

Returns to Education in India

Varun Kharbanda*

November 8, 2012

Abstract

An individual's choice of schooling level represents an equilibrium between the marginal rate of return to education and the marginal cost. To estimate the marginal rate of return at each schooling level, I use a Mincer type regression that is more general than in the standard literature, as well as a semiparametric approach. In light of the question raised by Heckman et. al. (1998) about the claim by Card and Krueger (1992) that the rate of return to education is constant, my results show that the rate of return to education varies non-monotonically across years of schooling in India. Based on Becker's (1967) theory of human capital accumulation, the non-monotonic marginal rates of return suggest the presence of heterogeneity in marginal costs among individuals seeking primary education. This heterogeneity in turn implies that a large fraction of Indian society faces a credit constraint for obtaining primary education. I also compare the returns to education using hourly wages versus annual wages and find that the use of hourly wages conceals the role of education in providing stable and secure employment.

1 Introduction

While the constitution of India considers education to be a fundamental right, the lack of commitment is evident in the heterogeneity of educational attainment across the population. The theory of human capital accumulation in Becker (1967) suggests that the choice of educational attainment is based on the intersection of the marginal rate of return and the marginal cost of education. The observed differences in educational attainment in India convey important information about the distribution of marginal costs and abilities. I link the trends in the marginal rates of return for different levels of education with the heterogeneity in the marginal cost of attaining education. These trends are identified by estimating the marginal rate of return to education at each schooling level,

*Currently a PhD Candidate at University of Iowa, Email:- varun-kharbanda@uiowa.edu.

Labor economists use the Mincer (1974) log-linear regression model to estimate the returns to education. Formally, the returns to education are represented by the slope of the years of schooling in a linear model with the log of earnings as the dependent variable. The Mincer model is based on the assumption that each additional year of schooling increases lifetime earnings at a constant rate. However, returns to education may vary with the level of educational attainment. I examine the linearity of the log of earnings with respect to years of education and use different approaches to capture their relationship in a Mincer type regression. I use micro level labor data on males from Employment and Unemployment Survey (EUS) for India in 2004–05 and the Current Population Survey (CPS) for USA in 2005.¹ My findings shows that the rate of return to education varies across years of schooling. There are also vast differences in the marginal costs for attaining primary education in India, which is reflected by the downward trend in the estimates of the marginal rate of return to education during primary school.

Earnings may be measured as hourly wages or as annual income. Using hourly wages is typically preferred as it removes the effect of market fluctuations and individual preferences for leisure. I analyze the effect of using annual verses hourly wages in estimating the rate of return to education.² The result show that, relative to annual wages, hourly wages understates the returns to education by hiding its role in providing stable and secure employment.

The contribution of education in economic growth as investigated by Krueger and Lindahl (2001) shows that the return to education is positive for small and medium size countries but close to zero or negative for countries which already have higher levels of education. This suggests that developing countries like India, where only 15% of the employed population have earned a high school degree or more, gain by investing in education. Knowing the returns to education at different levels of schooling, together with the economic theory explaining the observed trends behind these numbers, could help governments to effectively plan their educational expenditures.

The Mincer equation is a standard in estimating returns to education (Heckman et al., 2003b). Card and Krueger (1992) use micro data for different states in America and show that the relationship between earnings and edu-

¹The National Sample Survey Organization’s EUS 2004–05 data is used for India, and the March 2005 supplement of the CPS is used for the USA.

²Given the survey design of the NSSO unemployment and employment data, weekly earnings are calculated for each individual. Since no reliable variable is found to differentiate between hourly and annual wages, weekly wages are used to calculate hourly wages. The USA data is used to analyze the role of annual wages verses hourly wages in estimating rate of return to education.

cation is mainly log-linear. The standard Mincer regression provides a single estimate for returns to education. However, many studies use a dummy variable approach to find the non-linear relationship between years of education and the log of wages. The linearity of years of schooling in Mincer regressions is rejected by Cawley et al. (1998). Given the Becker (1967) type model discussed in Section 2, Mincer (1997) suggests that log earnings may be a convex or a concave function of schooling. Empirical work done by Deschênes (2001) suggests a rise in convexity between the log of earnings and schooling in 1980's and 1990's compared to 1970's. My analysis seeks to capture the full relationship between wages and schooling, which leads to a more general form of the Mincerian regression.

This paper starts with a basic Mincer style regression: a model in which log earning is linear in schooling and quadratic in potential experience³. I generalize it by adding higher order polynomials for years of schooling and potential experience. Chebychev polynomials transformation are used to avoid the problem of high correlation between the independent variable and its higher order polynomial values. To further ensure that the non-linearity of years of education in explaining wages is fully captured, I use a semiparametric method known as a partial linear model. This semiparametric method contains a non-parametric component of years of education along with a parametric component of the remaining variables in the standard Mincer regression to explain the log of wages.

In parametric regression, the function is specified and data are used to estimate the parameters of the specified functional form. Whereas in non-parametric regression, no functional form is pre-specified and the data estimate the functional form (Härdle, 1990). In my application, this ensures that there is no restriction on the functional form of the relationship between years of education and the log of wages. Estimates for the returns to education at different schooling levels are compared between different models for cross checking. While multiple models are used, the primary focus is to estimate returns to education for each schooling level without unduly restricting the relationship between years of schooling and the log of wages.

Recent evidence on the impact of credit constraints on education suggests an increasing role in its importance for household behavior (Lochner and Monge-Naranjo,

³A quadratic function of potential experience (defined as Age-years of education-6 in most of the studies) is often used in Mincer's log-linear model to capture the effect of on-the-job training on earnings, which is based on the assumption that it is hump shaped over time. However, Murphy and Welch (1990) find the use of a quartic function for experience, instead of quadratic, more appropriate for fitting a model on CPS data from 1964 to 1987. The use of a quartic function for experience is also validated on recent data by Lemieux (2006).

2011). Cameron and Heckman (1998, 1999); Carneiro and Heckman (2002) find little effect of family income on college enrollment using 1980's data. Belley and Lochner (2007) show that family income become a much more important determinant of college attendance over time. Cameron and Taber (2004); Keane and Wolpin (2001); Johnson (2010) estimate structural models to explore the importance of borrowing constraints for education related choices. While, Keane and Wolpin (2001) and Johnson (2012), using white male data from NLSY79 and recent male high school graduates data from NLSY97, respectively, find parental transfers and unobserved heterogeneity are important determinants of school. Cameron and Tabler (2004) studied male data from NLSY79 and find no evidence of discount rate heterogeneity. Stinebrickner and Stinebrickner (2008) use the direct survey of students enrolled at Berea college in Kentucky and find that the drop out rates are about 13% higher for 'constrained' students. Based on intergenerational data from Health and Retirement Survey, Brown et al. (2011) suggests that the total human capital investment is more sensitive to tuition subsidies among constrained youth. Although, these studies reinstate the role of credit constraints on education, my study identifies the stages of education that are impacted by credit constraints.

My research on India follows the work done by Duraisamy (2002) and Dutta (2006) which presents the estimates for returns to education for India using national level NSSO data for the period of 1983–1994 and 1983–1999, respectively. Other studies that estimate the returns to education are based on different data sources (Kingdon, 1998). Agarwal (2011) uses the India Human Development Survey (IHDS) 2005 data with quantile regression to examine the effect of education across the wage distribution. However, these studies use parametric Mincer regressions, using dummy variables for different schooling levels to capture the non-linear effect of schooling on education. The estimates in these studies ignore people who drop-out during primary schooling and are only available for a few specific schooling levels. These studies also included other variables like occupation, unionized worker, personal attributes in the regression which serves the purpose of modeling earnings instead of evaluating the rate of return on schooling (Becker, 1964). This paper uses NSSO data for 2004–05, and estimates the marginal rate of return to education at each schooling level for more years of schooling than are commonly reported. Selection bias into the workforce and measurement error in reported schooling are ignored as many studies found the overall bias in the estimates are negligible and are not statistically significant (Psacharopoulos and Patrinos, 2004; Ashenfelter et al.). Due to the lack of data, I ignore the tax rate and tuition fees in formulating my specification for the returns to education. My estimates for returns to education

are qualitatively comparable to estimates of Agarwal (2011).⁴

In the next section, I present Becker's theory (1967) on human capital accumulation and its relationship to schooling choice and the returns to education. Section 3 introduces the data and presents some preliminary findings for India. Since I also compare my results and test my empirical techniques on data from the USA, this section also describes the USA data. The different model specifications and an estimation procedure for the semiparametric model are discussed in Section 4. Results for India are presented in Section 5. Section 6 presents result for the USA and include a discussion on the use of hourly verses annual wages. Section 7 summarizes the findings with concluding remarks.

2 A Note on Becker's Theory of human Capital Accumulation

Becker (1967) associates the choice of human capital investment as the equilibrium between the marginal return on human capital (abilities) and the marginal rate of interest (opportunities) faced by that individuals. Thus, an individual's choice shows rational behavior in choosing level of education based on the cost and benefits of attending one extra year of education. At the aggregate level, this convey information about the aggregate demand and aggregate supply curves of the society.

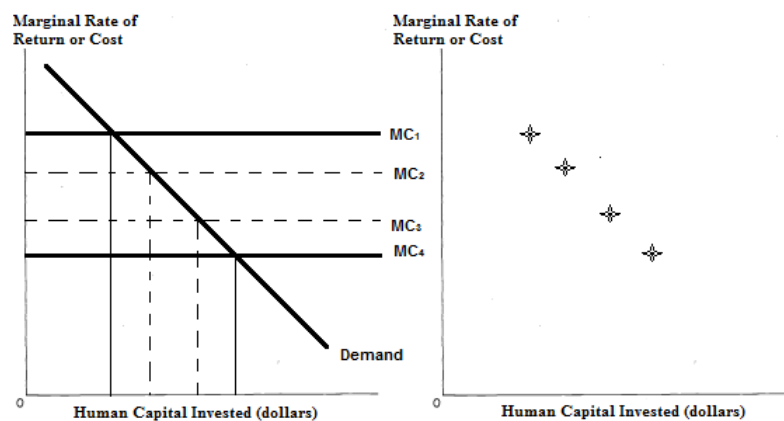


Figure 1: Equilibrium levels of investment in human capital resulting from differences in marginal cost curves. A negative correlation can be observed between marginal rate of return and amount of human capital invested . Source: Becker (1975).

The demand for education is derived from the ability of an individual in committing himself for each schooling level. This is decided by nature and fac-

⁴Check Table 9 in Appendix B

tors outside the control of individuals. The supply is presented by the cost of attaining each schooling level (along with tuition, fees, and the opportunity cost of time and energy). Since education and training are provided by the market, we can control the supply or marginal cost of attaining education to some extent. Different people have different sets of demand and supply curves and this gives them different equilibrium points. While the demand and supply curves are not observed from the data, we can observe the equilibrium points. The correlation of these points convey information about the positions of demand and supply curves of the society. Motivated by Becker's (1967) work, I show cases where the correlation between the marginal rate of return and years of education is positive, negative or not correlated and follow it with a discussion on the corresponding policy implications. Since the majority of the educational sector is funded by the government in most countries, I focus on policy implications.

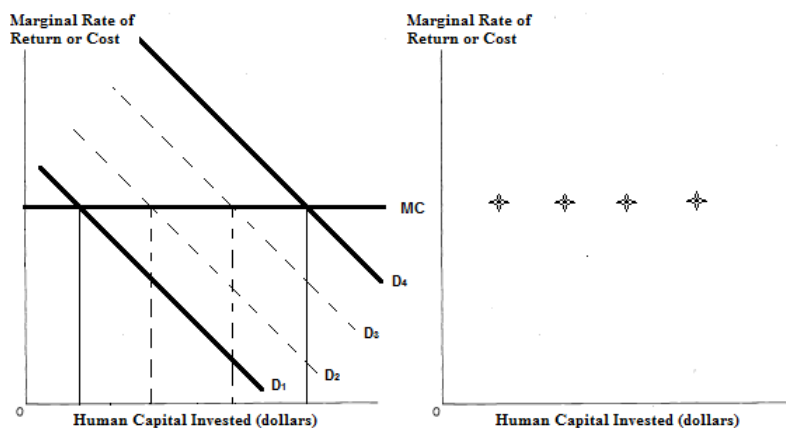


Figure 2: Equilibrium levels of investment in human capital resulting from differences in demand curves and horizontal marginal cost curve. A horizontal equilibrium points can be observed between the marginal rate of return and the amount of human capital invested. Source:Becker (1975).

When the cost of education differs across individuals, due perhaps to differences in access to credit markets, the rich face a lower cost of funds and thus finance education at a lower price. The poor face a higher cost of funds in the capital market and thus face higher cost of education. Given no heterogeneity of ability types in society, individuals who faces lower costs demand higher level of education than those who faces higher cost of education. This gives us a negative relationship between marginal rate of returns to education and schooling (as shown in Figure 1)(?). A similar result would imply a high degree of inequality in access to capital. For a government with a goal of improving human capital accumulation in the economy, some kind of intervention would be required to decrease the cost of capital faced by the poor so that they can

access education.

When individuals are heterogeneous with respect to ability but face no differences in the cost of education, high ability individuals demand higher levels of education and low ability individuals demand lower levels of education. As is shown in Figure 2, when society faces a horizontal marginal cost curve, there will be no correlation between the marginal rate of returns to education and years of schooling. A horizontal marginal cost curve may occur when all individuals face the same constant cost of funds over their lifetime. In practice, an

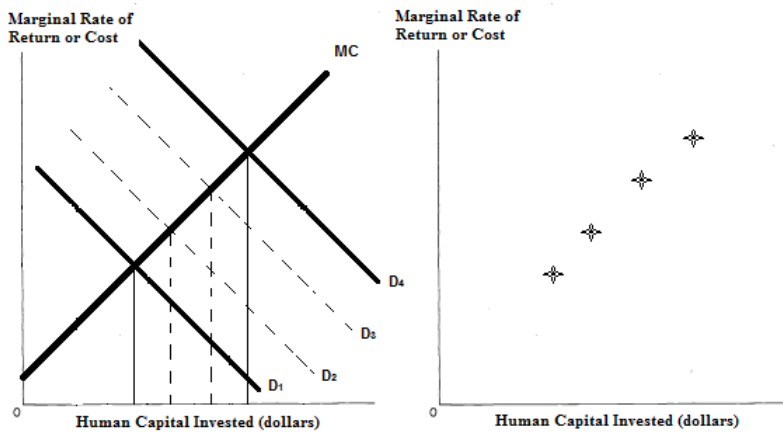


Figure 3: Equilibrium levels of investment in human capital resulting from differences in demand curves and increasing marginal cost curve. A positive correlation can be observed between the marginal rate of return and the amount of human capital invested. Source:Becker (1975).

increasing marginal cost curve for society is more common than a horizontal one. This results in a positive correlation for the case of heterogeneity in demand of education(as shown in 3). Since heterogeneity in ability is natural in a society, one should see a positive correlation as a natural case when all have access to education in the society and face an identical increasing marginal cost curve. The situation can be improved by reducing the cost of education for everyone by innovating new pedagogical techniques (e.g. online teaching).

One can argue for the case when both effects: heterogeneity in demand and heterogeneity in marginal cost curve, are present in the society. I try to graph this situation and found that one would see a fluctuations in equilibrium points, indicating no correlation(as shown in Figure 4). Since, it reflects the presence of both kind of heterogeneity, for the policy implications, a government have an option of easing out the credit constraints and work to provide access to education for all.

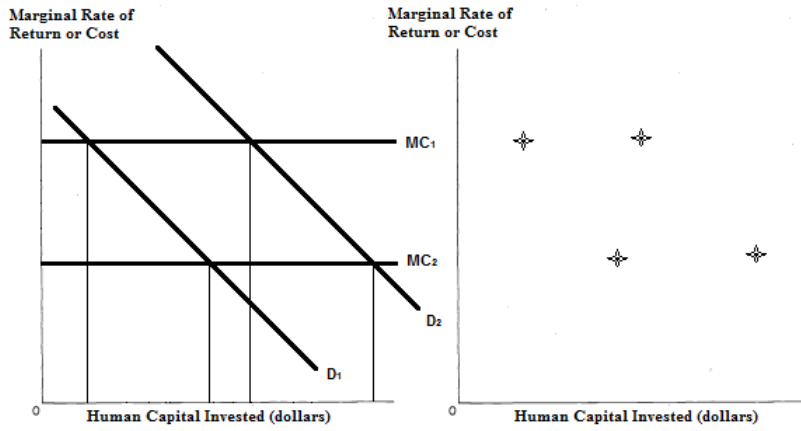


Figure 4: Equilibrium levels of investment in human capital resulting from differences in demand curves and marginal cost curves. No specific correlation can be seen when both effects are present. Even at average levels, these points are more likely to show no correlation. Source: Becker (1975).

3 Data

In India, the National Sample Survey Organization (NSSO) conducts surveys on employment and unemployment. I use the latest available data from the 2004–2005 survey. This survey encompasses over rural and urban areas and covers 124,680 households and 602,833 persons. Weights are used to replicate national level figures. For application of the Mincer equation, I focus on individuals age 16–64, who have positive wage income at the time of the survey⁵. Because literacy rates are low in India, government and non-government organizations run different programs to educate individuals. Since the most illiterates are adults, some programs provide informal education at the village level. NSSO data records the source of education for informally educated individuals, and for the more educated, the highest grade enrolled or completed. Analysis of the distribution of employed individuals by education in Table 1 shows that only 8% of the total working population are college graduates, 15% have passed high school, and 2.4% of the working population is reported to be informally literate.

⁵Agarwal (2011), Duraisamy (2002), Dutta (2006) are some of the studies that focus on the 16 to 64 age group.

Table 1: Distribution of Employed population in India by the highest level of educational attainment (based on NSSO Employment and Unemployment Survey 2004–05).

Education Levels	% of Employed Population	
	All	Male
Non-Literate	37.5	29.3
Literate without Formal Schooling	2.4	2.6
Below Primary School	8.9	9.6
Completed Primary Education	13.3	15.1
Completed Middle School	14.8	17.4
Completed Secondary School	8.1	9.5
Completed / High School	4.5	5.2
Diploma/Certificate Course	2.3	2.4
College Graduate	6	6.5
Postgraduate and Above	2.3	2.4

NSSO records all economic and non-economic activities an individual is engaged in over the week previous to the date of the survey. Because a person can be engaged in more than one job during the week, I focus on his total earnings for the week and use it to calculate hourly wages. The survey also identifies the primary job of each respondent which I use for identifying job type. To make the study comparable to similar studies in other countries, I base my study on the wage earning male population and exclude individuals with informal education, pensioners, the disabled, the unemployed, and those who are engaged in “non-economic” activities like prostitution, begging, and own household workers.⁶ Since the study is based on years of education and the measure we have in the NSSO survey is given by category, I need to assign a number of years that require to complete the given category of education. Observing the education system in India, I assign 0 years of school for illiterate individuals, 3 for below primary, 5 years if primary education has been completed, and so on until 18 years of school for post graduate and above.⁷ After all of the exclusions, I left with 61,473 observations that represent individuals across India.

⁶Jaeger and Pagé (1996), Card and Krueger (1992) Cameron and Taber (2004); Keane and Wolpin (2001); Johnson (2010) use male only data. I could not find any study that includes informal education in estimating the returns to education (informal education represents roughly 2.5% of India’s working population). Duraisamy (2002) only look at individual who are wage earners.

⁷The category for ‘post graduate and above’ is recoded as 18 years of schooling, which is a common length of time for post graduate work in India. The proportion of students going for higher degrees is generally very low, so on average the number of years required for ‘post graduate degree or above’ would be close to 18.

I define potential experience as $age - years\ of\ schooling - 6$.⁸ Given that 37.5% of the sample has no formal schooling and many individuals drop out of school early, I calculated the potential experience for individuals with less than 7 years of schooling as $age - 14$, because Indian Constitution prohibits children below the age of 14 years from working in any factory or mine or engaging in any other hazardous employment. By the year 2000, laws were enacted that made employing or facilitating in child labor a criminal act, punishable with a prison term. Given the poor enforceability of law in India, children from poor families usually start working at an age younger than 14. However, the experience accumulated before the age of 14 is not considered legal and on-the-job learning is also not high. This makes us use 14 years of age as the age when one can start accumulating any experience.

For the USA, I use the 2005 March supplement of the Current Population Survey (CPS). A number of individuals are excluded from the data such as Hispanics, Native Americans and Asian Americans as well as members of the armed forces. These exclusions were motivated by Heckman et al. (2003b) and Gelbach (2009). For analysis purposes, I structure the data similar to the NSSO data. To keep the standard of comparison at a similar level, I keep individuals who are in the full time work force, not working in the armed forces, ages 16 to 64, and focus on the male population who describe themselves as “White Only” or “Black Only” in the survey. To further clean the data, only individuals whose total wage and salary is greater than \$1, hours worked per week greater than 1, and weeks work in a year is between 1 and 52, are kept. After all the exclusions I left with 39,019 observations that are used for the analysis. CPS includes the Highest Grade attained for each respondent and, in order to keep it comparable to the analysis for India, the same number of grade levels are created for US schooling. Table 2 shows the distribution of employed population by highest education attained in USA, similar to Table 1 for India.

⁸In India, a child should be at least 5 years old before being admitted to primary school. Duraisamy (2002) uses 5 years as the minimum age to calculate potential experience. Previous literature on the USA uses 6 years as a minimum age to calculate the same (Jaeger and Pagé, 1996). To keep the study comparable and since 5 is a minimum age by law in India, I uses 6 years as the age to enter primary school, assuming that the average age would be greater than 5 and close to 6

Table 2: Distribution of the employed population in the USA by the highest level of educational attainment (based on CPS March Supplement 2005 data).

Education Levels	% of Employed Population	
	All	Male
" <1st Grade"	0.2	0.3
"1st to 4th Grade"	0.6	0.9
"5th or 6th Grade"	1.5	1.9
"7th and 8th Grade"	1.3	1.6
"9th Grade"	1.5	1.8
"10th Grade"	1.7	1.8
"11th Grade"	2.3	2.5
"12th Grade No Diploma"	1.1	1.4
"High School graduate- with diploma"	31.4	32.8
"Some College But no degree"	18.4	17.3
"Assc degree-occupation/vocation"	5.3	5.0
"Assc degree-academic program"	4.4	3.6
"Bachelor's degree"	20.2	19.3
"Masters degree"	7.5	6.6
"Professional school degree"	1.5	1.8
"Doctorate degree"	1.3	1.5

Some of the facts that we should know before comparing the results are about the differences in the labor market and the schooling system between India and the USA. In the USA, schooling is funded mainly by the government, and thus the law related to compulsory education in USA is implemented strictly. This is evident from the distribution of employed individuals based on their highest level of educational attainment as all but 10% of the employed population are high school graduate.

To keep the similarity in the methodology for USA and India, I combine the education levels of the USA to approximate the levels for India (as there are fewer levels in the NSSO data). The cluster on the USA education levels are build to reflect the same years of schooling that is required in India. Table 3 shows the mapping of the education level with years of education and potential experience for both India and the USA. The general rule in the USA, according to the Fair Labor Standards Act (FLSA), sets 14 years of age as the minimum age for employment, which is same as I use for India.

Table 3: Mapping of educational levels, as given in the NSSO survey, to the CPS data, years of schooling and potential experience.

Education Levels		Years of	Potential
India (NSSO Data)	USA (CPS Data)	Schooling	Experience
Non-Literate	"<1st Grade"	0	Age-14
Below Primary School	"1st to 4th Grade"	3	Age-14
Completed Primary Education	"5th or 6th Grade"	5	Age-14
Completed Middle School	"7th, 8th or 9th Grade"	7	Age-14
Completed Secondary School	"10th, 11th Grade or 12th Grade No Diploma"	10	Age-16
Completed High School	"High School graduate- high school diploma" or "Some College But no degree"	12	Age-18
Diploma/Certificate Course	"Assc degree-occupation/vocation" or "-academic program"	14	Age-20
College Graduate	"Bachelor's degree (BA,AB,BS)"	16	Age-22
Postgraduate and Above	"Masters degree (MA,MS, MENG, MED, MSW, MBA)" or "Professional school degree (MD, DDS, DVM, L)" or "Doctorate degree"	18	Age-24

New dummy variables are created to identify the Race/Class and the location for each individual. In the USA data, the variable "Black" is created with value 1 for individuals with race given as "Black Only" and 0 for individuals with "White Only"; other races are not included in the analysis. The location variable "Non-Metropolitan" identifies if the individual is from non-metropolitan or undefined metropolitan area (1) or a metropolitan area(0). For India, the dummy variable "Under Representative" is created to identify if the individual is from the Backward caste/class in the Indian caste/class system, (1 for backward caste/class individual and 0 for upper class/caste individual). Similar to "non-metropolitan" for USA, the "Rural Area" variable is created for India, taking the value 1 if the individual is from a rural area and 0 if the

individual is from an urban area.

Table 4: The summary (mean or percentage) of key variables used in the study for USA and India based upon the data used.

Variable	India	USA
Potential Experience	19.1	21.0
Years of Education	5.8	13.0
Log of hourly wages	2.05	2.84
% Rural or Non-metropolitan	67.1	16.2
% Under Representative or Black	71.5	10.6

Table 4 present the summary of key variables used in this study for both countries. The mean of potential experience, years of education and log of hourly wages are lower in India than in the USA. The population living in rural or non metropolitan areas and the size of discriminated social groups also reflect opposite situation in these two countries. The urbanization is under 35% in India and above 80% in the USA. The social group that faces discrimination is above 70% in India but around 10% in the USA. Higher numbers for average schooling years and potential experience reflect the different phases of economic development that the two countries are undergoing.

4 Model

Most models for estimating returns to education use the standard Mincerian earnings equation. Heckman et al. (2003a) summarizes the work done with Mincer regression. They also examine the importance of relaxing functional forms assumptions in Mincer regression and analyze its modifications by accounting for taxes and tuition. Their work is useful in understanding the fundamental basis of Mincer regression (Mincer, 1958) (Mincer, 1974). I contribute to this literature by generalizing the Mincer regression to obtain robust estimate for returns to education⁹. The focus of my study is on the returns to education for each schooling level. I first specify the Mincer regression and build upon it to compare the results of different modifications.

For estimating the returns to education, one needs to model the relationship between wages and years of schooling. It could be the case that some

⁹Although, Heckman et al. (2003a) uses an alternative non-parametric approach to estimate returns to schooling by regressing experience on the log of earnings for each schooling level separately, Gorodnichenko and Peter (2005) use the semiparametric method to construct counterfactual wage distribution for a cross-country comparison over the returns to education between Russia and Ukraine.

factors other than schooling explains wages. One need to ensure that these factors are controlled for and also keep the effect of schooling on wages intact from the effect of these other factors. The basic structure of the model is given by

$$\log(wages) = g(YED, PEX, X) + \varepsilon$$

Where YED =Years of Education
 PEX =Potential Experience
 X = Other factors

The function $g()$, if left unspecified along with the error distribution, represents a nonparametric regression. A commonly known issue with nonparametric regression is the “curse of dimensionality”. Ramsay and Silverman (2005) describe this “curse” as a combination of an increase in computation cost, a decline in the optimal rate of convergence, and potential problems with model identification. To overcome the “curse of dimensionality” some researcher limit the number of independent variables to one. Another suggestion in the literature is to assume that the $g()$ function is separately additive in each independent variable. In this case, the function $g()$ is specified as

$$g(YED, PEX, X) = f_1(YED) + f_2(PEX) + f_3(X)$$

where $f()$ represents the unspecified functional form.¹⁰ With the addition of some structure to the model and no pre-specified functional form for each independent variable, we move into the area of semiparametric models. This generalization helps us in preserving the non-linear relationship of each variable with the dependent variable and avoid some pitfalls of nonparametric estimation. Local linear regression and spline functions are some of the methods used in applied econometrics to estimate the unspecified function over each variable.

Horowitz (1998) mention some other semiparametric methods to resolve the “curse of dimensionality”. Single index models use a parametric specification to reduce the dimension of the model to a single index and then apply the nonparametric regression using that index. In terms of my model, this is given by the following specification of $g()$:

$$g(YED, PEX, X) = f(\beta_1 YED + \beta_2 PEX + \beta_3(X))$$

¹⁰For more on the generalized additive models see Hastie and Tibshirani (1986). A survey on nonparametric and semiparametric methods for economics is done by Delgado and Robinson (1992).

where $f()$ is a function to be specified by the nonparametric regression and the β 's are the parameters to be estimated. The parametric specification aggregates the effect of all the independent variables. The nonparametric model captures the non-linearity between the dependent variable and the aggregate effect. One other method that is frequently used in the empirical literature is a partial linear model, which uses an additive model with parametric restrictions on a limited number of independent variables. In my study, the partial linear model can be specified as

$$\log(wages) = f_1(YED) + \beta_2PEX + \beta_3(X)$$

where $f()$ is again an unspecified function and β are the parameters for the model. I will use the *partial linear model* to capture the non-linearity between years of education and the log of earnings.

The standard log-linear model suggested by Mincer for estimating the returns to education is shown below:

$$g(YED, PEX, X) = \beta_{10} + \beta_{11}YED + \beta_{21}PEX + \beta_{22}PEX^2 + \beta_{31}X \quad (1)$$

Higher degree polynomials can also be used to capture the non-linear effect of years of education. Some researchers suggest using this method for potential experience as well(Heckman et al., 2003b). In an additive model, the general polynomial specification for f_1 and f_2 is the following:

$$\begin{aligned} f_1(YED) &= \beta_{11}YED + \beta_{12}YED^2 + \dots + \beta_{1j}YED^j \\ f_2(PEX) &= \beta_{21}PEX + \beta_{22}PEX^2 + \dots + \beta_{2k}PEX^k \end{aligned}$$

Unfortunately, this specification will lead to a high degree of correlation between the exogenous variable and its higher degree polynomials, which can be problematic. To ensure I avoid the issue of high correlation and capture the linear and non-linear effects of education and experience in my model, I use the Chebychev's orthogonal polynomial transformation (of the first kind) over these

variables. The transformation is described below for T :

$$\begin{aligned}
T_0 &= 1 \\
T_1 &= T \\
T_2 &= 2T^2 - 1 \\
T_3 &= 4T^3 - 3T \\
&\dots = \dots \\
T_{n+1} &= 2T * T_n - T_{n-1}
\end{aligned}$$

To determine the appropriate number of polynomials to include, I check the value of the root mean square error (RMSE) of each model with different polynomial combinations up to order 10. I choose the model where the RMSE value is minimized and closest to the origin (1,1). In general this is the j^{th} order polynomial for years of education (YED) and k^{th} order polynomial for potential experience (PEX).

With these specifications I have following model.

$$\begin{aligned}
\log(wages) = & \beta_0 + \beta_{11}YED_1 + \beta_{12}YED_2 + \dots + \beta_{1j}YED_j \\
& + \beta_{21}PEX + \beta_{22}PEX_2 + \dots + \beta_{2k}PEX_k \\
& + \beta_3X + \varepsilon
\end{aligned} \tag{2}$$

where $j, k \in \{2, 3, 4, \dots\}$.

The returns to education will be different at each schooling level in Equation 2 because of the higher polynomial order for YED. The standard Mincer model given in Equation 1 gives only one estimate for all levels of education.

Since the polynomial model in Equation 2 has its own challenges (e.g. the choice of the number of polynomial terms to include and overfitting), I use a partial linear model which generalizes the Mincer regression further. This is shown below:

$$\begin{aligned}
\log(wages) = & \beta_0 + f_1(YED) + \beta_{21}PEX + \beta_{22}PEX_2 + \dots + \beta_{2k}PEX_k \\
& + \beta_3X + \varepsilon
\end{aligned} \tag{3}$$

The structure of the function $f_1()$, along with the other parameters of the model, are estimated using the data. The estimation of this semiparametric model is discussed later in more detail.

On other factors X , I have used different factors to control for social and geographic influences on wages. As suggested in Psacharopoulos and Patrinos (2004), factors related to work profile, such as occupation type or industry, are ignored as it steals the effect of education and should not be included in the model to get a Mincerian rate of return. In the literature, location as the external factor is used in Duraisamy (2000) to capture differences between the rural and urban economies in India. For the USA, Gabriel and Rosenthal (1999) show the importance of a location variable in a returns to education regression and suggested the use of Standard Metropolitan Statistical Area (SMSA) as a dummy variable to ensure that the estimates do not suffer from omitted variable bias. The other aspect of the Indian economy is the caste system. Most papers on returns to education use race as the variable to show discrimination in the job market. In India the situation is different; race discrimination is substituted with discrimination based on social class. Individuals from backward classes (low castes) are often considered to be discriminated against in the job market. The government of India recognizes this discrimination and work towards eliminating it by providing reservations or favorable conditions for these classes in education. I use a dummy variable to indicate if an individual belongs to one of these classes and refer to them as ‘Under Represented’.¹¹

The X takes the two variables $UR =$ Dummy for Under Represented or race(1 if under represented class or race, else 0)

$RA =$ Dummy for Rural Area or location(1 if rural area or location, else 0)

4.1 Estimating Semiparametric Regression

The estimation of a partial linear model includes an estimation of parameters for the parametric portion, and the estimation of parameters and the functional form of the non-parametric portion. In the literature, I find that a similar model is used by Tobias (2003) to estimate the returns on ability. He uses a two step method which separates the estimation for the parametric and non-parametric model. In the first step, OLS is used to estimate the parametric model by differencing out the nonparametric part. In my model, the first step of this process can be done by differencing observation $(i + 1)$ from i , where

¹¹Bhaumika and Chakrabarty (2009) cite that in 2005, religion is not significant in explaining wage differences in India. Madheswaran and Attewell (2009) suggest, that social class does have an effect on wage discrimination in India. For detail discussion check Abraham (2012)

$YED_i = YED_{i+1}$. This gives us following:

$$\begin{aligned} \log(wages_i) - \log(wages_{i+1}) &= \beta_{21}(PEX_i - PEX_{i+1}) + \beta_{22}(PEX_{i2} - PEX_{(i+1)2}) + \dots \\ &+ \beta_{2k}(PEX_{ik} - PEX_{(i+1)k}) + (f_1(YED_i) - f_1(YED_{i+1})) \\ &+ \beta_3(X_i - X_{i+1}) + \varepsilon \end{aligned}$$

Assuming the continuity of the function f_1 , the nonparametric part will cancel out from the model and OLS can be used to estimate the β values. These estimates are used in the second step to remove the effect of variables specified in the parametric model from the dependent variable. The residual is then used as the dependent variable in a local linear regression (a nonparametric regression technique) where YED is the independent variable.

I use different methods to estimate the partial linear models and found all methods, including Tobias (2003) produce similar qualitative results.¹² Here I discuss the result of Generalized Additive Models (GAM) since the SAS procedure used for this is readily available to researchers and it shows the linear and non linear part of non-parametric model separately. The SAS procedure Proc GAM is used to estimate the partial linear model. The model I estimated has the following form:

$$\begin{aligned} \beta_1(YED) &= \beta_{11}YED + s_{12}(YED) \\ \beta_2(PEX) &= \beta_{21}PEX_1 + \beta_{22}PEX_2 + \dots + \beta_{2k}PEX_k \\ \beta_3(UR) &= \beta_3UR \\ \beta_4(RA) &= \beta_4RA \end{aligned}$$

Here s_{12} is an unknown function that captures the non-linearity of YED over the log of wages.

The Proc GAM procedure uses a two loop iteration process referred to as Backfitting and Local Scoring algorithm. It starts with a loop using the Local Scoring algorithm by initializing an unknown function (s_{12}) and estimates some other values (such as the weights and the adjusted dependent variable). This loop has a nested loop, a backfitting algorithm, that uses these values, to get

¹²Other than the Tobias (2003) method, I estimated semiparametric model by using an iterative process between parametric and non-parametric regression until the until the change in root mean square error (RMSE) of the parametric model is less than 10^{-5} . I also try an additive model with an unspecified functional form for years of schooling and potential experience. A PROC GAM procedure in SAS is used to estimate a semiparametric model. The parameter estimates of each method give the same qualitative results with some methods show estimates at different scale than others.

an estimate of β and s_{12} . For my model, the inner loop runs until it fails to decrease or satisfies the convergence criterion over the following equation.

$$RSS = \frac{1}{n} \|\log(wages) - \beta_0^{(m)} - \beta_{11}^{(m)} YED - s_{12}^{(m)}(YED) - \beta_{21}^{(m)} PEX_1 - \beta_{22}^{(m)} PEX_2 - \dots - \beta_{2k}^{(m)} PEX_k - \beta_3^{(m)} UR - \beta_3^{(m)} RA\|^2$$

where m is the number of the iteration. Based on the estimated values of β from the weighted backfitting algorithm, the outer loop calculates a new set of weights and an adjusted dependent variable. The outer loop runs until the estimates satisfies the convergence criterion(which is 10^{-8}) or ceases to decrease.

For estimating the unknown function, I use the cubic smoothing spline as smoothers to find the function with two continuous derivatives. Cubic smoothing splines are the unique minimizer to the penalized least squares: a method to measure the fit of a function on the data using least square with a penalty on curvature. Proc GAM does the entire process and gives the final estimates for the following model as the output.¹³

$$\log(wages) = \beta_0 + \beta_{11} YED + s_{12}(YED) + \beta_{21} PEX_1 + \beta_{22} PEX_2 + \dots + \beta_{2k} PEX_k + \beta_3 UR + \beta_4 RA$$

5 Results and Analysis

In this section, I present and discuss the results for the three models specified in the previous section for India.¹⁴

5.1 Results

For the Polynomial Model (Equation 2), the appropriate degrees of polynomial are $j = 9$ and $k = 7$. This means we are using up to 9 degrees of polynomials for YED and 7 degrees of polynomials for PEX. With the high degree of polynomials in this model, there may be a problem of overfitting where the model starts explaining noise instead of the underlying relationship due to the high number of variables. Avoidance of this problem is necessary to get smoother curves for the marginal returns to education. To mitigate the problem, I use the Schwarz

¹³My results are robust to the choice of method for nonparametric estimation. For more details on Proc GAM refer to the SAS support documentation on Proc GAM.

¹⁴The derivation used for finding the returns to education for each model is given in Appendix A.

Baysian Criteria (SBC) for model selection with the stepwise function in SAS.¹⁵ The model selected by this procedure is also used as the base model for the semiparametric specification (for PEX, the semiparametric model only includes the polynomial degrees that are significant in the polynomial specification).

The measures of model fit and parameter estimates of variables other than YED and PEX are given in Table 5 with standard errors in parentheses. The adjusted R-square and error term shows that the Mincer specification decently explains the log of wages. The more generalized polynomial specification adds only 0.02 points in adjusted R-square value and reduces the error deviation by 3% in comparison to the standard mincer regression. A semiparametric specification reduce the error by <0.01% in comparison to polynomial specification. The comparison of coefficients other than PEX and YED, across different models have intuitive effects: rural location or Under Represented class effect wages negatively, and the value of these coefficient does not show much variation across models.

Table 5: Parameter estimates for variables other than PEX and YED, and key goodness-of-fit indicators for the different models discussed in Section 4 using the NSSO Data for India 2004–05.

Variables	Standard Mincer	Polynomial Mincer	Semiparametric
Intercept	1.3892 (0.0118)	1.2124 (0.0239)	1.1147 (0.0567*)
UR=Under Representative	-0.1585 (0.0064)	-0.1392 (0.0063)	-0.1394 (0.0149*)
RA=Rural Area	-0.4614 (0.0634*)	-0.4617 (0.0062)	-0.4617 (0.0148*)
Adj R-Square	0.4644	0.4837	
F-Value	10661.5	4114.3	
Model (Error)	49,208,716	47,431,409	47,426,655
DF (Error)	61,467	61,458	

Each j^{th} degree polynomial variable is multiplied by $10^{-(j-1)}$ to get standardized coefficient value. All estimates are significant at the 1% level.

* Multiplied by 10^{-2} .

Table 6 and 7 show the parameter estimates across models for YED and PEX, respectively. The comparing coefficients across the different models shows that potential experience and year of schooling have a non-linear relationship with the log of hourly wages. The presence of a high degree of polynomials for

¹⁵SBC method uses likelihood functions and penalizes the number of parameters in the model. The “stepwise” method of model selection in SAS ensures that a model does not over fit the data by checking the contribution of each variable in the model and its own significance. This method include and exclude the variable from the model at each stepwise, based on the user specified criteria for significance value of inclusion, exclusion and stay level, and the gain a variable brings in explaining the model.

YED in the polynomial model suggest a considerable amount of non-linearity between schooling and the log of wages. The linear estimate of YED in semi-parametric model mirrors the linear effect of YED in the Standard Mincer regression. However, the non-parametric portion is also significant in explaining the log of wages, as the smoothing parameter given in Table 6 is significant at the 1% level.

Table 6: Parameter estimates of YED and its Chebychev's polynomials for the different models discussed in Section 4 using the NSSO Data for India 2004-05.

	Standard Mincer	Polynomial Mincer	Semiparametric
YED	0.0917 (0.0589*)	0.0744 (0.007)	0.0922 (0.1387**)
(YED) ₂		-0.0448 (0.0092)	
(YED) ₃		0.0227 (0.0031)	
(YED) ₈		-0.6516** (0.109**)	
(YED) ₉		0.1442** (0.0254**)	
smoothing parameter (YED)			0.5192

Each j^{th} degree polynomial variable is multiplied by $10^{-(j-1)}$ to get standardized coefficient value. All estimates are significant at the 1% level.

* Multiplied by 10^{-2} .

** Multiplied by 10^{-4} .

The polynomial specification also captures the non-linear relationship between PEX and the log of wages, more than the standard Mincer specification. Comparing the coefficients of the polynomial and the semiparametric model shows that the semiparametric model retains the non-linearity captured by polynomial specification as the values of the coefficients do not vary much across these two models.

Figure 5 shows the estimates for the marginal rate of return to education across years of schooling in India for all the models. In the standard Mincer regression, the estimate for the marginal rate of return is a constant 9.2% per year of schooling. Conversely, the more generalized versions of the Mincer regression, both polynomial and semiparametric, show changes in the estimates of the marginal rate of returns to education across years of schooling. Estimates for the polynomial and semiparametric models show similar trends except at the tails. As mention by Green and Silverman (1994) the polynomial regression suffer from various drawbacks, on of them is overfitting, which can be seen in the graph where it tries to overfitt at the tails.

The points on the graph can be interpreted as the marginal return to ed-

Table 7: Parameter estimates for PEX and its Chebychev's polynomials for the different models discussed in Section 4 using the NSSO Data for India 2004–05.

	Standard Mincer	Polynomial Mincer	Semiparametric
PEX	0.047 (0.0871*)	0.2023 (0.0121)	0.2014 (0.0288*)
(PEX) ² (Squared term)	-0.007 (0.0194*)		
(PEX) ₂		-0.1438 (0.0138)	-0.1432 (0.0328*)
(PEX) ₃		0.06016 (0.0075)	0.05997 (0.0179*)
(PEX) ₄		-0.01414 (0.0021)	-0.0141 (0.5095**)
(PEX) ₅		0.186* (0.032*)	0.186* (0.0772**)
(PEX) ₆		-0.0128* (0.248**)	-0.0127* (0.0059**)
(PEX) ₇		0.0353** (0.0075**)	0.0352** (0.0002**)

Each j^{th} degree polynomial variable is multiplied by $10^{-(j-1)}$ to get standardized coefficient value. All estimates are significant at the 1% level.

* Multiplied by 10^{-2} .

** Multiplied by 10^{-4} .

ucation of attending school for an additional year. Since the marginal return is the derivative at a particular year of schooling, it reflects the slope at that point of the total return curve. For instance, an individual with seven years of schooling will get a return of around 8% by attending school for an extra year. The same returns increase to 12% per year after this individual complete ten years of schooling.

The marginal rate of return to education is around 6% when an individual first enroll in school. This declines to 5% in next couple of years before going back to 6% after finishing Primary education. The return increases to 8% after finishing middle school, 12% after finishing secondary school and 14% after high school (or higher secondary). After 12 years of education, the return varies around 14% – 15% during college and higher studies.

5.2 Analysis

With the background of the theory in Section 2, the graph reflects equilibrium points of demand and supply curves of education for the society. Since ability is largely determined by nature, the equilibrium points provide information about the supply of education in the society. The supply includes the marginal cost

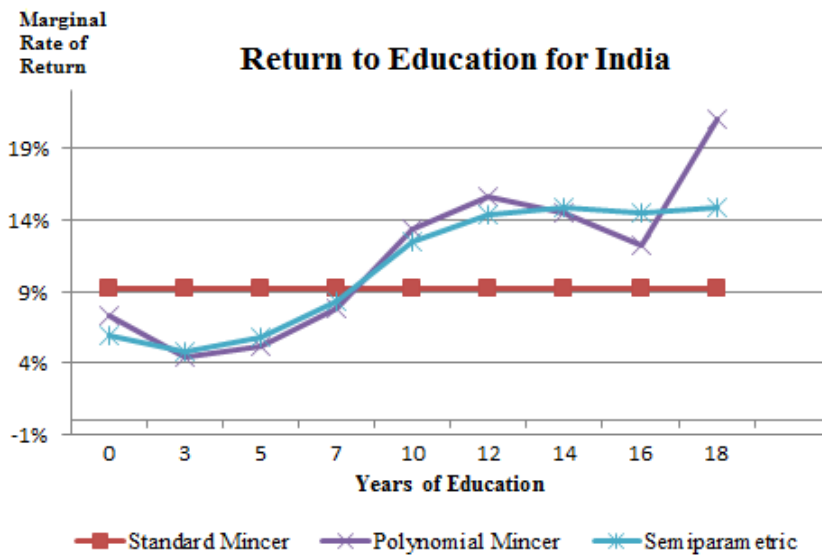


Figure 5: Marginal returns to education by years of education for India based on the NSSO Employment and Unemployment Survey 2004-05.

curves faced by individuals in the society along with the government policies at each schooling level. These curves can help to understand the effect of government policies and thus relate the actions with the outcomes of each policy. It can also suggest a policy change, if any required.

The entire curve can be divided into three parts based on the observed correlation: before primary education, primary education through high school, and after high school. Each part show a different correlation between return to education and the years of schooling. This result is at the core of this paper and further findings and policy implications are based on this result.

The first part reflects a declining trend (a negative correlation case) in the returns to education and can be explained either by declining marginal cost curves faced by society at an aggregate level or heterogeneity in cost curves faced by the individuals in the society. The individuals in this group are mostly children, ages of 6 to 10, who cannot participate in the labor market. Even the lack of outside options in the labor market does not add to the possibility that this group collectively faces a declining marginal cost curve. Hence, a downward trend in the returns to education can only be explained by the heterogeneity in the cost of funds faced by individuals at this level as shown in Figure 1. Since the cost of education for a 6 to 12 year old child is borne by the family, this reflect the constraint faced by families in providing education to child. This explains why the dropout rate from primary schools has been as high as 50% in past

years, and is still above 25% during 2005-06 (Department of Higher Education, 2008). It also reflects the lack of enforceability of free and compulsory education as provided by the constitution of India.¹⁶

An upward trend (the positive correlation case) in returns to education for the second part reflect a much more homogenous upward sloping marginal cost curve faced by the society as the upward trend can only be explain by an increasing marginal cost curve (see Figure 3). The reason for seeing less heterogeneity in the marginal cost of individuals in this group is the self selection of individuals at earlier stages of schooling. Families who cannot afford to send a child to school have already dropped out within the first five years. The upward trend also reflects the heterogeneity in abilities and shows that individuals with higher abilities choose higher levels of schooling.

A nearly horizontal trend (the no correlation case) in education after high school reflected in Figures 2 and 4. It may be the case that all individuals are facing the same constant marginal cost after high school and heterogeneity in ability is showing this trend (as in Figure 2). However, this reasoning can be refuted as we see that there is some fluctuation in the estimates in the semiparametric and polynomial models. Even with overfitting, this seems to reflect the case seen in Figure 4, which is high heterogeneity in both ability type and marginal cost curves for the society.

Government policy on education is crucial in any country and this certainly holds for India. During 2005–06, the government of India spent 1.61% of GDP on elementary education, 0.89% on secondary education, 0.28% on technical education and 0.67% on university and higher education (Department of Higher Education, 2008). This reflects the importance place of primary education in government policy as more than half of educational expenditures go to elementary education (Primary and Middle School). The Indian government has since enacted laws to provide free and compulsory education to all children between 6 to 14 years of age. The government's need for these policies confirms my analysis as well as the need for government intervention in providing education for all children during the first five years of school. However, there is still a need to provide large scale merit based scholarships and financial assistance to students to ensure low drop-out rates at higher schooling levels.

¹⁶The failure of free education policy is also argued by Tilak (1996).

6 Tests of Empirical Methodology

As a test for the methodology, the same estimation techniques used to study India are also applied to data for the USA. I further test these techniques by checking the estimation results with annual wages instead of hourly wages.

6.1 The USA

The selection procedure for choosing the appropriate degree of polynomials for YED and PEX gives $j = k = 9$. Using SBC in stepwise method for model selection, results in a smaller polynomial model. Table 8 shows parameter estimates and some key measures for each model with the USA data. The polynomial model merely adds 0.01 value in adjusted R-square value over the standard Mincer regression, again reflecting the relevance of the Mincer regression in modeling wages. The polynomial model reduce the model error by 1.34% over the standard Mincer specification, and the semi parametric specification further reduces the error by less than 0.05%. The PEX, race and location are important factors in the model and do not change much with the specification.

The linear estimate for return to education remains around 10% for both standard and semiparametric estimation. This is above the 9% estimate for India. The absence of linear estimates for YED and the presence of a second order polynomial for YED is because of the presence of higher order polynomial and the stringent model selection procedure¹⁷. This result do not suggest the irrelevance of schooling as linear form in the Mincer model and should be seen as just an outcome of the default model selection procedure. Even in the selected model the estimate of marginal returns to education in polynomial model and semi parametric model does not change much which indicate the robustness of my estimates.

Figure 6 shows the results for the USA. The standard Mincer regression gives a constant estimate for each year of schooling compared to the variant estimates from my other models. Once again there is overfitting at the tails of the data in the polynomial model. Overall, the curves show the upward trend in returns to education for the first 12 years of schooling, reflecting a homogenous increasing marginal cost curve in the society. This is quite possible as education is free and compulsory in USA and attendance is enforced by law. The proportion of students who drop out before completing high school is less than 10%, which reflect that these students really struggle in school and possibly are of low ability type. The lack of a trend after high school is also seen in the estimates for India and likely reflects the presence of heterogeneity in abilities and varying marginal cost curves in the society. This appears feasible since

¹⁷The exclusion of some degree of polynomials on YED add the linear YED back into the model.

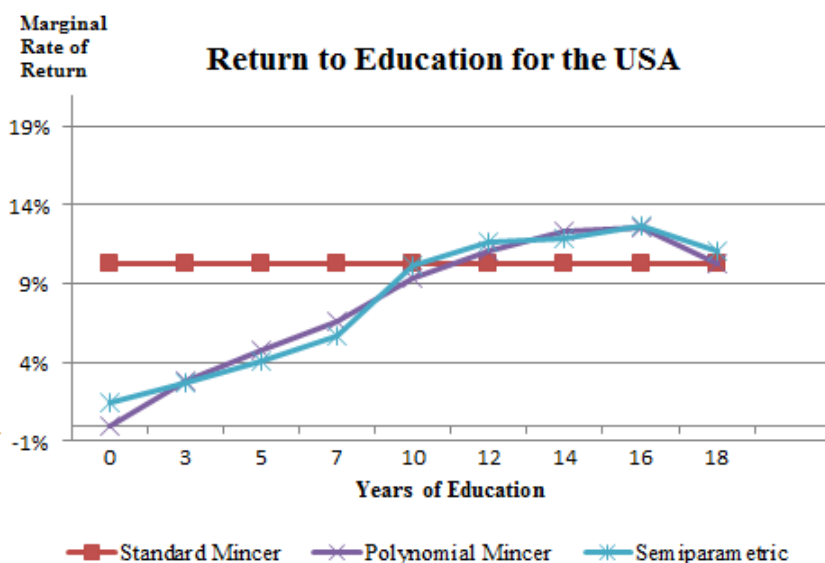


Figure 6: Marginal returns to education by year for the USA (based on the 2005 March Supplement of the CPS survey)

the cost of college is high in the USA and not everyone can afford to attend. The high cost of college education increases the marginal cost of additional funds required for graduate and professional degrees. High heterogeneity in the marginal cost curves is reflected by a slight negative trend around 18 years of schooling (additional years of education are recoded as 18 to match the data for India). To analyze this further, extended estimates are created for the USA and the results are presented in Figure 7.¹⁸ The polynomial model once again shows how bad it estimate at the tails. The presence of a continuous negative trend which starts after 14 years of education reflect the presence of a credit constraint similar to the case of primary education in India. This identifies the phase of schooling that can be improved by policy changes. One policy improvement would be to provide an interest waiver and deferment of repayment on education loans for students who pursue further studies after college. This and other policy changes (e.g. financial assistance and scholarships) can lower the marginal cost for individuals who face credit constraints. Innovations that make higher education more affordable will likely have a similar impact. Such policies would attract high ability individuals to pursue higher studies.¹⁹

¹⁸Individuals listed as “12th Grade with no diploma” are excluded from the analysis as some papers suggest measurement error in this category (Singer and Ennis (2003), Scanniello (2007), Heckman and LaFontaine (2010)). The “Professional school degree” and “Doctorate degree” categorize are combined into schooling that requires 21 years of education. Individuals with “some college but no degree” are assigned 13 years of schooling. The “7th or 8th Grade”, “9th Grade”, and “11th Grade” are assign 7.5, 9, and 11 years of schooling, respectively.

¹⁹Dropping the category “12th Grade with no diploma” affects the estimates of the standard

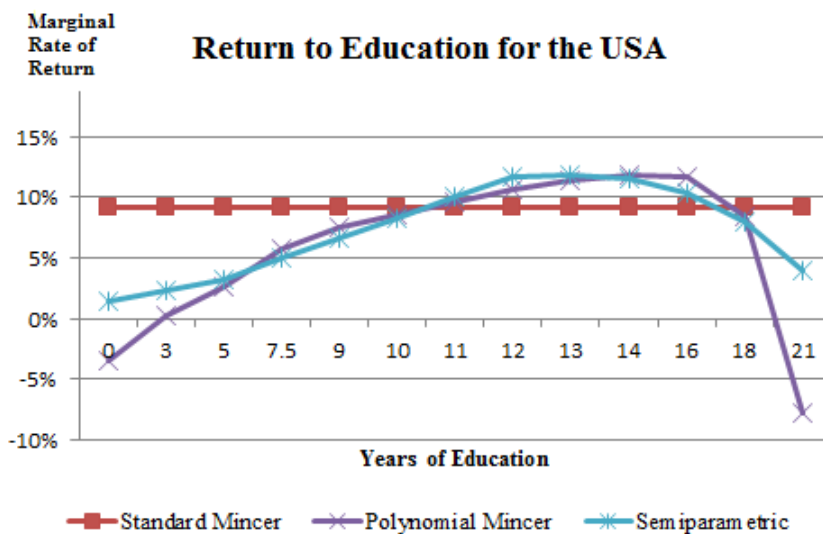


Figure 7: Semiparametric model estimates of the returns to education by year of schooling for the USA with extended map f years of schooling

In all, my methodology does reflect the social and economic scenario of the society and gives insights into a society’s marginal cost curves. As a further test of these empirical tools, I examine the use of hourly wages in comparison to annual wages in the semiparametric specification.

6.2 Annual vs Hourly Wages

In many studies, hourly wages are used to estimate the returns to education in order to avoid the effect of market fluctuations. Although annual wages can be used to derive hourly wages, as is done in this paper for the USA, hourly wages conceal the contribution of education in providing stable and secure jobs in comparison to temporary and part time jobs that require less educated workers. The use of annual wages should increase the marginal return curve, making wage returns more correlated to education. To be more precise, the estimates for marginal returns to education based on annual wages should be higher for higher schooling levels than the estimates based on hourly wages. For the USA, annual wages data is available. For India, however, this information is not reliable as data on weekly wages are provided in the survey and not enough variation is found in the data for days worked in a week.²⁰ Since our estimates are based on hourly data (derived from weekly data), they might be understating

Mincer regression by a reduction of 1%.

²⁰However, we have variables that indicate stable and secure job like ‘Availability for additional work’, ‘full time or part time job’ for our data on India. The mean of these variables by years of schooling indicate more educated person are less likely to be available for additional work and more likely to be in full time job.

the results. To approximate the bias and as a test of the methodology, I estimate the returns to education for annual wages in the USA, and compare it with the estimates based on hourly wage data.

Figure 8 shows the marginal returns for the semiparametric model with the different dependent variables.

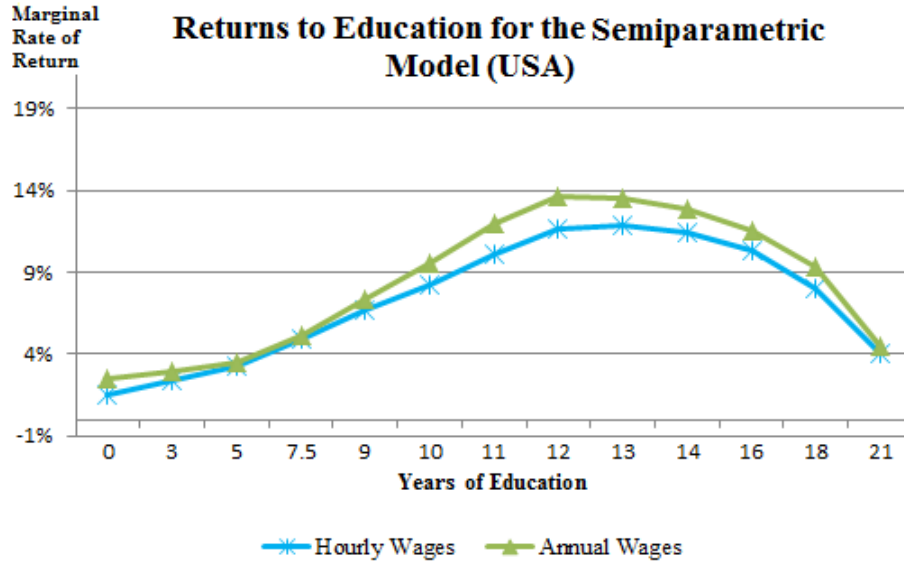


Figure 8: Semiparametric model estimates of the returns to education by year for the USA with hourly and annual wages as the dependent variable

The graph shows that, as expected, the returns are higher for later years of education in annual wages compared to hourly wages. On average, the estimates based on annual wages are 14% higher than the estimates based on hourly wages. This reflects our estimates for India are understated to some extent.

7 Conclusion

The semiparametric and polynomial specifications are used with a Mincer framework to estimate the returns to education for each schooling level in India. Compared to the standard Mincer regression, my estimates identify the trends in the marginal rate of return to education more accurately. These generalized specification provide much smoother estimates at each given year of schooling compared to the standard dummy variable approach based on the Mincer regression. My estimates reconfirm that the marginal rate of return on education is non-linear with respect to years of schooling.

Similar to other studies on India, my results show that the returns to education increase at a higher rate after the first seven years of schooling. Moreover, I discover the schooling level where the impact of credit constraints on receiving education is the highest. The negative trend in the estimates for the marginal rate of return to education during the first five years of school signals the presence of a binding credit constraint. It appears that the individuals with credit constraints drop-out during these first five years and the remaining individuals face no credit constraints until college (which is shown by the positive trend in the estimates from grades five to twelve). There is no clear trend after high school.

I cross-check my methodology for estimating the returns to education with data from the USA. The effect of compulsory and free education through high school in the USA is evident from the positive trend between the marginal rate of return to education and years of schooling. The non-positive trend between years of schooling and the marginal rate of return to education at the top end of the education spectrum in the USA reflects the presence of credit constraints after high school. These credit constraints are exaggerated for graduate and professional education. I find that the use of annual wages instead of hourly wages reflects the added benefit of education in providing stable and secure employment and may increase the estimates of returns to education by up to 14%.

The methodology suggested in this paper can be used for analyzing the hidden marginal cost curves for obtaining education in other sub-populations. Policymakers can draw on this information and build policies according to the needs of the different groups in their society. I suggest the use of targeted financial aid as well as encouraging innovations to decrease the cost of education.

References

- ABRAHAM, V., “Wages and earnings of marginalized social and religious groups in India: Data sources, scope, limitations and suggestions,” (2012).
- AGARWAL, T., “Returns to Education in India: Some Recent Evidence,” Indira Gandhi Institute of Development Research, Mumbai, WP-2011-017, 2011.
- ASHENFELTER, O., C. HARMON AND H. OOSTERBEEK, “A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias,” NBER Working Paper 7457, January, 2000, ????
- BECKER, G. S., “Human Capital: A Theoretical and Empirical Analysis,” National Bureau of Economic Research, 1964.
- , *Human Capital and the Personal Distribution of Income: An Analytical Approach* (Woytinsky Lecture no.1. Ann Arbor: University of Michigan, Institute of Public Administration, 1967).
- , “Investment in Human Capital: Rates of Return,” in *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education* (NBER, 1975).
- BELLEY, P. AND L. LOCHNER, “The changing role of family income and ability in determining educational achievement,” *Journal of Human Capital* 1 (2007), 37–89.
- BHAUMIKA, S. K. AND M. CHAKRABARTY, “Is education the panacea for economic deprivation of Muslims?: Evidence from wage earners in India, 19872005,” *Journal of Asian Economics* 20 (2009), 137–149.
- BROWN, M., J. SCHOLZ AND A. SESHADRI, “A New test of Borrowing Constraints for Education,” University of Wisconsin, Working paper, 2011.
- CAMERON, S. AND J. J. HECKMAN, “Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males,” *Journal of Political Economy* 106 (1998), 262–333.
- , “Can tuition policy combat rising wage inequality?,” in *Financing College Tuition: Government Policies and Educational Priorities* (American Enterprise Institute Press, Washington, 1999), 76–124.
- CAMERON, S. AND C. TABER, “Estimation of educational borrowing constraints using returns to schooling,” *Journal of Political Economy* 112 (2004), 132–82.

- CARD, D. AND A. B. KRUEGER, "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States," *The Journal of Political Economy* 100 (Feb 1992), 1–40.
- CARNEIRO, P. AND J. J. HECKMAN, "The evidence on credit constraints in post-secondary schooling," *Economic Journal* 112 (2002), 705–732.
- CAWLEY, J., J. HECKMAN AND E. VYTLACIL, "Cognitive Ability and the Rising Return to Education," NBER Working Paper 6388, 1998.
- DELGADO, M. A. AND P. M. ROBINSON, "Nonparametric and Semiparametric Methods for Economic Research," *Journal of Economic Surveys* 6 (September 1992), 201–249.
- DEPARTMENT OF HIGHER EDUCATION, G. O. I., MHRD, "National Level Educational Statistics - At a Glance," Technical Report, Ministry of Human Resource Development, Government of India, 2008.
- DESCHÊNES, O., "Unobserved Ability, Comparative Advantage and the Rising Return to Education in the United States: A Cohort-Based Approach," Princeton University Industrial Relations Section Working Paper No. 465, August, 2001.
- DURASAMY, P., "Changes in the returns to education in India, 1983-94: by gender, age-cohort and location," *Economics of Education Review* 21 (2002), 609–622.
- DUTTA, P., "Returns to Education: New Evidence for India, 1983-1999," *Education Economics* 14 (2006), 431–451.
- GABRIEL, S. A. AND S. S. ROSENTHAL, "Location and the effect of demographic traits on earnings," *Regional Science and Urban Economics* 29 (1999), 445–461.
- GELBACH, J. B., "When Do Covariates Matter? And Which Ones, and How Much?," mimeo, Eller College of Management, University of Arizona, 2009.
- GORODNICHENKO, A. AND K. S. PETER, "Returns to schooling in Russia and Ukraine: A semiparametric approach to cross-country comparative analysis," *Journal of Comparative Economics* 33 (2005), 324–350.
- GREEN, P. J. AND B. W. SILVERMAN, *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*, first edition (Chapman and Hall, 1994).
- HÄRDLE, W., *Applied Nonparametric Regression* (Cambridge University Press, 1990).

- HASTIE, T. AND R. TIBSHIRANI, “Generalized Additive Models,” *Statistical Science* 1 (1986), 297–318.
- HECKMAN, J. J. AND P. A. LAFONTAINE, “The American High School Graduation Rate: Trends and Levels,” *The Review of Economics and Statistics* 92 (May 2010), 244–262.
- HECKMAN, J. J., L. J. LOCHNER AND P. TODD, “Fifty Years of Mincer Earnings Regressions,” NBER Working Paper 9732, May 2003a.
- HECKMAN, J. J., L. J. LOCHNER AND P. E. TODD, “Fifty Years of Mincer Earnings Regressions,” NBER Working Paper No. 9732, May 2003b.
- HOROWITZ, J. L., *Semiparametric Methods in Econometrics* (Springer, 1998).
- JAEGER, D. A. AND E. A. PAGÉ, “Degree Matter: New Evidence on Sheepskin Effects in the Returns to Education,” *The Review of Economics and Statistics* 78 (1996), 733–740.
- JOHNSON, M., “Borrowing constraints, college enrolment, and delayed entry,” Mathematica Policy Research, Working Paper, 2010.
- KEANE, M. AND K. WOLPIN, “The effect of parental transfers and borrowing constraints on educational attainment,” *International Economic Review* 42 (2001), 1051–1103.
- KINGDON, G., “Does the Labour Market Explain Lower Female Schooling in India?,” *Journal of Development Studies* 35 (1998), 39–65.
- KRUEGER, A. B. AND M. LINDAHL, “Education for Growth: Why and For Whom?,” *Journal of Economic Literature* 39 (2001), 1101–1136.
- LEMIEUX, T., “The Mincer Equation Thirty Years after Schooling, Experience, and Earnings,” in *Jacob Mincer A Pioneer of Modern Labor Economics* (Springer Science+Business Media Inc., 2006).
- LOCHNER, L. AND A. MONGE-NARANJO, “Credit Constraints in Education,” NBER Working Paper No. 17435, 2011.
- MADHESWARAN, S. AND P. ATTEWELL, “Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey,” *Economic and Political Weekly* 42 (2009), 4146–4153.
- MINCER, J., “Investment in Human Capital and Personal Income Distribution,” *Journal of Political Economy* 66 (1958), 281–302.
- , “Schooling, Experience, and Earnings,” New York, NBER Press, 1974.

- , “Changes in Wage Inequality, 1970–1990,” *Research in Labor Economics* 16 (1997), 1–18.
- MURPHY, K. M. AND F. WELCH, “Empirical Age-Earnings Profiles,” *Journal of Labor Economics* 8 (1990), 202–229.
- PSACHAROPOULOS, G. AND H. A. PATRINOS, “Returns to Investment in Education: A Further Update,” *Education Economics* 12 (2004), 111–134.
- RAMSAY, J. O. AND B. W. SILVERMAN, *Functional Data Analysis*, second edition (Springer, 2005).
- SCANNIELLO, N., “Comparison of ACS and ASEC Data on Educational Attainment: 2004,” Technical Report, U.S. Census Bureau, Housing and Household Economics Statistics Division; Washington, DC, 2007.
- SINGER, P. AND R. S. ENNIS, “Census 2000 Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview,” Technical Report, United States Census Bureau; Washington, D.C., 2003.
- STINEBRICKNER, T. AND R. STINEBRICKNER, “The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study,” *American Economic Review* 98 (2008), 2163–84.
- TILAK, J. B. G., “How Free is ‘Free’ Primary Education in India?,” *Economic and Political Weekly* 31 (1996), 355–366.
- TOBIAS, J. L., “Are Returns to Schooling Concentrated Among the Most Able? A Semiparametric Analysis of the Ability-earnings Relationships,” *Oxford Bulletin of Economics and Statistics* 65 (2003), 1–29.

Appendix

A Deriving The Return to Education

Calculation the returns to education is often based on finding the effect of a change in years of education on wages. In the standard Mincerian regression model total returns on education is represented by

$$\text{Total Return to Education} = \beta_1 YED$$

The marginal return is derived by taking the derivative of $\log(\text{wages})$ with respect to YED . In standard Mincer regression (Equation 1), it is simply represented by the value of the coefficient for years of education:

$$\frac{d(\log(\text{wages}))}{d(YED)} \Big|_{YED} = \beta_1$$

This is a single estimate for all education levels and is interpreted as the returns for an additional year of schooling. Since the focus of this paper is to examine the changes in returns to education at different schooling levels, Chebychev's polynomial (Equation 2) and semiparametric (Equation 3) Models are more interesting. Chebychev's polynomial model includes several degrees of polynomials, so the total returns to education is given as follows:

$$\text{Total Return to Education} = \beta_{11} YED + \beta_{12} YED_2 + \dots + \beta_{1j} YED_j$$

The marginal rate of return to education at each schooling level can be calculated by using the following equation

$$\frac{d(\log(\text{wages}))}{d(YED)} \Big|_{YED} = \beta_{11} + \beta_{12} \frac{d(YED_2)}{d(YED)} + \dots + \beta_{1j} \frac{d(YED_j)}{d(YED)}$$

Where the derivatives for Chebychev's orthogonal polynomial of first kind

are

$$\begin{aligned}
\frac{d(YED_2)}{d(YED)} &= 4YED \\
\frac{d(YED_3)}{d(YED)} &= 12YED^2 - 3 \\
\frac{d(YED_4)}{d(YED)} &= 32YED^3 - 16YED \\
&\dots = \dots \\
\frac{d(YED_{n+1})}{d(YED)} &= 2 \left\{ YED_n + YED \frac{d(YED_n)}{d(YED)} \right\} - \frac{d(YED_{n-1})}{d(YED)}
\end{aligned}$$

Based on the estimated structure for the non-parametric function of $\beta_1()$, the total returns to education for the semiparametric specification is given by

$$\text{Total Return to education} = \beta_1(YED) = \beta_{11}YED + s_{12}(YED)$$

Proc GAM output in SAS provides a linear estimate for YED, β_{11} , as well as a non-parametric estimate of YED, the value of s_{12} for each schooling level. To derive the marginal returns to education from this information I use MATLAB with the the estimated total return to education across schooling years. First, I use the “spline” function to get the piecewise polynomial form of the cubic spline that fits total returns to education for each schooling level. Then a first derivative function of the piecewise polynomial function is found by using the “fnder” function. The value of this first derivative, an estimate of the marginal returns to education at each schooling level can then be found by using the “ppval” function. This method ensures that I get an estimate for each schooling level.

B Estimates from Other Studies on India

Table 9: Estimates of marginal rate of return to education for males in some of the previous studies for India

Schooling Level	This Paper	Agarwal(2011)*	Dutta (2006)**
Non-Literate	5.88		
Below Primary	4.78		
Primary	5.84	5.47	5.6
Middle	8.38	6.15	3.5
Secondary	12.47	11.38	6.1
Higher Secondary	14.36	12.21	
Diploma/Certificate Course	14.85		
College Graduate	14.49	15.87	12.3
Postgraduate and Above	14.8		

* Male and female both included. Based on India Human Development Survey (IHDS) 2005

** For year 1999-2000. Male and regular workers only.

Table 8: Parameter estimates and key goodness-of-fit indicators for the different models discussed in Section 3 using the CPS March 2005 Data Supplement for the USA.

Variables	Standard Mincer	Polynomial Mincer	Semi Parametric
Intercept	0.9975 (0.0184)	1.3595 (0.0209)	0.8210 (0.0574*)
YED	0.1022 (0.0011)		0.1042 (0.2953**)
(YED) ₂		0.0237 (0.0461*)	
(YED) ₉		-0.0027** (0.0459*)	
smoothing parameter (YED)			0.7912
PEX	0.047 (0.0011)	0.0831 (0.0028)	0.0831 (0.7265**)
(PEX) ² (Square term)	-0.0075 (0.0254*)		
(PEX) ₂		-0.0137 (0.0707*)	-0.0137 (0.1841**)
(PEX) ₃		0.0749* (0.5186**)	0.0749* (0.0135**)
UR=Black	-0.2156 (0.011)	-0.2004 (0.011)	-0.2003 (0.0286*)
RA=Non- Metropolitan	-0.1737 (0.0092)	-0.1571 (0.0092)	-0.1572 (0.0239*)
Adj R-Square	0.2442	0.2543	
F-Value	2522.6	1901.8	
Model (Error)	25,659,704	25,316,253	25,306,755
DF (Error)	39,013	39,011	

Each j^{th} degree polynomial variable is multiplied by $10^{-(j-1)}$ to get standardized coefficient value. All estimates are significant at the 1% level.

* Multiplied by 10^{-2} .

** Multiplied by 10^{-4}

** Multiplied by 10^{-4} .