Education and health outcomes for social minorities in India: An analysis using SUR model

Ganita Bhupal[1]

and

Abdoul G. Sam[2]

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington, August 12-14, 2012.

[1] Graduate Student, Agricultural, Environmental and Development Economics, The Ohio State University. (email: bhupal.1@osu,edu).

[2] Associate Professor, Department of , Agricultural, Environmental and Development Economics, The Ohio State University.(email:sam.7@osu.edu).

Education and health outcomes for social minorities in India: An analysis using SUR model

Ganita Bhupal and Abdoul G. Sam

Abstract

The current study analyzes the health and education outcomes of fifteen year old children in India and investigates the question of inequality of such outcomes for socio-religious categories. To study the effect of health on education, SUR estimation has been undertaken. The comparison of SUR and OLS results shows that SUR estimates have smaller standard errors than the OLS estimates. Of the three categories analyzed in the data, STs have worst outcomes for both education and health and SCs lag behind in the health field. The results have important implications for policy regarding education and health of the socio-religious minorities.

I. Introduction

The questions of caste, tribe, religion and region have been pertinent to the development debate in India. The significant caste-class overlap (Deshpande, 2001) further alleviates the social and economic disparity associated with the caste system in India. Not only is there stark inequality in economic outcomes, the deprivation of religious and social minorities extends to education and health as well. In such a context, the analysis of health and education outcomes is of crucial importance to the discussion of caste outcomes in India.

The literature on education and health inequality in India is scant and only a handful of studies exist. Many of the studies on education and caste analyse the situation from the

perspective of affirmative action policy in India and present mixed evidence. Desai and Kulkarni (2008), using longitudinal data from four rounds of the National Sample Survey find that the primary school enrolment and completion rates have shown convergence for almost all socio religious categories in India over a period of almost 20 years, but the middle school and secondary school rate convergence, is not significant. They also identify muslims as the only class which has not shown much improvement. Dreze and Kingdon (2003) use the primary data collected from the PROBE study of 1999 look at the predictors of school participation in rural India and note that SC-ST children have an "intrinsic disadvantage" meaning that they have a relatively low chance of going to school even after controlling for household wealth, parental education and motivation, school quality, and related variables[3]. Dreze and Kingdon use cross-sectional data and resort to using ordered logit in order to get around the problem of non-random assignment of children to school. The study by Basant and Sen(2010) utilizes data from the 61st round of NSSO and they find that the relative deprivation of SCs and STs is not significant across all categories of individuals and all categories of outcomes. Borooah and Iyer (2005) looks at the primary enrolment ratios in India using the India Human Development Survey, and they find significant differences in the school enrolment of socio religious minorities and other caste hindus. They also establish that the community effects are more pronounced when the attribute effects are smaller. Their main findings suggest that the size of the religion or caste effect depends on the non community circumstances in which the children are placed. Under favorable circumstances (for example, when parents are literate), the size of the community effect is negligible. Under less favorable circumstances, the size of the community effect is considerable.

[3] Certain lower castes and tribes in India were placed under special schedules of constitution to provide them with positive affirmative action. Hence the classification- Scheduled caste(SC) and Scheduled Tribe (ST).

The studies which pertain to health and caste in India reach the same conclusions about the relative deprivation of SC and ST. There are very few studies which explore this issue in details. The studies also range across various demographic profiles. Mahindra et al.(2006) look at the self reported health status and limitations in daily activities of women in rural Kerala and find that lower caste women bear a "double deficit" which points to the social inequality persistent in India relating to caste and gender. A study on malnutrition and socio-economic groups in India was carried out by Poel and Speyboreck (2009) and the study concludes that the scheduled tribe children have lower health outcomes

(measured by the height for age z scores) than the other caste groups. They also conclude that the gap was primarily caused by SC/ST's lower wealth, education and use of health care services. A more recent study conducted by Sabharwal (2011) looks at incidence of child malnutrition and concludes that SC, ST and muslim children have significantly worse health outcomes than the children from other socio-religious backgrounds.

With such limited evidence on the education and health outcomes in the context of socio-religious diversity in India, the present study finds its relevance in analyzing both these outcomes for children belonging to different groups. Most of the aforementioned studies identify certain groups for specific focus. The groups being: Scheduled castes (SCs), Scheduled tribes (STs), Other backward classes (OBCs), muslims, hindus etc. For the purpose of this discussion, we restrict our focus to three categories: SC, ST and muslim.

The current study looks at the health and education outcomes of 15 year old children in Andhra Pradesh, India. To allow for the effects of health on education outcomes, we use a

system of equations and to account for the cross equation co-variance between the error terms, we use an SUR model to estimate the outcomes.

This paper's contribution to the literature is many fold. We analyze the education and health outcomes with respect to caste. We estimate the model using an SUR model which takes into account the non- zero covariance between error terms in simultaneous equation setting. The model also helps to analyze the effect of health on education outcomes, something which has never been attempted before in the health and education literature in India. To take into account the sample selection bias due to the school drop- outs from one round of survey to the other, we use propensity score matching.

The results from the analysis of health and education outcomes suggest that social status of the children remains an important factor in the determination of such outcomes. The education and health outcomes are significantly lower for STs, whereas the SCs face relative deprivation only for the health outcomes. Equally important is to note that health plays a significant role in the determination of education outcomes. The economic status of households and children's involvement in paid work to support the income of household remain two most important reasons for dropping out of the school. The SUR estimates generate more efficient and, in most of the cases, larger estimates.

The results from this exercise throw light on some serious issues pertaining to social minorities in India. The results also throw interesting challenges to the policy for provision of universal education and health to all children across all categories.

The rest of the paper is divided as follows: Section 2 discusses the data and related issues; Section 3 describes the methodology used; Section 4 presents the results; and Section 6 concludes.

II. Data

The data for this study has been taken from the Young Lives Study Project which aims at following the lives of around 12,000 children in four developing countries- Peru, Vietnam, Ethiopia and India. The data collection started in the first round in 2001-02 and three rounds have been conducted so far (2006 and 2009 were the other two).The data provides a range of information on the two cohorts of children that the study targets. In India, the survey was conducted in Andhra Pradesh, a semi-arid state in the South of India. The state can be divided into three agro climatic zones: Rayalseema, Coastal Andhra and Telangana. The 23 districts comprising the state were ranked poor-non poor according to economic, human and infrastructure development indicators and purposive sampling was carried out to choose 7 poor and 5 non-poor districts. 20 sentinel sites were chosen from these districts again on the basis of poor and non-poor criterion. The selection of villages from the sentinel sites was random and a total of 20 villages were chosen from these 20 sentinel sites such that 50 children for older cohort were selected from each village to make up the total sample of 1000 eight year old children in the first round.

The survey has been thoroughly conducted and has very low attrition rates, thus eliminating the possibility of a sample selection bias. This study makes use of the third round of data to analyze health and education outcomes for fifteen year old children. The variable used to measure education outcome is the school grade that the child was studying in when the survey

was conducted in 2009. The variable used to measure health is the height for age z score of the child. The height for age z score is the normalized measure for height and is representative of malnutrition because it is a long term measure of health (Poel and Speybroeck, 2009). A detailed description of the variables that have been used is provided below and main statistics corresponding to those have been presented in Table 1.

Discussion of variables:

The two dependent variables that have been used to measure education and health outcomes for the estimation purposes is: Grade that the child was studying in both the rounds of data and the height for age z score.

The height for age z score is a measure of malnutrition and height is an indicator of long term health of the individual. The literature on health and education states that healthier individuals have better educational outcomes because they are better able to concentrate and have better attendance in schools (Behrman, 1996). Since health is a cumulative characteristic, the effect of poor health accumulates over time and may affect both short-term educational performance (e.g. grades) and long –term schooling outcomes (e.g. total number of years of schooling). Adequate evidence of effects of health on both long term and short term education outcomes is available in the literature. Galler et al . (1990) and Richardson et al. (1972) find that previously malnourished children often have poorer grades in school. Case et al. (2005) find that children with poor health have significantly lower educational attainment, poorer health and lower social status as adults. In some of their recent work, Case and Paxson (2008) prove that the returns to height are returns to cognitive ability of the individuals. Further refinement of their results (Case and Paxson, 2010) suggests that height is a proxy for in –utero and early

childhood health, which in turn is positively associated with cognitive ability. In the context of developing countries, a number of studies suggest that health of children is important in determining the age at which the children get enrolled in schoolg the school.

Independent variables:

Socio-religious categories: The three socio religious categories that have been used are: the Scheduled castes (SC), the scheduled tribes (ST), muslims . The muslims constitute a very small percentage of the sample and have been included due to the influence of religious affiliations on education. Indeed, a lot of studies have pointed out that the emphasis on traditional schooling (madrassas) lowers the probability of muslim children getting conventional education. A separate dummy for females was included to capture any distinction between males and females. Unlike Dreze and Kingdon (2003), we could not run separate regressions for males and females because of small number of observations and lack of large variability in the dependent variable for education.

Household variables: The most important household variables include the education of mother and father of the child and the wealth index of the household. The birth order of the child constitutes the fourth such variable.The wealth index is a composite index which is composed of a simple average of three other indices that have been calculated in the data and are directly available for use. These are: Household quality index which looks at the physical condition of the house: number of rooms, quality of roof etc.; the service index- which looks at the amount of services and the consumer durables that the household possesses. The education level of the mother has been coded as a dummy because of the startling fact that approximately 60% of the mothers in the data are illiterate. On the other hand, 42.35% of the fathers in the data

were illiterate. To ensure that the years of education are captured in the regression, the education of the father beyond high school was coded as 13 years of schooling if they answered to the question as post secondary or vocational training and 15 years if they answered university. (Typically in India, a bachelor's or undergraduate degree is for three years). The adult literacy for the fathers was coded as 1 year of education as there is no reliable information on the adult literacy in the data.

Individual variables: A measure of cognitive ability of the variable has been provided in all the rounds of the data. To account for such skills, this variable has been included in the estimation. Since cognitive ability can be endogenous to schooling, the data for this variable comes from the second round of the survey.

Equation specific variables:

School variables: (denoted as s_i in equation (1))Since the performance of the child can be assumed to depend on a variety of school related factors, some of the factors which affect the quality of school and hence education outcomes for children were included in the estimation. These are: if the child missed school for more than a week in the last school year, how many days a teacher was absent from the school in a typical week.

Genetic factors: (denoted as g_i in equation (2)).For the health equation, a few more variables were taken into account, including the height of child's mother and short term health.

Table 1: Description of Variables and Means

Variable	Description	Mean : Age 15

Enrol	1if the child is enrolled in school; 0 otherwise	0.772
Z score	Height for age Z score.	-1.662
Grade	School grade that the child is currently enrolled in.	9.662
SC	Equals 1 if the child belongs to a SC category; 0 otherwise	0.209
ST	Equals 1 if the child belongs to ST category; 0 otherwise	0.108
Muslim	Equals 1 if the child is a muslim; 0 otherwise	0.066
Female	Equals 1 if the child is a female; 0 otherwise	0.504
Wealth Index	Wealth Index of the household; takes values between 0 and 1; 1 indicates the highest level of wealth	0.334
Mother Educated	Equals 1 if the mother is educated; 0 otherwise	0.401
Father's Education	Number of years that the father received education	4.588
Order	Birth order of the child	1.545
Age School start	Age at which the child started school	5.041
Mother's Height	Mother's height (in centimeters)	150.737
Paidwork	Equals 1 if the child is involved in paid work; 0 otherwise.	0.278

III. Methodology

The two equations determining education and health outcomes have been estimated as follows:

$$e_i = \alpha + \beta_1 sc + \beta_2 st + \beta_3 muslim + \gamma_1\, x_{it} + \gamma_2\, s_i + \gamma_3\, h_i + \varepsilon_1 \qquad (1)$$

$$h_i = \alpha + \beta_1 sc + \beta_2 st + \beta_3 muslim + \gamma_1\, x_i + \gamma_2\, g_i + \varepsilon_2 \qquad (2)$$

The above presented equations (1) and (2) identify the main equations for estimation in the methodology. The dependent variables e and h represent the education and health outcomes respectively. The corresponding individual and time subscripts are denoted by i and t respectively. The variables sc, st and muslim represent the socio religious categories that we are interested in, namely Scheduled castes, scheduled tribes and muslim. The household variables are represented by the vector x, and are three important variables related to determination of health and education outcomes of children: wealth index, education of the father and education of the mother. The school related variables are expressed in the vector s.

The estimation methodology for this system of equations is a panel data SUR (Seemingly Unrelated Regression) model. Zellner's SUR or SUR model is undertaken in the estimation of a system of equations when the error terms across the equations can be assumed to be correlated. The system of equations for health and education outcomes can be viewed as a system of demand equations for these two human capital variables as has been often done in the literature on health and education. Also interesting to note is that the equations have same explanatory variables and that might increase the likelihood of having correlations in the error terms. In such a situation, the estimation of equations by classical OLS renders the estimates of betas and

gammas inefficient. Therefore, to recover efficiency, this system of equations is estimated with a FGLS (Feasible Generalized Least Squares) and this is known as the Zellner's SUR estimates.

With each round of data collection, the number of children attending school goes down as there are drop outs from the school. This creates selection bias in the data and to account for this sample selection bias, we use propensity score matching. The technique of propensity score matching was developed in the medical studies to account for the selection on observables.

Since the schooling equation shows drop outs from one round to the other, a probit was undertaken to identify the determinants of school enrolment and to correct for any bias arising due to non randomness of school enrolment selection, propensity score matching was used. This technique, even though, does not account for any unobservable determinants of school enrolment, presents a valid method to address the sample selection bias arising out of non randomness. Heckman's two step selection process which relies on finding an exclusion restriction for the correction of sample selection could not be undertaken because of the lack of a good instrument in the data.

IV. Results

The Probit results : the results from the probit to estimate the predictors of school enrolment at both the stages have been presented in the Appendix table 2[4]. The important results to note from the probit regressions are that the education of both the parents, the socioeconomic status of the households and involvement of child in paid work remain important predictors of the probability of school participation. One extra variable on migration was added to the regression for fifteen year olds and the coefficient carries a negative sign and is significant in the estimation, indicating

[4] The probit results were calculated with sentinel site fixed effects.

that migration of family/ child is negatively associated with the probability of school enrolment. This may happen because migration is correlated with (or is driven by) economic and other shocks to the households and hence, represents the effects of those shocks on the school participation. The cognitive ability test scores were included in the regressions but were not found to be significant in any of the regressions and hence, were removed from the estimation for final reporting.

The final SUR and OLS results have been reported in Tables 2 and 3 for education and health outcome equations respectively.

The results from both the empirical exercises of SUR and OLS exhibit that the health outcomes of the children are positively and significantly correlated with their education outcomes. The cumulative health outcomes are, therefore, important in determining the school outcomes. The results are consistent with the large literature on education and health (Case et al., 2005).Though a large amount of this literature pertains to developed countries, a few studies undertaken in the context of developing countries also exhibit the same results.

Education outcomes: The comparison of results from the SUR and OLS in this equation suggests that almost all the variables have their expected signs and the same level of significance in both the specifications. An important thing to note is that the SUR estimates have slightly smaller standard errors. The results conform to the theory that taking account of cross –equation correlation between the cross equation error terms leads to more efficient estimates. Also except for the health outcome in the equation, all the other coefficients have larger estimates, implying that OLS slightly overestimates the effect of health on education.

Of the main socio- economic categories of out regression, the SC, ST and muslim , only ST show a relative deprivation in terms of education outcomes. And the OLS results are slightly lower than the SUR estimates. Hence, neither muslims nor SCs have significantly lower outcomes for eduation, after controlling for a range of household and individual characteristics.

Table 2: SUR and OLS results for the Education Outcome

Variables	Grade (SUR)	Grade(OLS)
SC	-0.00256	-0.000374
	(0.0919)	(0.0942)
ST	-0.389***	-0.387***
	(0.138)	(0.142)
Female	0.100	0.101
	(0.0669)	(0.0686)
Muslim	-0.0693	-0.0684
	(0.186)	(0.190)
Wealth Index	0.418	0.415
	(0.264)	(0.270)
Father's Education	-0.0168*	-0.0169*
	(0.00861)	(0.00882)
Mother Educated	0.0699	0.0685
	(0.0850)	(0.0871)
Z score	0.178***	0.184***
	(0.0336)	(0.0344)
Cognitive Ability	0.00411***	0.00411***
	(0.000907)	(0.000929)
Order	0.0315	0.0315
	(0.0201)	(0.0206)
Age School Start	-0.212***	-0.212***

	(0.0515)	(0.0528)
Rural	-0.371	-0.371
	(0.361)	(0.370)
Constant	9.328***	9.338***
	(0.453)	(0.464)
Observations	685	685
R-squared	0.244	0.244

Note: (1) Standard errors in parentheses (2)*** p<0.01, ** p<0.05, * p<0.1(3) The estimations were carried out with sentinel site fixed effects.

Health Outcomes: The health equation also shows the expected signs for all the variables. The SUR estimates are slightly larger and have smaller standard errors here. But the comparison of the main socio-economic categories shows that SCs and STs both have significantly lower health outcomes for health. The wealth index shows a positive and significant coefficient in the SUR formulation on account of smaller standard errors. The coefficient for mother's height is positive and highly significant conforming to the genetic theory. Unlike in the case of education outcome, mother's education has a positive and significant effect on the health of the child. The females perform no worse or better than males in both the aspects of education and health.

Table 3: SUR and OLS results for the Health Outcome

Variables	(2) Z score(SUR)	(4) Z score(OLS)
SC	-0.377***	-0.363***
	(0.103)	(0.111)
ST	-0.373**	-0.340**
	(0.155)	(0.149)

Female	-0.0878	-0.0663
	(0.0755)	(0.0747)
Muslim	-0.151	-0.106
	(0.210)	(0.168)
Wealth Index	0.524*	0.453
	(0.293)	(0.277)
Father's Education	-0.000423	0.00219
	(0.00973)	(0.00968)
Mother Educated	0.238**	0.217**
	(0.0948)	(0.0982)
Order	-0.0116	-0.00478
	(0.0227)	(0.0199)
Rural	-0.0802	-0.114
	(0.406)	(0.420)
Mother's Height	0.0125***	0.0128***
	(0.00301)	(0.00418)
Constant	-3.633***	-3.647***
	(0.493)	(0.643)
Observations	685	713
R-squared	0.144	0.140

Note: (1) Standard errors in parentheses (2) *** p<0.01, ** p<0.05, * p<0.1(3) The estimations were carried out with sentinel site fixed effects.

V. Conclusions

The results from probits estimation of enrolment exhibit interesting patterns/ explanations of dropping out of school. As previously mentioned, the economic status of the household,

represented by the wealth index is associated with a higher probability of school enrolment. The education levels of both the parents are significant predictors of the school enrolment probability.

Comparing the results from the SUR and OLS suggests that the SUR estimates are slightly larger and have larger standard errors. Thus, the importance of carrying out an SUR is realized in the results.

The results reveal that after controlling for important factors, the schooling and health outcomes are significantly different for Scheduled Tribes even as SCs suffer relative deprivation only in health outcomes. The Muslims , on the other hand, have significantly lower probability of school enrolment at age 15 but after controlling for that, do not suffer any consequences in school achievement. The health outcomes results, too, are the same for them.

Hence, the research carries important policy implications for the education and health outcomes of the socio religious minorities in India. Even though the reservation exists for higher level of education in India, the secondary school outcomes show the deprivation of such classes. Hence an effective policy, addressing these concerns, is a must for betterment of such classes in India.

References:

1. Basant, R. and Sen, G.(2010). Who Participates in Higher Education in India? Rethinking the Role of Affirmative Action. Economic and Political Weekly. 255(9):62-70.

2. Behrman, Jere R. (1996). The Impact of Health and Nutrition on Education. World Bank Research Observer. 11(1). 23-37.

3. Borooah, V.K. and Iyer, S. (2005). Vidya, Veda and Varna: The Influence of Religion and Caste on Education in Rural India. *Journal of Development Studies*.141(8). 1369-1404.

4. Case, A., Fertig, A.and Paxson. C. (2005). The Lasting impact of Childhood Health and Circumstance. Journal of Health Eocnomics. 24:365-389.

5. Case, A. and Paxson, C. (2008): Stature and Status: Height, Ability and Labor Market Outcomes. *Journal of Political Economy*. 116(3):499-532.

6. Desai, S. and Kulkarni, V. (2008). Changing Educational Inequalities in India in the context of Affirmative Action. *Demography*.45:2.pp. 245-270.

7. Deshpande, A.(2000).Recasting Economic Inequality. *Review of Social Economy.* 58(3). 382-399.

8. Dreze, J. and Kingdon, G.G. (2001). School Participation in Rural India. *Review of Development Studies*.5(1):1-24.

9. Mohindra, K.S., Haddad, S. and Narayana, D. (2006). Women's Health in a rural Community in Kerala: DO caste and Socio-Economic Position matter? *Journal of Health and Social Behavior*.13:276-283.

10. Sabharwal, N.S.(2011). Caste Religion and Malnutrition Linkages. *Economic and Political Weekly*. 46(50)

11. Poel, E. Van de and Speybroeck. (2009). Decomposing Malnutrition Inequalities between Scheduled Castes and Tribes and the remaining Indian Population. *Ethnicity and Health*. 14(3). 271-287.

Appendix

Table 1 Probit results for school enrolment of 15 year olds

Variables	(1) Enrol
SC	0.121
	(0.140)
ST	-0.432*
	(0.230)
Female	-0.144
	(0.107)
Muslim	-0.800***
	(0.243)
Wealth Index	1.388***
	(0.451)
Father's Education	0.0604***
	(0.0162)
Mother Educated	0.571***
	(0.143)
Order	-0.0676**
	(0.0293)
Migrate	-0.812***
	(0.170)
Paidwork	-1.437***
	(0.255)
Rural	-0.866**
	(0.391)
Constant	1.558***
	(0.416)
Observations	959

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1