by incorporating a series of dummy variables referring to the completion of education level in place of schooling variable s_i, to estimate returns at different levels. The second stage wage equation can be written as:

$$\ln w_i = \alpha + \sum_k \beta_{i,k} s_{i,k} + \gamma_1 \exp_i + \gamma_2 \exp_i^2 + \delta x_i + \theta\widehat{\lambda_\ell} + \varepsilon_i \qquad \text{... (3)}$$

where, $s_{i,k}$ represents a dummy variable for kth level of education, x is a set of other (demographic and family background) variables assumed to affect earnings, and ε ~ N (0, σ^2_ε). The equation also includes the inverse Mills ratio as an additional regressor obtained after the estimation of the first stage. This stage estimation is carried out only for the uncensored observations, i.e., only for those who participate in wage work.

---

8 In the absence of exclusion restriction, the sample selection problem cannot be addressed appropriately and the estimates of the returns cannot be used to make inferences for the entire population (Checchi 2006, pp. 202-203). If one allows all variables in the selection equation to also appear in the wage equation, the Heckman estimates become very imprecise (Wooldridge 2002, p. 565).

The dependent variable selected for the wage equation is the natural logarithm of the hourly wage.[9] Usually, it is difficult to get information on the actual labour market experience of each individual; therefore, potential experience is used as a proxy for the actual experience. The measure does not reflect labour market experience, rather the combined evolution of schooling and age (Machado and Mata 2001). A set of demographic and family background variables includes control for gender, marital status, social groups and household head's education. Some studies also use occupation of individuals and industry affiliation as control variables apart from other regressors. These studies examine wage determinants or wage inequality in a different context. Since our purpose is to estimate the returns to education, therefore, we do not control for occupation.[10]

By fitting such an earnings function, the average rate of return per year to each education level can be obtained by comparing the coefficients of the adjacent dummy variables:

$$r_k = (\beta_k - \beta_{k-1}) / \Delta n_k \qquad \text{... (4)}$$

where, β_k is the coefficient of kth education level, β_{k-1} is the coefficient of the previous education level, and Δn_k is the difference in years of schooling between kth and (k-1)th schooling levels.

5.2 Quantile Regression

The quantile regression method was introduced by Koenker and Basset (1978). The quantile regression model in form of a wage equation can be written as:

$$\ln w_i = x_i'\beta_\theta + \varepsilon_{\theta_i} \text{ with } \mathrm{Quant}_\theta(\ln w_i \mid x_i) = x_i'\beta_\theta \qquad \text{... (5)}$$

where θ is the specified quantile, x_i is a vector of the covariates, and $E[\varepsilon_{\theta i}|x_i]=0$.

Quantile regression minimises the weighted absolute values of the residuals.[11] One can assess the entire distribution by setting different quantile and get different parameter estimates of the conditional distribution of the dependent variable (wage rate). The method also allows to examine whether the effect of explanatory variables differ across the conditional wage distribution. The coefficients of the quantile regression can be interpreted conceptually in the same way as in the OLS regression. In this paper we are not able to address sample selection in the quantile regression case.

---

[9] The log transformation has various advantages: it reduces the effects of earnings outliers so that the distribution is closer to a normal distribution and is easier to interpret.

[10] Psacharopoulos and Patrinos (2004) suggest one has to be very careful while selecting the appropriate variables and interpreting the rate of return: "many researchers feel obliged to throw in the regression whatever independent variables they seem to have in the data set, including occupation. In effect, this procedure leads to stealing part of the effect of education on earnings that comes from occupational mobility. Of course, researchers who include occupation dummies in earnings functions do so because they are interested in modeling earnings, not necessarily in evaluating the rate of return to schooling. Obviously, such practice creates a problem when people other that the authors of these studies interpret the schooling coefficient as a Mincerian rate of return" (Psacharopoulos and Patrinos 2004, p. 116).

[11] OLS (ordinary least square) method, as the name indicates, involves the minimization of the sum of squared residuals. A special case of quantile regression is the least absolute deviation (LAD) estimator, which is obtained by fitting θ = 0.5 (median). LAD estimation is an appealing option when one believes that the median may be a better measure of location than the mean (Buchinsky 1998).

Due to data limitations, we are not able to control for ability of an individual. We also do not have any measure to control for quality of schooling. The analysis is based on the assumption that the quality of schooling is the same across the states as well as within the rural and urban sectors. Our estimates are restricted to wage earners and cannot be generalized to the entire population.

The wage distribution is trimmed by 0.1% at the top and bottom tails of the distribution to eliminate the possibilities of outliers. The analysis of the paper is restricted to individuals aged 15 and 65, since this group matches well with the labour force. Appendix I gives the description of variables used in the estimation. Table 1 gives the mean and standard deviation for the variables used in the analysis for the OLS and Heckman estimation based on the IHDS data.

Table 1. Mean and Standard Deviation of Variables used in Estimation

Variables	*OLS*			*Heckman*		
	All	*Rural*	*Urban*	*All*	*Rural*	*Urban*
Log Hourly Wage	2.148	1.909	2.653	2.148	1.909	2.653
	(0.82)	(0.68)	(0.85)	(0.82)	(0.68)	(0.85)
Hourly Wage	12.675	8.990	20.458	12.675	8.990	20.458
	(14.62)	(9.80)	(19.31)	(14.62)	(9.80)	(19.31)
Work Participation	-	-	-	0.467	0.542	0.362
Educational Level						
Illiterate & Below Primary	0.391	0.478	0.207	0.326	0.416	0.199
Primary	0.143	0.153	0.121	0.127	0.141	0.108
Middle	0.150	0.145	0.158	0.158	0.157	0.161
Secondary	0.164	0.139	0.216	0.203	0.175	0.243
Higher Secondary	0.065	0.047	0.102	0.095	0.069	0.132
Graduate	0.087	0.037	0.194	0.091	0.043	0.158
Age	35.944	35.493	36.896	33.494	33.354	33.691
	(11.8)	(12.09)	(11.39)	(13.90)	(14.05)	(13.68)
Experience	25.552	26.295	23.980	22.343	23.454	20.784
	(13.38)	(13.69)	(12.54)	(15.94)	(16.24)	(15.36)
Experience squared	831.794	878.894	732.305	753.216	813.956	667.966
	(771.55)	(803.45)	(688.93)	(880.44)	(924.69)	(806.63)
Female	0.273	0.313	0.189	0.504	0.484	0.532
Urban	0.321	-	-	0.416	-	-
Married	0.827	0.834	0.812	0.702	0.714	0.685
Social Groups						
Others	0.244	0.184	0.372	0.314	0.240	0.419
OBC	0.381	0.378	0.389	0.384	0.390	0.376
SC	0.258	0.287	0.196	0.218	0.252	0.170
ST	0.116	0.151	0.043	0.084	0.118	0.035
Household Head Education						
Illiterate & Below Primary	0.589	0.672	0.415	0.434	0.550	0.271
Primary	0.173	0.167	0.184	0.169	0.180	0.154
Middle	0.097	0.080	0.135	0.128	0.109	0.154
Secondary & Higher Sec.	0.109	0.071	0.188	0.198	0.135	0.286
Graduate	0.032	0.010	0.078	0.071	0.026	0.135
Household Size	-	-	-	5.958	6.214	5.597
				(2.93)	(3.13)	(2.58)
Number of Children	-	-	-	1.668	1.874	1.379
				(1.68)	(1.79)	(1.46)
Non Labour Income	-	-	-	0.047	0.023	0.082
Observations	46965	31875	15090	99900	58336	41564

Notes: The sample consists of individuals aged 15-65 in IHDS (2005) data. Standard deviation in parentheses and not reported for dummy variables. Refer Appendix I for a description of variables.

6. Estimates and Discussion

6.1 Estimates of the Augmented Mincer Function

Table 2 presents the OLS and Heckman estimates of the augmented wage equation. The selectivity term (inverse Mills ratio) is statistically significant indicating that sample selection could be a problem and therefore the OLS estimates will be biased and inconsistent. A positive inverse Mills ratio indicates that a shock to the selection equation that increases labour force participation also increases the conditional expectation of wages (Arrazola and Hevia 2008). The exclusion restrictions selected for identifying the selectivity term are statistically significant and have expected signs on all the coefficients which suggest these are the reasonable identifying variables.

All the variables in the wage equation, except marital status, are statistically significant at the 1% level of significance. The coefficients of all education dummies are positive and size of the coefficients increase with educational levels. This indicates a convex-shaped relationship between wages (log hourly) and educational level. There is a substantial earnings difference among persons with different educational levels. For example, an individual with primary education earns about 18% higher than an illiterate or individual with less than primary education.[12] The magnitude of the coefficients differs substantially between the rural and urban sectors. For example, an individual with primary education in the rural sector earns 15% higher than a person with no or below primary education whereas in the urban sector an individual with the same level of education earns 22% higher than those with no literacy and below primary schooling. Higher experience contributes to higher wages as confirmed by the presence of a positive sign on the coefficient. An additional year of experience increases the wages by 5%. A negative coefficient of experience squared shows that marginal returns from experience tend to decline over time. Our estimates indicate that wages are at the maximum level at 39 years of experience.[13] This maximum value of experience lies in our sample of individuals.

There is a substantial wage differential between males and females. Females earn 38% less than males. Another important dimension is the wage differential among the social groups. The estimates yield that STs, OBCs and SCs are likely to earn less by 14, 13 and 7%, respectively with reference to 'Others' category. In the rural sector, wages for STs, OBCs and SCs are significantly lower by 17, 14 and 8%, respectively than those for 'Others'. One notable point is that STs earn more than 'Others' category by 11% whereas OBCs and SCs earn less by 13 and 4%, respectively in the urban sector. This is perhaps due to relatively well-off STs group in the north-east states of the country.[14] When we drop these states from the analysis, we find that the STs earn less than 'Others' category in the urban sector too. This wage differential may be because these groups are associated mainly with those kinds of occupation which are low paid or

---

12 Since, the dependent variable is in the logarithmic form, the coefficient of dummy variable is adjusted by ($e^{coefficient}$ -1). See, Halvorsen and Raymond (1980) for the interpretation of dummy variables in a semi-logarithmic equation.

13 This can be computed as $\gamma_1/(-2\gamma_2)$ = [0.04902/ (2*0.00063)] using Equation 3 and Table 2.

14 The north-east states include Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim and Tripura. We find in our sample difference in the mean hourly wages for STs and 'Others' is about 10 rupees in the north-east states. In fact, for 'Others' category the mean hourly wage is lowest among all the groups in these states.

they are paid lower wages than their 'Others' counterpart. There could also be discrimination in the labour market.

Table 2. OLS and Heckman Estimates of the Wage Equation

Variables	*OLS*			*Heckman*		
	All	*Rural*	*Urban*	*All*	*Rural*	*Urban*
Human Capital Variables:						
(Ref: Illiterate)						
Primary	0.173***	0.152***	0.206***	0.164***	0.139***	0.198***
	(0.009)	(0.010)	(0.019)	(0.009)	(0.010)	(0.020)
Middle	0.356***	0.323***	0.394***	0.349***	0.313***	0.384***
	(0.009)	(0.011)	(0.019)	(0.009)	(0.011)	(0.019)
Secondary	0.584***	0.531***	0.649***	0.576***	0.519***	0.639***
	(0.010)	(0.011)	(0.018)	(0.010)	(0.012)	(0.019)
Higher Secondary	0.824***	0.721***	0.936***	0.820***	0.711***	0.932***
	(0.013)	(0.017)	(0.023)	(0.013)	(0.017)	(0.023)
Graduate	1.285***	1.195***	1.371***	1.296***	1.204***	1.386***
	(0.013)	(0.019)	(0.021)	(0.013)	(0.019)	(0.022)
Experience	0.044***	0.038***	0.052***	0.049***	0.047***	0.058***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.002)	(0.003)
Experience squared	-0.001***	-0.000***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Demographic Variables:						
(Ref: Male)						
Female	-0.420***	-0.429***	-0.406***	-0.478***	-0.522***	-0.500***
	(0.007)	(0.007)	(0.014)	(0.016)	(0.016)	(0.047)
(Ref: Unmarried)						
Married	-0.041***	-0.075***	0.036**	-0.021**	-0.036***	0.060***
	(0.009)	(0.011)	(0.017)	(0.011)	(0.013)	(0.020)
(Ref: Others)						
OBC	-0.147***	-0.156***	-0.142***	-0.141***	-0.139***	-0.136***
	(0.007)	(0.009)	(0.012)	(0.007)	(0.009)	(0.013)
SC	-0.084***	-0.108***	-0.046***	-0.074***	-0.083***	-0.037**
	(0.008)	(0.009)	(0.015)	(0.008)	(0.010)	(0.016)
ST	-0.174***	-0.232***	0.084***	-0.156***	-0.191***	0.103***
	(0.010)	(0.011)	(0.027)	(0.011)	(0.013)	(0.028)
(Ref: Rural)						
Urban	0.354***	-	-	0.343***	-	-
	(0.006)			(0.007)		
Family Background Variable:						
(Ref: Head- Illiterate)						
Head -Primary	0.040***	0.040***	0.045***	0.033***	0.026***	0.036**
	(0.008)	(0.009)	(0.015)	(0.008)	(0.009)	(0.015)
Head -Middle	0.059***	0.052***	0.077***	0.045***	0.026**	0.062***
	(0.010)	(0.012)	(0.017)	(0.010)	(0.013)	(0.019)
Head -Secondary	0.121***	0.092***	0.157***	0.099***	0.047***	0.133***
	(0.010)	(0.013)	(0.016)	(0.011)	(0.015)	(0.020)
Head -Graduate	0.368***	0.376***	0.373***	0.338***	0.306***	0.343***
	(0.018)	(0.032)	(0.024)	(0.019)	(0.034)	(0.028)
Intercept	1.649***	1.835***	1.707***	1.617***	1.738***	1.691***
	(0.017)	(0.020)	(0.032)	(0.019)	(0.025)	(0.033)
Mills Lambda	-	-	-	0.075***	0.152***	0.088**
				(0.020)	(0.024)	(0.043)
R-squared	0.490	0.339	0.446			
Wald Chi2	-	-	-	43661.00	14596.00	10419.00
Total Observations	46965	31875	15090	99900	58336	41564

Notes: Dependent variable is the natural logarithm of hourly wage. *, **, *** indicate significance levels at 10, 5 and 1% level of significance, respectively. Standard errors are in parentheses. Chi^2 statistics are significant at p-value less than 0.00. For Heckman model, only estimates of the wage equation are reported. Exclusion restrictions included in the Heckman model are household size, number of children in a household and non-labour income of the households. All the exclusion restrictions are statistically significant at the 1% level of significance. Refer Appendix I for a description of variables.

The control variable used to proxy for family background is positive and statistically significant. Other things remaining the same, increase in household head's educational level is directly associated with increase in hourly wages. For instance, having a household head with a graduate degree is associated with a 40% wage advantage compared to having an illiterate or below primary household head.

6.2 Private Rates of Return to Education

The private rates of return are computed using the estimates of the wage equation and presented in Table 3.[15] We find that the rates of return to education increase with educational level, i.e., returns are lower for primary level and higher for graduate level. The rates of return to education for the primary, middle, secondary, higher secondary and graduate levels are 5.5, 6.2, 11.4, 12.2 and 15.9% respectively.[16] These values are the marginal return to each extra year of education at that particular level. For example, rate of return to primary schooling can be interpreted as: each year of additional schooling after no schooling or below primary schooling would get 5.5% increase in wages for an individual who finishes primary schooling.

Table 3. Private Rates of Return to Education (%)

Educational Level	*OLS*			*Heckman*		
	All	*Rural*	*Urban*	*All*	*Rural*	*Urban*
Primary	5.75	5.07	6.87	5.47	4.64	6.59
Middle	6.11	5.69	6.25	6.15	5.80	6.20
Secondary	11.40	10.41	12.76	11.38	10.29	12.73
Higher Secondary	12.00	9.50	14.33	12.21	9.60	14.67
Graduate	15.38	15.80	14.52	15.87	16.43	15.12

Notes: The results are computed using Table 2. For example, private rate of return for middle level (using the Heckman) can be computed as: $r_{middle} = (\beta_{middle} - \beta_{primary})/\Delta n_{middle}$ = (0.349-0.164)/3= 0.061 or 6.15%. For primary level of education, Δn is taken as three years instead of five years.

The private rates of return differ between the rural and urban sectors. The results show that the rates of return for primary, middle, secondary and higher secondary are higher in urban areas whereas those for graduation are higher in rural areas. In both the sectors, the returns are lowest for primary education and highest for graduation level. There is a sharp rise in the returns after middle level of education in both the sectors. The difference in the rates of return between secondary and middle level are 5 and 6 percentage points in the rural and urban sectors, respectively.

---

15 Private rates of return to education are 'per year' returns to education. These are computed using Equation 4. Psacharopoulos (1994) mentions that primary school children do not forego earnings during their entire study-period, hence it is not advisable to assign them six (in our case five) years of forgone earnings. Therefore, for primary level of education, Δn is taken as three years instead of five years.

16 These estimates differ from those obtained when family background characteristics are not controlled for. In that case, the estimation of the wage equation seems to overestimate returns and for the corresponding levels, returns are 5.7, 6.3, 12.0, 13.2 and 18.0%, respectively.

Therefore, our findings do not support the hypothesis of diminishing returns to education. The results are in contradiction to studies (Psacharopoulos 1985, 1994; Psacharopoulos and Patrinos 2004) which show that returns decline by the level of schooling in developing countries. Recent evidence indicates that this pattern is changing; primary education is likely to be associated with lower wage returns that those accrued with other higher levels of education (Colclough *et al*. 2010). Some other studies, for example, Mwabu and Schultz (2000) for South Africa, and Siphambe (2000) for Botswana also find the increasing pattern of returns. The finding of low returns for primary education is also evidenced by studies of Duraisamy (2002) for women wage workers for India, Dutta (2006) for regular and casual male workers for India and Moll (1996) for South Africa.

Higher returns for the higher levels of education could be attributed to technological advancement and industrial structure change. These have resulted in an increasing demand of skilled workforce in the country. Graduate workers are paid considerably higher wages than their other counterparts. However, the demand has remained unfulfilled due to unavailability of skilled workforce (Agrawal 2012). As a consequence, wage inequality among skilled and unskilled workers has also risen in the past decade (Ramaswamy and Agrawal 2012).

Quality of schooling is another factor that can be attributed to low returns to primary education. Moll (1996) finds various qualitative factors such as: very high pupil-teacher ratios, poorly qualified teachers and low financing levels explain the low level of primary returns compared to secondary schooling in South Africa. Duraisamy (2002) also argues that the low returns to primary education in India may be due to the declining quality of primary education. A recent report of 'Pratham' shows that learning levels of children in rural India are not good. For instance, the report shows only 57.5 and 46.5% students in the standard III-V can read the standard I text or more and do subtraction or more respectively (Pratham 2012). Quality of education, apart from the other factors, could also explain the difference in the returns between the rural and urban sectors.[17]

Our results indicate that there is an incentive for individuals to achieve high levels of education. This result also has implication for the public funding on education. High and rising returns to tertiary education indicate that large public subsidies on higher education should be avoided (Colclough *et al*. 2010). Subsidies are needed only for the poor section of the society who in the presence of credit market failures faces difficulty. Though we do not control for household status (poor or rich) in our analysis, findings of the increasing returns with educational level may be linked to the status of a family. If private returns to education increase at higher levels of education, poorer families who generally educate their children at the primary level will face low returns whereas richer families who generally educate their children to secondary or beyond will face higher returns. As a result the poor families are motivated to invest less per child than the rich. Further, families would like to invest on education of those children who are more likely to reach a higher level to get higher returns. This may result in inequality between education and earnings, which may increase overtime both between families and within family (Schultz 2004).

[17] However, it may be noted that migrants in urban areas have had their education in rural areas and due to employment opportunities or some other reasons they have migrated to urban areas. In this case, the urban estimates would also reflect the school quality in rural areas (Duraisamy 2002).

6.3 Estimates of Quantile Regression

Quantile regression estimates across different quantiles (θ = 0.1, 0.25, 0.5, 0.75 and 0.9) are shown in Tables 4 and 5 for the rural and urban sectors, respectively. The positive coefficients on education dummies indicate that education has a positive impact on the wage distribution. However, the effect of education on wages differs across the wage distribution; the effect is smaller at lower quantiles and larger at higher quantiles. This suggests that education is relatively more valued for highly paid jobs. As a result it has a positive impact on wage inequality.

These findings are in accordance with studies of Blom *et al.* (2001) for Brazil, Hartog *et al*. (2001), and Machado and Mata (2001) for Portugal, Falaris (2004) for Panama, Martins and Pereira (2004) for many European countries, and Tansel and Bodur (2012) for Turkey. Higher returns at the top end of the wage distribution can be understood as education and ability are complementary. If the residuals in the wage regressions are interpreted as unobserved ability and returns increase across quantiles of the wage distribution, this indicates that schooling and ability are complements in enhancing worker productivity (Mwabu and Schultz 1996).

An interesting fact is that the effect of education on wages is lower in the rural sector than in the urban sector till the 75th quantile but for the top quantile the effect is higher in the rural sector. Specifically, in the first four quantiles for each educational level returns are higher (except graduation level in the 75th quantile) in the urban sector than in the rural sector, whereas in the 90th quantile returns for each educational level are higher in the rural sector.[18]

This could be possibly due to the following reasons. In the rural sector, individuals have low levels of education than those in the urban sector. One can also see high drop-out rates in schooling in rural areas. The individuals residing in the rural sector do not see much advantage or higher marginal returns to education in the labour market thus they prefer to drop-out at the early stages of schooling and start looking for work. As a result, a meager proportion of the individuals choose to continue for higher education. In addition, schooling infrastructure particularly at higher level of education is concentrated in urban areas this also limits participation of rural people in higher education. Consequently, controlling for other factors these individuals not only get much higher wage premium than their other rural counterparts who have the low levels of education but also likely to get high wage premium than their urban counterparts with the same levels of education. Further, a large proportion of graduates in urban areas are employed by the government, where wage-setting follows a very different process from the market, i.e., through the Government of India’s Pay commissions.[19]

One may also note that the wage dispersion (difference between spread of 90th and 10th quantiles) is higher in the rural sector and particularly it is highest for the higher secondary and graduation levels (Table 4). This indicates that the wage differential among the individuals with lower and higher levels of education is more pronounced in the rural sector. This also suggests that higher levels of education in the rural sector make substantial contribution to within-group wage inequality. The wage differentials between the rural and urban sectors are much larger at

---

[18] The average hourly wages for graduation level in the rural and urban sectors are Rs. 26.67 and Rs. 39.89, respectively.

[19] I am grateful to the referee for providing me this plausible explanation.

the top than at the bottom of the wage distribution. It has also been noted that in urban India the increase in the mean wages between years 1999-2000 and 2009-10 was heavily influenced by wage increases at the upper end of the wage distribution (Ramaswamy and Agrawal 2012).

Table 4. Estimates of Quantile Regression for Rural Sector

Variables	*q10*	*q25*	*q50*	*q75*	*q90*
Human Capital Variables:					
(Ref: Illiterate)					
Primary	0.046**	0.056***	0.086***	0.215***	0.313***
	(0.015)	(0.011)	(0.010)	(0.014)	(0.020)
Middle	0.143***	0.166***	0.232***	0.401***	0.559***
	(0.017)	(0.012)	(0.010)	(0.017)	(0.023)
Secondary	0.253***	0.269***	0.395***	0.678***	0.891***
	(0.020)	(0.012)	(0.014)	(0.020)	(0.024)
Higher Secondary	0.223***	0.329***	0.537***	1.054***	1.291***
	(0.032)	(0.020)	(0.030)	(0.034)	(0.038)
Graduate	0.535***	0.710***	1.196***	1.629***	1.716***
	(0.036)	(0.047)	(0.039)	(0.031)	(0.036)
Experience	0.018***	0.021***	0.027***	0.045***	0.053***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
Experience squared	-0.000***	-0.000***	-0.000***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Demographic Variables:					
(Ref: Male)					
Female	-0.353***	-0.438***	-0.479***	-0.426***	-0.438***
	(0.010)	(0.008)	(0.007)	(0.011)	(0.014)
(Ref: Unmarried)					
Married	(0.007)	-0.030*	-0.052***	-0.107***	-0.129***
	(0.018)	(0.013)	(0.011)	(0.014)	(0.022)
(Ref: Others)					
OBC	-0.159***	-0.146***	-0.144***	-0.163***	-0.146***
	(0.012)	(0.008)	(0.011)	(0.014)	(0.018)
SC	-0.095***	-0.081***	-0.095***	-0.106***	-0.120***
	(0.015)	(0.010)	(0.011)	(0.015)	(0.019)
ST	-0.238***	-0.255***	-0.216***	-0.240***	-0.189***
	(0.020)	(0.017)	(0.012)	(0.017)	(0.021)
Family Background Variable:					
(Ref: Head- Illiterate)					
Head -Primary	-0.034**	(0.016)	0.024*	0.058***	0.079***
	(0.011)	(0.009)	(0.010)	(0.014)	(0.018)
Head -Middle	0.003	0.015	0.036*	0.051**	0.074**
	(0.018)	(0.011)	(0.014)	(0.018)	(0.027)
Head -Secondary	0.000	0.025	0.089***	0.100***	0.139***
	(0.020)	(0.019)	(0.020)	(0.017)	(0.030)
Head -Graduate	0.125*	0.218*	0.381***	0.398***	0.471***
	(0.050)	(0.088)	(0.060)	(0.072)	(0.075)
Intercept	1.528***	1.861***	2.046***	1.992***	2.153***
	(0.034)	(0.026)	(0.022)	(0.031)	(0.040)

Notes: Dependent variable is the natural logarithm of hourly wage. *, **, *** indicate significance levels at 10, 5 and 1% level of significance, respectively. Bootstrap standard errors (with 600 repetitions) are in parentheses. *F*-test is carried out to check equality of coefficients on education dummies across quantiles. The test statistics $F(4, 31858)$ for primary, middle, secondary, higher secondary and graduate dummies are 50.04, 76.06, 177.51, 195.78 and 189.43, respectively. Refer Appendix I for a description of variables.

Table 5. Estimates of Quantile Regression for Urban Sector

Variables	*q10*	*q25*	*q50*	*q75*	*q90*
Human Capital Variables:					
(Ref: Illiterate)					
Primary	0.071**	0.121***	0.217***	0.273***	0.295***
	(0.025)	(0.021)	(0.023)	(0.022)	(0.035)
Middle	0.217***	0.284***	0.424***	0.505***	0.476***
	(0.026)	(0.022)	(0.023)	(0.021)	(0.031)
Secondary	0.360***	0.489***	0.704***	0.802***	0.776***
	(0.026)	(0.024)	(0.024)	(0.023)	(0.029)
Higher Secondary	0.549***	0.753***	1.037***	1.100***	1.049***
	(0.038)	(0.032)	(0.030)	(0.026)	(0.032)
Graduate	1.035***	1.291***	1.481***	1.496***	1.448***
	(0.041)	(0.034)	(0.028)	(0.026)	(0.032)
Experience	0.048***	0.054***	0.059***	0.051***	0.042***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Experience squared	-0.001***	-0.001***	-0.001***	-0.001***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Demographic Variables:					
(Ref: Male)					
Female	-0.632***	-0.557***	-0.442***	-0.304***	-0.205***
	(0.023)	(0.017)	(0.019)	(0.020)	(0.026)
(Ref: Unmarried)					
Married	0.120***	0.040	0.006	0.036	0.033
	(0.032)	(0.023)	(0.019)	(0.021)	(0.030)
(Ref: Others)					
OBC	-0.159***	-0.148***	-0.142***	-0.124***	-0.116***
	(0.021)	(0.018)	(0.016)	(0.015)	(0.022)
SC	-0.050*	-0.065***	-0.091***	-0.043*	-0.026
	(0.021)	(0.019)	(0.017)	(0.020)	(0.024)
ST	(0.079)	(0.011)	0.051	0.161***	0.179***
	(0.058)	(0.044)	(0.037)	(0.037)	(0.043)
Family Background Variable:					
(Ref: Head- Illiterate)					
Head -Primary	0.003	0.052*	0.044*	0.052**	0.035
	(0.024)	(0.021)	(0.019)	(0.019)	(0.028)
Head -Middle	0.041	0.089***	0.043*	0.085***	0.043
	(0.031)	(0.023)	(0.020)	(0.019)	(0.030)
Head -Secondary	0.104***	0.134***	0.122***	0.193***	0.208***
	(0.027)	(0.023)	(0.021)	(0.023)	(0.025)
Head -Graduate	0.290***	0.353***	0.373***	0.395***	0.344***
	(0.043)	(0.041)	(0.030)	(0.032)	(0.035)
Intercept	1.506***	1.600***	1.680***	1.870***	2.206***
	(0.054)	(0.043)	(0.040)	(0.041)	(0.050)

Note: Same as Table 4. The test statistics F(4, 15073) for primary, middle, secondary, higher secondary and graduate dummies are 13.12, 25.02, 58.89, 49.47 and 34.81, respectively.

Using the estimates of Tables 4 and 5, per year return to different educational levels across different quantiles are computed in Table 6. The rates of return to education are low for lower levels of education and high for higher levels of education. The rates of return within educational levels differ across the wage distribution. For primary, middle, secondary and higher secondary levels returns increase across the quantiles (except higher secondary level in the urban sector). For graduation level, rates of return across quantiles are of an inverted 'U shape'. This shows that the highest paid graduate workers possess lower returns than the lower paid graduate workers. Blom *et al.* (2001) also find returns for wealthier quantiles (75^{th} and 90^{th}) were lower than the less wealthy quantiles for tertiary education in Brazil.

Table 6. Per Year Quantile Rates of Return by Educational Level (%)

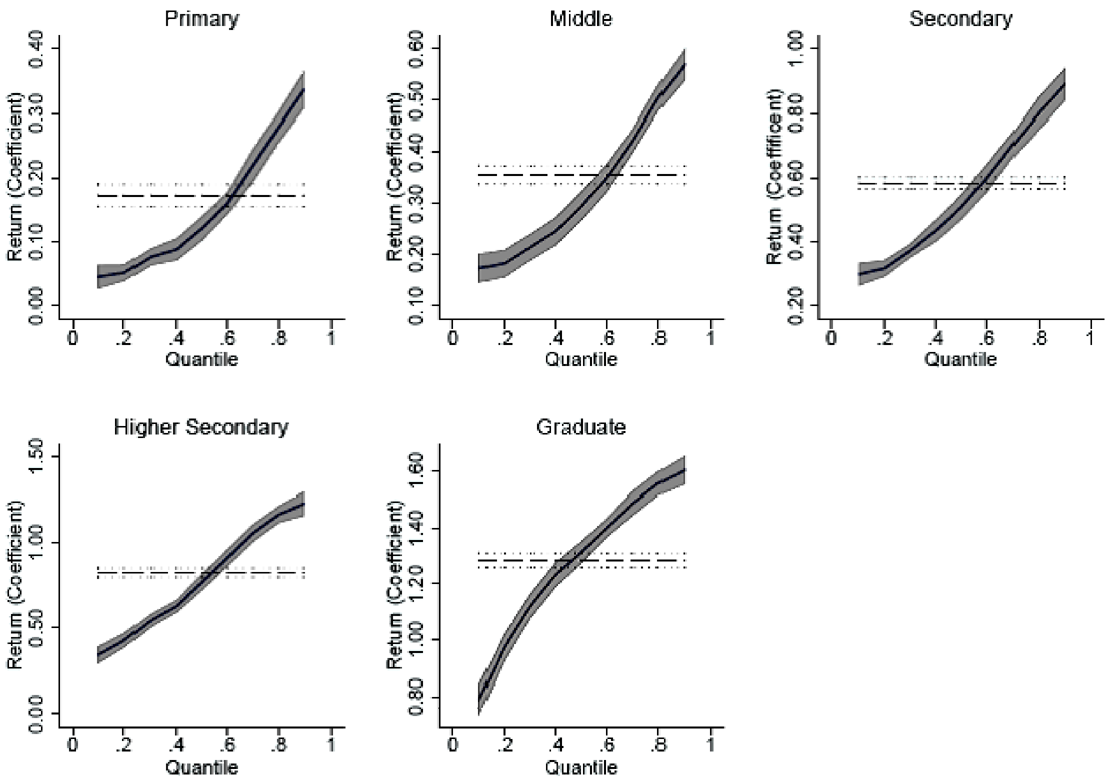
Educational Level	*Quantile Group*				
	0.10	*0.25*	*0.50*	*0.75*	*0.90*
Rural					
Primary	1.52	1.87	2.86	7.17	10.42
Middle	3.24	3.65	4.88	6.21	8.22
Secondary	5.50	5.15	8.16	13.86	16.58
Higher Secondary	-1.49	3.00	7.08	18.79	20.01
Graduate	10.41	12.71	21.96	19.16	14.18
Urban					
Primary	2.35	4.03	7.24	9.11	9.84
Middle	4.88	5.44	6.90	7.73	6.01
Secondary	7.16	10.25	13.99	14.87	15.03
Higher Secondary	9.45	13.19	16.68	14.86	13.65
Graduate	16.20	17.92	14.80	13.21	13.31

Note: The results are computed using Tables 4 and 5.

These results show that there is no location model; the slope coefficients and intercept term are not the same in the quantile regressions.[20] We also test hypothesis of equality of the regression coefficients (of education dummies) at different quantiles using *F*-test. The test statistics show that the null hypothesis of equality among the slope coefficients can be rejected at the 1% level of significance. Finally, we show a comparison of OLS and quantile regressions estimates for each level of education graphically in Figures 1. The figure confirms that both the mean and median regressions are quite different. The quantile regression estimates of each educational level lie outside the confidence intervals of the OLS regression. Quantile regression methods capture a large disparity along the wage distribution and in this manner these are quite helpful over OLS regression which assumes identical returns to education in the same education group.

Our results are based on a specific cross-sectional data, therefore, we cannot say much about change in patterns of the returns in India. However, recent evidence from other developing countries indicates that the concave pattern of returns to education does not hold more. Higher education is becoming scarce and workers with secondary level are becoming abundant (Mehta *et al*. 2011). There are some possible reasons which could be attributed to higher returns for higher education like increased openness to trade and foreign investment and introduction of new technologies which promote the demand for skilled labour especially those with higher education.

[20] If the model is truly a location model, then all the slope coefficients would be the same (Buchinsky 1998).



Note: In each figure, the dashed (horizontal) line and the continuous line show the OLS estimate and quantile regression estimates, respectively. The two dotted lines and the shaded region around the continuous line depict 95% confidence intervals for the two estimates.

Figure 1. Comparison of OLS and Quantile Regression Estimates by Educational Level

6. Conclusions

This paper estimates the private returns to education across different educational levels in India. After correcting for the possibility of sample selection bias, we find that the private returns to education increase with the level of education. The findings in the paper thus do not support the hypothesis of diminishing returns to education. The results for the earnings function show that there is a substantial earnings difference between males and females. Hourly wages of females are significantly lower than those of males. Across the social groups, wages for STs and OBCs are significantly lower than those of ‘Others’ in both rural and urban areas. Family background, as measured by household head’s education is an important explanatory variable in explaining the wage equation. This indicates it is important to identify individuals from poor family background and to support their education. We find omitting the family background characteristics may bias the returns upwards.

The increasing pattern of private rates of return suggests that for an individual, as a private decision, there is an incentive to invest at higher secondary and graduate levels. The increasing pattern of returns by level of education could be due to quality of schooling among the other reasons. One can expect that quality of schooling may be ameliorating as an individual ascends upwards in the educational hierarchy. Another reason which could explain this phenomenon is ability of the people. If people with higher ability attain more schooling then higher rates of return will be as a result of higher ability. However, we are not able to account for both these factors in our analysis.

Using quantile regression method, we analyse the returns at different points of the wage distribution. The returns to education differ along the wage distribution; the returns are higher at the upper end of the wage distribution. The returns to education within educational level also differ considerably. The rates of return increase for primary, middle, secondary and higher secondary levels across the wage distribution. For graduate workers per year returns are higher in the bottom quantiles. This shows that education is not rewarded in a uniform manner in the labour market.

References

Agrawal, T. (2012) “Vocational Education and Training in India: Challenges, Status and Labour Market Outcomes”, *Journal of Vocational Education & Training*, 64(4), 453-474.

Arrazola, M. and J. Hevia (2008) “Three Measures of Returns to Education: An Illustration for the Case of Spain”, *Economics of Education Review,* 27(3), 266–275.

Azam, M. (2012) “Changes in Wage Structure in Urban India, 1983–2004: A Quantile Regression Decomposition”, *World Development,* 40(6), 1135–1150.

Becker, G.S. (1964) *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, University of Chicago Press, Chicago.

Blom, A., L. Holm-Nielsen and D. Verner (2001) “Education, Earnings, and Inequality in Brazil, 1982-1998: Implications for Education Policy”, *Peabody Journal of Education,* 76(3/4), 180-221.

Borjas, G.J. (1995) “Ethnicity, Neighborhoods, and Human-Capital Externalities”, *The American Economic Review,* 85(3), 365-390.

Buchinsky, M. (1998) “Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research” *The Journal of Human Resources,* 33(1), 88-126.

Card, D. (1999) "The causal effect of education on earnings", in Ashenfelter, O. and D. Card (Eds.) *Handbook of Labour Economics,* North-Holland, Amsterdam.

Chamarbagwala, R. (2010) “Economic Liberalization and Urban–Rural Inequality in India: A Quantile Regression Analysis”, *Empirical Economics,* 39(2), 371–394.

Checchi, D. (2006) *The Economics of Education: Human Capital, Family Background and Inequality,* Cambridge University Press, Cambridge.

Colclough, C., G. Kingdon and H. Patrinos (2010) “The Changing Pattern of Wage Returns to Education and its Implications”, *Development Policy Review,* 28(6), 733-747.

Desai, S., A. Dubey, B.L. Joshi, M. Sen, A. Shariff and R. Vanneman (2010) *Human Development in India: Challenges for a Society in Transition*, Oxford University Press, New Delhi.

Deshpande, A. (2011) *The Grammar of Caste: Economic Discrimination in Contemporary India,* Oxford University Press, New Delhi.

Duraisamy, P. (2002) "Changes in Returns to Education in India, 1983–94: By Gender, Age-Cohort and Location", *Economics of Education Review,* 21(6), 609-622.

Dutta, P.V. (2006) "Returns to Education: New Evidence for India,1983–1999", *Education Economics,* 14(4), 431-451.

Falaris, E.M. (2008) "A Quantile Regression Analysis of Wages in Panama", *Review of Development Economics,* 12(3), 498–514.

Griliches, Z. (1977) "Estimating the Returns to Schooling: Some Econometric Problems" *Econometrica,* 45(1), 1-22.

Halvorsen, R. and R. Palmquist (1980) "The Interpretation of Dummy Variables in Semilogarithmic Equations", *The American Economic Review,* 70(3), 474-475.

Harmon, C., H. Oosterbeek and I. Walker (2003) "The Returns to Education: Microeconomics", *Journal of Economic Surveys,* 17(2), 115-155.

Hartog, J., P.T. Pereira and J.A.C. Vieira (2001) "Changing Returns to Education in Portugal during the 1980s and Early 1990s: OLS and Quantile Regression Estimators", *Applied Economics,* 33(8), 1021-1037.

Haveman, R., B. Wolfe and J. Spaulding (1991) "Childhood Events and Circumstances Influencing High School Completion", *Demography,* 28(1), 133-157.

Heckman, J.J. (1979) "Sample Selection Bias as a Specification Error", *Econometrica,* 47(1), 153-161.

Kingdon, G.G. (1997) "Labour Force Participation, Returns to Education, and Sex Discrimination in India", *The Indian Journal of Labour Economics,* 40(3), 507–526.

Kingdon, G.G. (1998) "Does the Labour Market Explain Lower Female Schooling in India?", *Journal of Development Studies,* 35(1), 39-65.

Kingdon, G.G. and N. Theopold (2008) "Do Returns to Education Matter to Schooling Participation? Evidence from India", *Education Economics,* 16(4), 329-350.

Koenker, R. and G. Bassett (1978) "Regression Quantiles", *Econometrica,* 46(1), 33-50.

Krishnan, P. (1996) "Family Background, Education and Employment in Urban Ethiopia", *Oxford Bulletin of Economics and Statistics,* 58(1), 167-183.

Machado, J.A.F. and J. Mata (2001) "Earning Functions in Portugal 1982-1994: Evidence from Quantile Regressions", *Empirical Economics,* 26(1), 115-134.

Madheswaran, S. and P. Attewell (2007) "Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey", *Economic and Political Weekly,* 42(41), 4146-4153.

Martins, P.S. and P.T. Pereira (2004) “Does Education Reduce Wage Inequality? Quantile Regressions Evidence from 16 Countries”, *Labour Economics,* 11(3), 355-371.

Mehta, A., J. Felipe, P. Quising and S. Camingue (2011) “Overeducation in Developing Economies: How Can We Test for It, and What Does it Mean?”, *Economics of Education Review,* 30(6), 1334– 1347.

Mincer, J. (1974) *Schooling, Experience and Earnings,* National Bureau of Economic Research, New York.

Moll, P.G. (1996) “The Collapse of Primary Schooling Returns in South Africa 1960–90” *Oxford Bulletin of Economics and Statistics,* 58(1), 185–209.

Mwabu, G. and T.P. Schultz (1996) “Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa”, *The American Economic Review, Papers and Proceedings,* 86(2), 335-339.

Mwabu, G. and T.P. Schultz (2000) “Wage Premiums for Education and Location of South African Workers, by Gender and Race”, *Economic Development and Cultural Change,* 48(2), 307-334.

Ramaswamy K.V. and T. Agrawal (2012) “Services-led growth, employment and job quality: A study of manufacturing and service sectors in urban India”, in Dev, S.M. (Ed) *India Development Report 2012-13*, Oxford University Press, New Delhi.

Pratham (2012) *Annual Status of Education Report (Rural) 2011,* Pratham, New Delhi.

Psacharopoulos, G. (1981) “Returns to Education: An Updated International Comparison”, *Comparative Education,* 17(3), 321-341.

Psacharopoulos, G. (1985) “Returns to Education: A Further International Update and Implications”, *The Journal of Human Resources,* 20(4), 583-604.

Psacharopoulos, G. (1994) “Returns to Investment in Education: A Global Update”, *World Development,* 22(9), 1325-1343.

Psacharopoulos, G. and H.A. Patrions (2004) “Returns to Investment in Education: A Further Update”, *Education Economics,* 12(2), 111-134.

Schultz, T.P. (2004) “Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education”, *Journal of African Economies,* 13 (supplement 2), ii95-ii148.

Siphambe, H.K. (2000) “Rates of Return to Education in Botswana”, *Economics of Education Review,* 19(3), 291-300.

Tansel, A. and F.B. Bodur (2012) “Wage Inequality and Returns to Education in Turkey: A Quantile Regression Analysis”. *Review of Development Economics,* 16(1), 107–121.

Tilak, J.B.G. (1987). *The Economics of Inequality in Education*, Sage Publications, New Delhi.

Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, London.

Appendix I. Description of Variables used in the Estimation

Variable	*Description*	*Base category*
	Explained (Dependent)Variables	
Log Hourly Wage	Natural Logarithm of hourly wages in rupees- Explained variable in the wage equation. Quantile Regression Wage Quantile: 0.1, 0.25, 0.5, 0.75, and 0.9	None
Work Participation	If an individual works more than or equal to 240 hours in a year, he/she is considered as part of the workforce - Explained variable in probit equation. [a]	If an individual works less than 240 hours in a year.
	Explanatory (Independent) Variables	
Human Capital Variables		
Educational Level (*5 dummies: Primary, Middle, Secondary, Higher Secondary, and Graduate)*	An individual belongs to one of the following educational level: Illiterate (includes literate with below primary also), Primary, Middle, Secondary, Higher Secondary, and Graduate. It is assumed that an individual spends 0, 5, 3, 2, 2 and 3 additional years, respectively in these educational levels.	Illiterate (and literate with below Primary)
Experience	Potential experience (proxy for the actual labour market experience) in years, defined as: Age-Years of schooling-5.[b]	None
Experience squared	Square of Experience	None
Demographic Variables		
Gender (*Female*)	Sex of individual: Male or Female	Male
Sector *(Urban)*	Place of residence: Rural or Urban	Rural
Marital Status	Marital Status of individual: Never married (unmarried), and Married (also includes Divorced, Widowed and others)	Unmarried
Social Group (*3 dummies: SC, ST and OBC)*	Each household belong to one of the following social groups: Scheduled Castes (SCs), Scheduled Tribes (STs), Other Backward Classes (OBCs) and Others.	Others
Family Background Variable		
Household Head's Education *(4 dummies: Head-Primary, Head-Middle, Head-Secondary, and Head- Graduate*	Education of household head, which is grouped as Illiterate (and literate with below primary), Primary, Middle, Secondary and Higher Secondary, and Graduate.[c]	Head-Illiterate
Exclusion Restrictions		
Household Size	Number of members in a household	None
No. of Children	Number of children (aged 0-14) in a household	None
Non Labour Income	Non labour income includes income from renting property and/or income from interest, dividends, or capital gains.	Household does not have non-labour income.

Note: Description of variables is based on IHDS (2005) data.

[a] The criterion for selection of 240 hours in a year is based on work participation measure used in the IHDS data. [b] It is assumed that an individual starts schooling at the age of five and starts working immediately after schooling. [c] In case, where household head him/herself is considered as an individual his/her father's education is taken as household head's education.

♦♦♦