# SOCIOECONOMIC DETERMINANTS OF SCHOOL PROGRESSION IN PAKISTAN

SIDDIQUI, Anjum [*]
IRAM, Uzma

**Abstract**

Low enrollment and high drop out rates can best be understood by examining a range of socioeconomic factors that affect school progression from primary through secondary to post secondary schools in Pakistan. The study employs a sequential approach which captures the different opportunity costs of education at successive levels of schooling attained by students. The results show that child characteristics, parent's education and household level variables are important determinants of child school progression. Household income and parent's education are significantly and positively related to child schooling. The child's own age as well as the number of siblings (up to age 18) are negatively related to the schooling decision and are an important factor in low enrollment rates and high incidence of dropouts. It was also found that the provision of government schools appeared to be an important predictor of enrollment in Pakistan. The study thus infers that a number of socio economic variables which capture or affect "poverty" are intimately related to the school progression decision.

JEL Classification: I20, I21, I31, O15

Keywords: Child Schooling, School Progression, Socioeconomic Factors, Sequential Probit Model, Parent's education, Sibling composition

## 1. Introduction

South Asia has one of the most alarming education statistics in the world. Millions of children have never set foot in school. In India, one-third of all children, that is, 59 million children aged 6 to 14 do not attend school and an estimated 7 million children do not attend school in Pakistan[1] and about 50 percent of the population has never attended school[2]. With so many children and adults out of school it is not surprising that South Asian adult literacy rate (56%) has slipped behind both Sub Saharan Africa (62%) and Arab States (61%). The school enrollment and primary school completion statistics also paint a sorry picture. In Bangladesh while 60% of poor adolescents complete grade 1, only 36% complete grade 5. Pakistan is even worse off with only one third of primary school goers aged 5-9 years completing grade 5. In India while 62% school goers aged 6-9 complete grade 5, less than 30% adults finish eight years of schooling. Likewise, in Pakistan only 43% females enroll in primary schools compared to 63% males. The gender gap increases substantially in secondary enrollment with only 22% females and 62% males. Tertiary education shows even worst gender equality of opportunity with a mere 2% female enrollment compared to 11% for males[3].

There are no easy answers to explain the millions that never attend school, the low enrollment rates and the high drop out rates as early as grade 5, and the huge gender inequality. Supply side factors alone cannot explain the complex schooling decisions[4].

[*] The authors are Professor of Economics and Director, Center for Business Research and Consultancy at Gulf University for Science and Technology, Kuwait and Economist at Applied Economic Research Centre, University of Karachi, Pakistan respectively.

For example, the availability of a school is no guarantee either that all students in that area will attend or complete primary or secondary schooling[5].Both demand and supply side factors are required to explain the stylized facts on education. The general finding in developing countries is that the decision to go to school is intimately related to the decision to work. However a number of other factors such as parents' education, their employment and health as well as the child's age and the number of siblings and their age composition and the relative level of household poverty are important demand side factors affecting the decision to go to school or drop out. Studies on schooling decisions have investigated a number of determinants for low levels of participation in primary schools and high rates of dropout. Lam and Schoeni (1993) Knight and Sabot (1990) found that the employment status of parents and their schooling achievements increases their children's years of schooling. However mixed results have been obtained for whether a mother's or a father's schooling is more important in explaining school enrolment and completion by children[6] (Duraisamy 1988).

Not only parents' schooling but also their employment is an important variable affecting school enrollment. Jayachandran (2002) concludes that higher levels of work force participation by women could result in some children, particularly girls to stay at home tending to household chores and taking care of the younger siblings. On the other hand, higher rates of work force participation by women can also be expected to positively affect children's enrollment. The positive connection between employment, household income and schooling of children is confirmed in a number of studies (Alderman *et all* 1996). Poor families tend to either not enroll their children or withdraw them early from primary schools. It is not surprising that Jayachandran (2002) found that poverty has a negative and significant effect on child schooling. Schultz (1993) has similarly found that limited financial resources could negatively impact upon schooling choices. These studies provide support to explain why the economic contribution of children encourages parents to have more children and discourages investment in their schooling. Perhaps this also explains the high incidence of child labor in Pakistan and South Asia.

Another strand of literature examines the effect of the number of children and their ages on the labour supply of married women (Connelly, 1992). Various studies in both developing and developed countries have consistently confirmed the inverse relationship between number of siblings and the education of children (Downey 1995).[7] Our study also tests for this important variable as a subset of other socioeconomic factors.
Empirical studies have traditionally employed ordered logit or probit analysis to estimate the determinants of child schooling. There are two shortcomings of these studies. Firstly, they use a static approach by employing ordinary or two stage least squares regressions to determine completed years of schooling, whereas enrollment is gauged through univariate probit/logit models, and grades attained are estimated from ordered probit/logit models. Secondly, such an approach errs by using a common set of explanatory factors to understand the different household decisions based on self selection, as the child progresses through successive levels of schooling.

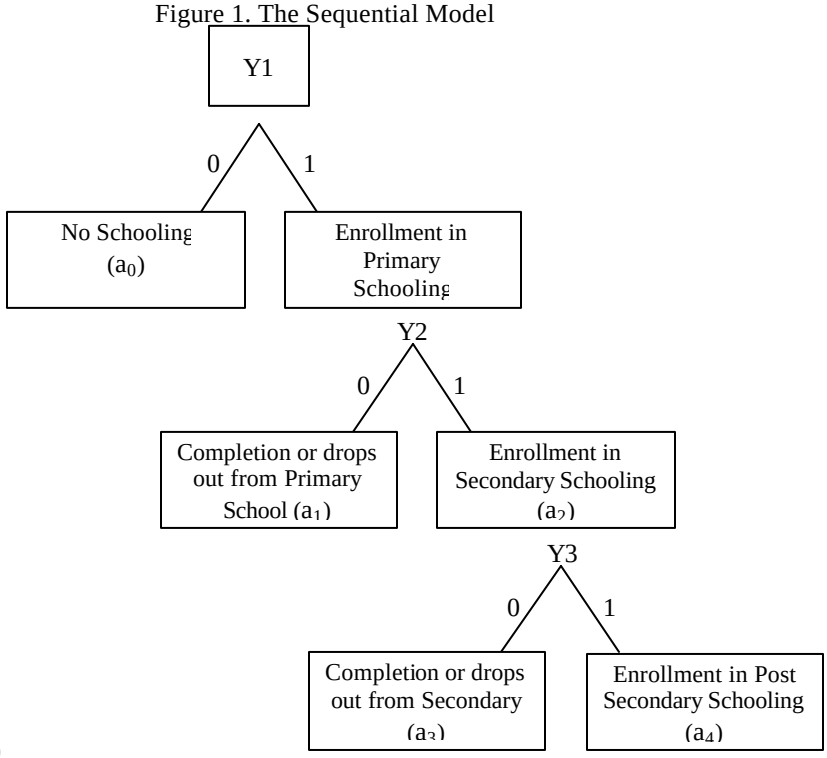
This study examines the effect of sociological factors including the age and number of siblings and some poverty related instrumental variables on the schooling decisions of

households. To circumvent the problem of endogeneity of household expenditure with schooling decisions, we use various instruments to proxy household expenditure, for example, various variables reflecting the household demographics, parents' education and household amenities – some of these amenity related variables also reflect and proxy for household poverty. Section 2 of the paper presents the sequential probit model. Section 3 discusses the data while Section 4 presents the estimated results and Section 5 concludes with some policy recommendations.

## 2. The Sequential Probit Modelling Approach

Earlier studies on developing countries including Pakistan have either used ordered probit or a univariate logit approach (Ray, 2000b; Dreze and Kingdon, 2001; Patrinos and Psacharapoulos 1997). The standard ordered probit or logit models with their specific structure restrict the distributional effects a-priori without the possibility "to let the data speak". We require a model that is sufficiently flexible such that the effect of socioeconomic variables on the probability distribution of child schooling is not fully determined by functional form. The sequential modeling approach of our study follows Waelbroeck (2003). The same methodology has been used by Alpu and Fidan (2004) to estimate the determinants of infant mortality and Pal (2004) to study child schooling in Peru. This study considers a simple sequential model with three qualitative variables, $y_1$, $y_2$ and $y_3$, which are observed sequentially. Let $y1,i$ 1if the child has some primary level of schooling and 0 otherwise. Similarly, $y2,i$ 1 if the child has a some secondary level of schooling and 0 if child drops out before completing secondary schooling. Finally let $y3,i$ 1 if the child has some post secondary level of schooling and 0 otherwise. The sequential model is illustrated in Figure 1.

Figure 1. The Sequential Model




Source: Ozlem(2004)

There are five possible outcomes: $a_0$, $a_1$, $a_2$ $a_3$ and $a_4$. If $y_1=0$ outcome $a_0$ is observed, otherwise depending on the value of $y_2$ there are two additional outcomes: $a_1$= {$y_1=1$, $y_2=0$) which shows completion or dropping out from primary school and $a_2$= ($y_1=1$, $y_2=1$) which also shows completion or dropping out from secondary school. If $y_2=0$ outcome $a_1$ is observed, otherwise depending on the value of $y_3$ there are further additional outcomes; $a_3$={ $y_2=1$, $y_3=0$} and $a_4$={ $y_2=1$, $y_3=1$}. A sample of n observations indexed by the binary value of 1 was considered. It follows that the sequential model has a special feature: the observed values of $y_{2i}$ and $y_{3i,}$ are conditionalon the fact that $y_{1i}=1$.
The mathematical sequential approach is described in the Annex. The empirical sequential probit equation for child schooling used in this study is:

Child Schooling (CEDU)=$\beta_0$ + $\beta_1$(CAGE1) +$\beta_2$(CAGE2) + $\beta_3$(CSEX) + $\beta_4$(SR1) +

$\beta_5$(SR2) + $\beta_6$(SR3) + $\beta_7$(MED) + $\beta_8$(MEMP) + $\beta_9$(FED) + $\beta_{10}$(I) + $\beta_{11}$(WATR) + $\beta_{12}$(SF) + $\beta_{13}$(GAS) +$\beta_{14}$(ELEC) + $\beta_{15}$(HSE) + $\beta_{16}$(DIST) + $\beta_{17}$(GS) + $\beta_{18}$(URBN) + $\mu$

Where the dependent Variables are: CEDU1 Child education (Primary versus no schooling), CEDU2 Child education (secondary versus Primary schooling), CEDU3 Child education (post secondary versus secondary schooling).
Independent Variables are:
CAGE1=Child ages (10-12); 1,Otherwise; 0,
CAGE2=child ages (13-17); 1, Otherwise; 0,
CSEX=Child sex (Equal one for girl, otherwise zero),
SR1=Siblings ratio between the ages 0-6 years.
SR2=Siblings ratio between the ages 7-18 years.
SR3=Siblings ratio between the ages 19-22 years,
MED=Mother's years of schooling, MEMP=Mother's own employment,
FED=Father's years of schooling, I=Log of household annual income,
WATR=Availability of piped water at house (piped water; 1, otherwise: 0),
SF=No sanitation facility at house; (1, otherwise; 0),
GAS=Availability of gas at house (gas; 1, otherwise; 0),
ELEC=Availability of electricity at house (electricity; 1 otherwise; 0),
HSE=Independent house; (1, otherwise; 0),
DIST=Distance from school (Km),
GS=Availability of Government school (school; 1, otherwise; 0),
URBN=Region (Urban; 1, otherwise;0)

## 3. Data

The data for this study is obtained from the Pakistan Integrated Household Survey (PIHS), of 2000-01**.** There are 42,696 male and female children in the age of 10-22 years in our sample. These constitute the total potential population (from this sample) of children who should be enrolled in various levels of schools i.e. primary secondary and post secondary. We consider three sequentially related transition decisions about schooling pertaining to enrolling in primary school, progressing to secondary school and finally pursuing post secondary education. It was calculated from the PIHS data that of the total sample of 42,696 children in the age group of 10-22 years, 28 % (11,949

potential students) have never attended school or have dropped out very soon after enrolling in a primary school and for all practical purposes the survey considers them illiterate. Even more disturbing is the statistic that of those 30,747 students who enroll in primary schools 63% (19,283) drop out or do not progress to secondary schools and a mere 4% (2,694) progress from secondary to post secondary institutions. Those who did not progress to post secondary levels included both those who could not finish secondary education as well as those who completed secondary schools but did not wish to pursue post secondary education. Aside from school progression which was calculated on the basis of the 30,747 students who enrolled in primary schools, if we now consider educational attainment on the basis of the total sample population of 42,696 potential students, the survey reveals that 72% of children have some level of primary education while only 26% have some level of secondary schooling and only 6% of these children have some post-secondary education. Table 1 presents frequencies of the variables from the PIHS data that have been used in estimations of this paper.

Table 1. Frequencies of Selected Socioeconomic Variables

| Group | Variables | Number | % |
|---|---|---|---|
| Child' Age | CAGE1 (10-12) | 10072 | 23 |
| | CAGE2 (13-17) | 19955 | 47 |
| | CAGE3 (18-22) | 12669 | 30 |
| | Total Sample Size | 42696 | 100 |
| Child' Gender | Male | 24806 | 58 |
| | Female | 17890 | 42 |
| Mother's Education (Years) | Illiterate | 5543 | 13 |
| | Primary | 23984 | 56 |
| | Secondary | 12481 | 29 |
| | High | 688 | 2 |
| Father's Education (Years) | Illiterate | 4645 | 11 |
| | Primary | 5413 | 13 |
| | Secondary | 29299 | 69 |
| | High | 3339 | 07 |
| Household Annual Income | Low income | 9357 | 22 |
| | Middle income | 3076 | 07 |
| | High income | 30263 | 71 |
| Dwelling Type | Independent House | 33239 | 78 |
| | Compound House | 3109 | 07 |
| | Other House Type | 6348 | 15 |
| Sources of Water Availability | Piped inside house | 26351 | 61 |
| | Piped outside house | 15679 | 37 |
| | Other sources | 666 | 02 |
| Sanitation Facilities | Flush Toilet | 20659 | 48 |
| | No flush toilet | 8649 | 20 |
| | No toilet at all | 13418 | 32 |
| Availability of Gas | Gas Connection (Yes) | 10781 | 25 |
| | Gas connection (No) | 31915 | 75 |
| Availability of Electricity | Electricity connection (yes) | 34645 | 81 |
| | Electricity connection (No) | 8051 | 19 |

Source: PIHS 2000-01

The table shows that in Pakistan the largest proportion of children fall in the 13-17 year bracket while the younger children in the age group of 10-12 years are 23% of the children population. Parent's education is also an important contributor towards children's school progression. The sample shows that 13% of mothers and 11% of fathers are illiterate. However, 69% fathers and 29% mothers have attained secondary education which clearly reflects the gender bias against women in secondary education in Pakistan.

A household's economic and social well being is measured by its income, the type of dwelling that they inhabit and the facilities of water, power and sanitation that are available to them and their children. In this sample 22 percent households belong to the low-income level group, 7 percent belong to the medium income group and 70 percent households are in the high-income group. While 78 percent households owned an independent house, only 61 percent households have access to drinking water within their residence, while 37 percent households have drinking water outside their residence. 32 percent of the total households have no toilets at all within their housing premises and only 48 percent households reported the presence of flush toilet in their residence. 25 percent households have gas connections and 81 percent households report the availability of electricity connection in their residence. The lack of amenities is then associated with lower educational access and enrollment as verified in the results of this study in next section.

## 4. Empirical Results

The selected socioeconomic factors used in the empirical equation can be grouped as those pertaining to gender and age of students, ages of siblings as a proportion of the household size, parents schooling and employment, and variables that capture the well being and amenities enjoyed by a household e.g. availability of utilities and home ownership. These later group of variables also capture in a broader sense the relative poverty of households. Table 2 presents the empirical results for overall Pakistan while Tables 3 and 4 in the Annex present further results for the urban and rural areas.

In low income developing countries like Pakistan parents engage their children not only in child labour for additional income but they also utilize their children, especially older siblings, in child rearing responsibilities. This practice has a negative impact both on enrollment as well as on school progression of children. The estimates of this study indicate the negative and significant impact of child age (CAGE) on school progression for all three levels i.e. primary, secondary and post secondary education. While age exerts a mild effect on the probability of dropping out of school at younger ages, this effect becomes important from about 13 years up to the age of 22. These measured effects of age possibly capture the urgency in poor families to put the children to work which intensifies as the child gets older. The effect of gender on schooling progression is examined through a dummy variable (CSEX). At the primary and secondary level of schooling there is a gender bias against females. There is a 72 percent probability that female students will not complete primary school and an 18 percent probability that they will not complete secondary school. However, females have a 13 percent probability of

completing post secondary education. In the South Asian and Pakistani context low post secondary completion by females can be explained by the common fact that they usually quit schooling in favour of marriage and child bearing. It needs emphasis that in South Asia the decision to attend school or drop out is not a choice which children make rather it's the parents who play a principal role in this decision. Strong social and cultural pressures are also responsible for parents to not send their teen age girls to school and to marry them off. This is an important factor which explains low secondary and post secondary educational attainment by females.

Table 2. Sequential Probit Estimates of Child Schooling (Pakistan)

| Variable | Primary schooling vs. no schooling | | From primary to secondary school | | From secondary to post secondary school | |
|---|---|---|---|---|---|---|
| | Coeff. | t Stat. | Coeff. | t Stat. | Coeffi. | t Stat. |
| CONSTANT | 1.08 | 6.67* | -0.62 | 3.84* | -3.08 | 13.15* |
| CAGE1 | -0.75 | 28.02* | -1.91 | 60.94* | -2.03 | 24.07* |
| CAGE2 | -0.21 | 10.29* | -0.41 | 16.62* | -1.16 | 32.72* |
| CSEX (Girl) | -0.72 | 40.59* | -0.18 | 9.63* | 0.13 | 4.38* |
| SR1 | -0.85 | 7.39* | -0.62 | 4.78* | 0.03 | 0.15 |
| SR2 | 0.17 | 2.69** | -0.11 | 1.64*** | -0.06 | 0.58 |
| SR3 | 0.28 | 2.68** | 0.63 | 4.94* | 1.07 | 6.96* |
| MED | -0.001 | 0.19 | 0.05 | 6.78* | 0.04 | 4.79* |
| MEMP | -0.03 | 1.07 | -0.18 | 5.19* | -0.25 | 4.16* |
| FED | 0.02 | 5.04* | 0.05 | 14.67* | 0.06 | 12.73* |
| I | 0.05 | 4.27* | 0.05 | 3.87* | 0.08 | 4.72* |
| WATR | 0.01 | 0.44 | -0.001 | 0.03 | -0.02 | 0.63 |
| SF | -0.29 | 13.47* | -0.22 | 8.73 | -0.22 | 4.53* |
| ELEC | 0.62 | 29.06* | 0.29 | 10.74* | 0.05 | 0.93 |
| GS | 0.41 | 14.56* | 0.21 | 8.38* | 0.21 | 6.03* |
| HSE | 0.01 | 0.28 | 0.03 | 0.80 | 0.04 | 0.74 |
| DIST | -0.16 | 41.57* | -0.02 | 7.72* | 0.04 | 11.72* |
| GAS | 2.04 | 39.15* | 0.61 | 24.31* | 0.85 | 24.0* |
| URBN | 0.08 | 3.10* | 0.10 | 3.78* | 0.03 | 0.89 |
| | N 36525<br>McFadden $R^2$ =0.42<br>LR Stat (18df)=19479<br>Probability(LRStat.)=0.000 | | N 24078<br>McFadden $R^2$=0.24, LR Stat(18df)= 7871<br>Probability (LRStat)=0.000 | | N 14158<br>McFadden $R^2$= 0.23, LR Stat(18df) = 3232<br>Probability (LR Stat)=0.000 | |

Note: at 1% level of significance* at 5% level of significance** at 10% level of significance***

Table 3. Sequential Probit Estimates of Child Schooling (Urban Pakistan)

| Variable | Primary schooling vs. no schooling | | From primary to secondary school | | From secondary to post secondary school | |
|---|---|---|---|---|---|---|
| | Coefficients | T. Statistics | Coefficients | T. Statistics | Coefficients | T. Statistics |
| Constant | 1.27 | 4.54* | -0.45 | 1.79*** | -3.36 | 10.22* |
| CAGE1 | -0.70 | 14.68* | -2.04 | 42.44* | -2.59 | 19.60* |
| CAGE2 | -0.26 | 7.47* | -0.46 | 11.78* | -1.35 | 27.99* |
| CSEX | -0.40 | 13.35* | 0.02 | 0.79 | 0.21 | 5.74* |
| SR1 | -1.41 | 6.92* | -0.38 | 1.85*** | 0.01 | 0.02 |
| SR2 | 0.01 | 0.12 | -0.14 | 1.42 | -0.002 | 0.01 |
| SR3 | 0.42 | 2.39** | 0.95 | 5.02* | 1.24 | 6.27* |
| MED | 0.02 | 1.82*** | 0.05 | 6.32* | 0.05 | 4.81* |
| MEMP | -0.14 | 2.97* | -0.24 | 5.10* | -0.35 | 4.58* |
| FED | 0.01 | 0.65 | 0.05 | 9.71* | 0.07 | 10.69 |
| I | 0.08 | 3.92* | 0.06 | 3.35* | 0.09 | 4.13* |
| WATR | 0.04 | 1.18 | -0.04 | 1.23 | 0.04 | 0.91 |
| SF | -0.51 | 7.79* | -0.26 | 3.32* | -0.19 | 1.36 |
| ELEC | 0.58 | 11.11* | 0.39 | 5.67* | 0.26 | 2.06** |
| GS | 0.39 | 11.96* | 0.20 | 6.75* | 0.21 | 5.30* |
| HSE | -0.17 | 3.34* | -0.07 | 1.54 | 0.10 | 1.39 |
| DIST | -0.18 | 28.90* | -0.45 | 11.77* | 0.01 | 1.79*** |
| GAS | 1.63 | 16.56* | 0.42 | 11.52* | 0.89 | 17.73* |
| | N 13955<br>McFadden R-sq 0.34<br>LR Stat (17df) 4748.274<br>Prob ( LR Stat) 0.000 | | N 11149<br>McFadden R-sq 0.24<br>LR Stat (17df) 3411.633<br>Prob ( LR Stat) 0.000 | | N 7593<br>McFadden R-sq 0.26<br>LR Stat (17df) 2156.288<br>Prob ( LR Stat) 0.000 | |

Note: at 1% level of significance* at 5% level of significance** at 10% level of significance***

It has been argued that sibling composition within a family may play an important role in a child's school progression, particularly if the child comes from a poor and resource constrained household. As the age of the siblings matters in schooling and dropout decisions, we classify siblings by age: SR4, the number of siblings in the age group (0 - 6), SR2, the number of siblings in the age group (7 - 18) and SR3, the number of siblings in the age group (19 - 22). It is found that sibling composition exerts a significant impact on child school progression. The number of relatively younger siblings (up to the age of 18 years) in a household has a negative effect on schooling of children. This points out that the larger the number of siblings, the smaller the amount of money that is invested in the education of each child. The study also indicates that the sibling variables SR2 and SR3 are positively and significantly related to a child's primary schooling. This means that the probability for the child to go to school is positively associated with the presence of prime-age siblings (above 18) in the household. The explanation may be that these older siblings lower the demand of younger children's time for doing household chores or earning income and free the children to go to school, while more infant siblings lower the attainment at the primary and secondary level of schooling. Our conclusions find support in an earlier study by Sawada and Lokshin (2001) who explained that education of school-age children is supported by the earning contribution of the elder siblings.

Table 4 Sequential Probit Estimates of Child Schooling (Rural Pakistan)

| Variable | Primary schooling vs. no schooling | | From primary to secondary school | | From secondary to post secondary school | |
|---|---|---|---|---|---|---|
| | Coefficients | t Statistics | Coefficients | t Statistics | Coefficients | t Statistics |
| Constant | 1.30 | 5.72* | -0.76 | 3.27* | -2.83 | 7.51* |
| CAGE1 | -0.78 | 23.82* | -1.87 | 43.96* | -1.39 | 12.06* |
| CAGE2 | -0.18 | 7.09* | -0.41 | 12.45* | 0.93 | 17.08* |
| CSEX | -0.89 | 40.17* | -0.35 | 13.22* | 0.01 | 0.14 |
| SR1 | -0.54 | 3.83* | -0.73 | 4.38* | 0.03 | 0.11 |
| SR2 | 0.28 | 3.62* | -0.06 | 0.64 | -0.09 | 0.59 |
| SR3 | 0.21 | 1.60*** | 0.39 | 2.30** | 0.86 | 3.45* |
| MED | -0.07 | 3.28* | 0.04 | 2.51** | 0.02 | 0.69 |
| MEMP | 0.06 | 1.47 | -0.09 | 1.74*** | -0.08 | 0.77 |
| FED | 0.04 | 6.11* | 0.06 | 10.69* | 0.05 | 6.48* |
| I | 0.03 | 2.08** | 0.03 | 2.02** | 0.07 | 2.71** |
| WATR | -0.01 | 0.44 | 0.05 | 1.85*** | 0.01 | 0.23 |
| SF | -0.29 | 12.07* | -0.23 | 8.49* | -0.25 | 4.88 |
| ELEC | 0.67 | 27.85* | 0.31 | 10.24* | 0.06 | 0.99 |
| GS | 0.29 | 4.59* | 0.15 | 2.67** | 0.18 | 2.25** |
| HSE | 0.11 | 2.91** | 0.12 | 2.56** | -0.07 | 0.78 |
| DIST | -0.14 | 28.84* | 0.01 | 2.57** | 0.07 | 14.74* |
| GAS | 2.16 | 35.35* | 0.77 | 21.16* | 0.63 | 12.08* |
| | N 22571<br>McFadden R-sq. 0.42<br>LR Stat (17df) 12963<br>Probability(LR Stat) 0.000 | | N 12929<br>McFadden R-sq. 0.22<br>LR Stat (17df) 1010<br>Probability(LR Stat)0.000 | | N 6565<br>McFadden R-sq. 0.20<br>LR Stat (17df) 1058<br>Probability(LR Stat) 0.000 | |

Notes: at 1% level of significance* at 5% level of significance** at 10% level of significance***

It is generally perceived that schooling of a child is positively related to the parent's education, indicating complementarity between the schooling of a child and parents' education. This complementarity is generated possibly by educated parent's positive perception for child schooling. As expected, the probability of school progression increases with parental education (both maternal and paternal). The results also show that mother's education (MED i.e. years of schooling), has a positive and strongly significant effect on various levels of child school progression, although the coefficient had a negative sign and it was insignificantly related with child primary schooling. Similarly, a father's years of schooling (FED) is positively associated with all levels of school progression. The results thus show that education of parents in Pakistan significantly enhances the likelihood of primary education of their children. As literacy of parents is very low, our study provides strong evidence of inter-generational persistence of illiteracy in families. This is not an ordinary finding and needs to be emphasized within the context of the South Asian Region. Illiterate parents do not have an appreciation for education for their children because they cannot see the link between education and earnings. For them the immediate and short term concern is economic survivability which is helped if their children are sent off for child labor. However, while it helps in the short

term it certainly creates a vicious circle of illiteracy and poverty in the long run. A related factor is the effect of mothers' poor education on higher fertility levels and consequently higher population rates. Using aggregate macro data Guisan and Exposito (2003) have examined the economic growth rate rates in a number of Asian countries and observe that education, fertility and growth rates are strongly interrelated. Our study which uses micro household level data supports these inferences and finds a negative effect of age and number of siblings on enrollment. One can thus conclude that low level of parents' education results in higher fertility of mothers, more siblings, and consequently negative effects on school enrolment.

Another intimately related issue and its effect on enrollment is a mother's employment. We found that a mother's employment participation (MEMP) has a negative and significant impact on all levels of child school progression and in explaining the high drop out rates from primary school. This result is not surprising in the context of relatively poorer families in South Asia and Pakistan as working women in rural areas involve their children in farm work and urban working women from poorer households also engage their children in work with them. The result shows some substitution between employment of poor mothers and their children's primary schooling.

One has to be careful in not jumping to erroneous conclusions regarding causality from mothers work to low enrollment. It needs emphasis that mothers work is not the "cause" of low enrollment and school drop outs; rather it is the consequence of poverty and poor social infrastructure in which the government has been negligent and can play an important role to correct the situation. When women go to work and they have babies (non primary school going), they usually take their younger children (5-10 years) with them to take care of the babies while they are busy in work. These younger children (who should infact be attending school) are also taken to work because they can get some food at the women's work place. If there were free public day care centers in Pakistan, the babies and toddlers would have gone to daycares; and if there were free primary schools provided by the government in the vicinity of the household as is the case in many developed countries, the primary school going children who accompany their mothers to work would have rightfully gone to schools.

Another factor that explains low enrollment is that poor working parents encourage their older children between 10 years and older to be employed in some kind of child labor in both urban and rural areas. 13 to18 year old children are in most demand for child labor and this is the age when a child should be attending secondary school. This fact explains why a number of children from poorer households do not attend secondary school. Thus, child labor is another factor in low school attendance and in high drop out rates. One important factor which lies at the heart of low enrollment is that due to generally high illiteracy as well as poverty the importance of education is not clear to millions of Pakistani families. The government has not been able to mount any worthwhile education awareness programs or to provide schools and incentives for the poor people to send their children to school, rather than be employed in child labor. It is interesting that when school lunch is provided parents have sent their children to school in larger numbers. One such success story has been in Bangladesh (see Siddiqui 2007).

As poverty is an important detriment to educational access, we specifically examined the impact of household poverty on child schooling by using a number of variables: annual household income (I, continuous variable), availability of electricity (ELEC), gas (GAS), safe drinking water (WATR) and sanitation facilities (SF) – four binary variables. The non-income variables are being used as proxies of household standard of living and wealth. We found a strong positive relation between a household's income and child school progression. The positive coefficient of income indicates that schooling is a normal good and the growth in income will increase enrollment as well as school progression. Households with higher incomes give more weight to education as they can afford the cost of schooling. Thus the probability for children to attend school from such households is high. Other than income our results also highlight the importance of standard of living variables or poverty related variables. It was found that amenities such as electricity (ELEC), gas (GAS) and sanitation facilities (SF) are positively related to schooling implying that decrease in relative poverty increases the probability of child schooling. Another standard of living variable is ownership of an independent house (HSE). However, this variable was found to have an insignificant impact on child school progression, though its sign was correctly positive. The result suggests that there are other factors such as availability of school in the vicinity of the residence which may have a much stronger bearing on the schooling decision.

We found that the probability of enrollment in school and school progression is significantly and positively associated with the presence of public schools in Pakistan. The significant and negative coefficient sign for distance of school from residence (DS) shows that the larger the distance from school, the lower the probability of a child's primary and secondary schooling. Young children cannot walk large distances and in the absence of an adequate transportation system parents are less willing to send their children to school. The lack of schools in rural areas and large distances of such schools is another factor in low attendance rates and higher drop out rates. In developed countries school buses are common and hence school distance is usually not a factor in low enrollment, but there is no such busing system in the rural areas of Pakistan. Similarly, in urban areas the government does not provide any busing and private hires make urban commuting expensive and out of reach of a number of poorer households with negative consequences for enrollment.

The study also found that urban locations (URBN) have a positive impact on all levels of schooling although it is statistically insignificant at the post secondary level. The positive coefficient on URBN signifies that residing in an urban area increases the probability of progressing through successively higher levels of schooling, thereby reducing the probability of urban illiteracy. The results that were found for overall Pakistan in Table 2 are now verified through a further examination in urban and rural areas. Tables 3 and 4 present sequential estimation results for urban and rural areas. The results substantiate the importance of parents' education on their children's educational achievement. The impact of poverty related variables is again substantiated with household income and facilities showing a positive association with enrollment and school progression at all levels of education.

## 5. Conclusion

This study investigates the determinants of child school progression in Pakistan among children 10-22 years through estimating a sequential probit model for the primary, secondary and post-secondary levels of schooling for overall Pakistan and also for urban and rural areas. The results suggest that a child's age, siblings' composition, parent's education, household income and standard of living variables and provision of government schools significantly affect child schooling in our sample. However, these factors affect various levels of schooling differently when we take account of the correlated nature of sequential school progression. Some important lessons and confirmations emerge from the results presented in the study. There is a positive association between household income and child school progression and literate parents are more likely to send their children to school- all three levels of schooling were found to be strongly associated with a parent's years of schooling.

Mother's employment status was found to have a negative impact on child school participation for overall Pakistan, while it was positively but insignificantly associated with child primary schooling in rural areas. This econometric finding needs to be interpreted with care so as to avoid erroneous conclusions. The true underlying reason for this result is the poor provision of social programs viz. day care centers and income subsidies in order to allow compatibility between mothers work and children schooling as it is the most common situation in other countries. The impact of child gender confirms gender discrimination against women in primary and secondary level of schooling in Pakistan. Regional effects bring out that the gender bias in school progression is particularly high in rural than urban areas. The low probability for girls schooling may also be the result of social and cultural barriers imposed by both poor as well as more affluent but conservative parents in rural and urban areas.

Furthermore, gender discrimination in schooling is also the consequence of childcare and domestic responsibilities borne by women.It is also found that, the number of children and their age groups (within a family) impacts on a child's schooling. The number of siblings (up to the age 18 years) in the household has shown a negative effect on education of school age children in Pakistan indicating that poverty increases in a household with more siblings. This study shows that probability for the child to go to school is positively associated with the presence of prime-age (above 18) in the household. It may be that such children are more likely to work, which improves family income and enhances the probability of younger school age children to attend school. The study also found that a household's standard of living and provision of government schools appeared to be important predictors of child school progression in Pakistan. Specifically we found that larger traveling distances from residences to schools lowered the probability of child schooling.

The principal policy implications of the results presented in this study pertain to the potential social and economic benefits of improving education, particularly mother's education, in rural areas. The availability of school going incentives and subsidies as well as the provision of public schools in rural areas will surely counteract the negative impact of mothers' employment on schooling in rural areas. Without economic subsidies

to the poorer segments of society, educational policies will surely not succeed as parents will continue to substitute child labor for education of their children. Pakistan spends the lowest percentage of its GDP on education amongst SAARC countries[8] which has resulted in the vicious cycle of poor education, low earnings, child labor and low enrollment. The study highlights that an increase in the number of younger siblings deters enrollment and encourages drop outs. This calls for a viable and effective fertility control programme by the government which has been neglected so far and has contributed to illiteracy and poverty. Moreover the study once again confirms that policy interventions are direly needed to focus on improving female enrollment and reducing the gender bias.

**References**

Alderman, H. J. Behrman, S. Khan and R. Sabot (1996). "Decomposing the Regional Gap in Cognitive Skills in Rural Pakistan", Journal of Asian Economics, **7**(1), 49 – 76.

Behrman, J. R. and B. L. Wolfe (1987). "Investments in Schooling in Two Generations in Pre-revolutionary Nicaragua: The Roles of Family Background and School Supply", Journal of Development Economics, 27(1-2), 395 – 419.

Connelly,R.,(1992). " The effect of Child care and Cost on Married Women's Labour Force Participation", Review of Economics and Statistics, 74(1)

Drèze, J., and Kingdon, G. G., (2001), "School Participation in Rural India", Review of Development Economics, 5(1), pp. 1-24.

Downey, Douglas B. 1995. "When Bigger is Not Better: Family Size, Parental Resources, and Children's Educational Performance." *American Sociology Review* 60:746-761.

Durais amy, P (1988), "An Econometric Analysis of Fertility, Child Schooling and Labour Force Participation of Women in Rural Indian Households" Journal of Quantitative Economics, 4.Formulation." IPAR Special Report, Nairobi. from India", The Development Economics Discussion Paper Series, STICERD,

Gao, C., Lahiri, K., and Wixon, B., (2001). "Value of Sample Separation Information in a Sequential Probit Model", Review of Economics and Statistics, 348-361.

Guisan, M.C. and Exposito P. (2003). "Education, Industry, Trade and Development in Asia-Pacific Countries in 1980-99" in Applied Econometrics and International Development, Vol.3-3, pp117-42.

Knight, J.B. and Sabot, R.H. (1990). Educational Productivity and Inequality: The East African Natural Experience. New York: Oxford University Press.

Lam, D. and Schoeni, R. (1993). "Effects of Family Background on Earnings and Returns to Schooling: Evidence from Brazil". Journal of Political Economy, 101, Number 4.

Özlem Alpu and Hatice Fidan (2004). "Sequential Probit Model for Infant Mortality Modelling in Turkey". Journal of Applied Sciences 4 (4): 2004

Özarici, Ö., (2002). "Bivariate Probit Model with Full Observability and Heteroscedasticity and an Application", Ph.D.Thesis, Osmangazi University, Turkey.

Patrinos, H. A. and G. Psacharopoulos. 1997. 'Family Size, Schooling and Child Labour in Peru – An Empirical Analysis', Journal of Population Economics, pp. 387-405.

Ray, R. 2000a. 'Analysis of Child Labour in Peru and Pakistan: A Comparative Study', *Journal of Population Economics* 13(1), pp. 3-19.

Ray, R. 2000b. 'Child Labour, Child Schooling and Their Interaction with Adult Labour: Empirical Evidence for Peru and Pakistan', *The World Bank Economic Review*, 14(2), pp. 347-67.

Sawada, Y. and M. Lokshin (2001), "Household Schooling Decisions in Rural Pakistan", mimeo, World Bank, Country Economics Department.

Schultz, P.T. (1993). "Investments in the Schooling and Health of Women and Men: Quantity and Returns". .Journal of Human Resources, 28, 694-734.

Schultz, T.P. (1993). "Returns to Women's Education" in "Women's Education in Developing Countries: Barriers, Benefits and Policies", eds. E.M. King and M.A. Hill, Baltimore and London: The Johns Hopkins University Press for the World Bank, pp. 51-99.

Sarmistha Pal (2004). "Child Schooling In Peru: Further Evidence from a Sequential Analysis of School Progression". Journal of Population Economics, vol. 17, issue 4, pp. 657-680.

Siddiqui Anjum (2007). "Human Capital Investment and Development" in "India and South Asia: Economic Developments in the Age of Globalization" ed. Siddiqui Anjum, M.E. Sharpe Publishers, New York.

Usha Jayachandran (2002). "Socio-Economic Determinants of School Attendance in India". Working Paper No.103. Centre for Development Economics.

Waelbroeck, P., (2000). "Effects of Information Search on Innovation Decisions: Bayesian Analysis of the Sequential Probit Model", GREQUAM, 25.

Waelbroeck, P., (2002). "Econometric Analysis of the Sequential Probit Model: Application to Innovation Surveys", ECARES and FNRS, 25.

Waelbroeck, P., (2003). "Comparison of Simulated Maximum Likelihood and Bayesian MCMC in the Sequential Probit Model: A Monte Carlo Study", ECARES and FNRS, 21.

On line Annex at the journal web site

Journal published by the EAAEDS: http://www.usc.es/economet/eaa.htm

**Appendix**

The sequential approach departs from earlier tradition to develop a dynamic sequential framework of school progression conditional on attainment at the previous level and self-selection into the next higher level of educational attainment (Pal 2004). This approach enables us to integrate both school enrolment and attainment in a single framework and yields selectivity corrected estimates with respect to successive schooling decisions from primary to higher levels of enrollment. However, the sequential nature of the observations raises econometric issues if the unobservable variables associated with each stage are correlated. In the sequential model it is assumed that the unobservable variables follow a multivariate normal distribution and is referred to as the sequential probit model.

The mathematical version of the sequential probit model is given here.
For notational convenience, it is assumed that data are sorted according to the values of $y_1$, $y_2$ and $y_3$. In other words the first $n_1$ observations correspond to outcome $a_0$ ($y_1$=0) and the next $n_2$ observations to outcome $a_1$ or $a_2$ depending on the value of $y_2$, and the next $n_3$ observations to outcome $a_3$ or $a_4$ depending on the value of $y_3$. We associate with stage j (1, 2 or 3) a latent variable $y^*_{j,i}$ such that

$$y_{1,i} = \begin{cases} 1 \; if \; y^*_{1,i} \geq 0 \\ 0 \; otherwise. \end{cases} \qquad (1)$$

Continuous latent variables are modeled as

$$\begin{aligned} y^*_{1,i} &= x_{1,i}\beta_1 + \varepsilon_{1,i} && i = 1,2\ldots,n \\ y^*_{2,i} &= x_{2i}\beta_2 + \varepsilon_{2,i} && i = n_1+1,\ldots n \\ y^*_{3,i} &= x_{3,i}\beta_3 + \varepsilon_{3,i} && i = n_1+2\ldots\ldots n \end{aligned} \qquad (2)$$

Where the decision to move from one level to the next higher level depends on a set of covariates Xi, which vary by the schooling decision, individual child, and also the household. Here $x_{1,i}$ , $x_{2,i}$ and $x_{3,i}$ are vectors of explanatory variables of respective dimension $k_{1*1}$, $k_{2*1}$ and $k_{3*1}$, $\beta_1$ , $\beta_2$ , $\beta_3$ are vectors of parameters to be estimated of respective dimensions $k_0, k_1$ and $k_2$ , and $\varepsilon_{1,i}$, $\varepsilon_{2,i}$ *and* $\varepsilon_{3,i}$ are vectors of error terms[1]. We can write the model in matrix notations[2]. Let $X_1 = (x_{1,1}, \ldots, x_{1,n})$ , $X_2 = (x_{2,1}, \ldots, x_{2,n2})$ $X_3 = (x_{3,1}, \ldots, x_{3,n})$ , $\varepsilon = (\varepsilon_{1,1}, \ldots, \varepsilon_{3,n})$ , $\varepsilon_1 = (\varepsilon_{1,1}, \ldots, \varepsilon_{1,n})$ , $\varepsilon_2 = (\varepsilon_{2,1}, \ldots, \varepsilon_{2,n2})$ , $(\varepsilon)_1 = (\varepsilon_{1,1}, \ldots, \varepsilon_{1,n})$ , $y_1^* = (y^*_{1,1}, \ldots, y^*_{1,n})$ $y_2^* = (y^*_{2,1}, \ldots, y^*_{2,n2})$ $y_3^* = (y^*_{3,1}, \ldots, y^*_{3,n})$

Now, the latent model can be written as

$$Y^* = X\beta + \varepsilon$$

where $y^* = (y^*_1, y^*_2, y^*_3) = (\beta_1, \beta_2, \beta_3) + (\varepsilon_1, \varepsilon_2, \varepsilon_3) = \beta + \varepsilon$ and $\varepsilon$

[1] See Özarici, (2002)
[2] see Waelbroeck, ( 2000)

$$X = \begin{bmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{bmatrix} (n_1 + n_2 + n_3)(k_1 + k_2 + k_3)$$

? $_1$? = (?$_{1i}$? ? $_{2i}$? ? $_{3i}$values? and covariance are? are assumed to be bivariate normal with mean ( 0,0,0) and covariance? independently and normally distributed with mean (0, 0, 0)

$$\begin{bmatrix} s_{11} & r_{01} & r_{02} \\ r_{01} & s_{22} & r_{12} \\ r_{02} & r_{12} & s_{33} \end{bmatrix}$$

Where $s_{11}, s_{22}, s_{33}$ denote the variances of the unobserved variable at each stage and $r$ represents the covariance between the error terms.

First, it is noted that is diagonal, all latent variables are independent and the coefficient? when of the model can be estimated by three standard probit regressions. Second, it is clear that multiplying each latent equation by a positive constant does not affect the qualitative variables $y_1$, $y_2$ or $y_3$. Hence, it is impossible to identify both location and scale parameters of these equations. Therefore, it is needed to impose three restrictions. We have restricted ($s_{11}, s_{22}, s_{33}$=1) so that coefficients $r$ have the natural interpretation of correlation coefficients[3].

The probabilities of the different options are written as follows[4].

$$\begin{aligned}
&P(y_{1,i}=0) = P(\varepsilon_{1,i} \le -x'_{1,i}\beta_1) = \Phi(-x'_{1,i}\beta_1) \\
&P(y_{1,i}=1, y_{2,i}=0) = P(e_{1,i} > -x'_{1,i}, e_{2,i} \le -x'_{2,i}\, b_2, r) = \Phi_2(x'_{1,i}\, b_1, -x'_{2,i}\, b_2, r) \\
&P(y_{1,i}=1, y_{2,i}=1) = P(e_{1,i} > -x'_{1,i}\, b_1, e_{2,i} > -x'_{2,i}\, b_2, r) = \Phi_2(x'_{1,i}\, b_1, x'_{2,i}\, b_2, r) \\
&P(y_{2,i}=1, y_{3,i}=0) = P(e_{2,i} > -x'_{2,i} b_2, e_{3,i} \le -x'_{3,i} b_3, r) = \Phi_3(x'_{2,i}\, b_2, -x'_{3,i} b_3, r) \\
&P(y_{2,i}=1, y_{3,i}=1) = P(e_{2,i} > -x'_{2,i} b_2, e_{2,i} > -x'_{3,i}, b_3\, r) = \Phi_3(x'_{2,i} b_2, x'_{3,i}, b_3, r)
\end{aligned} \tag{4}$$

where $\Phi$, $\Phi_2$ and $\Phi_3$ are cumulative distribution function (c.d.f.) of the univariate and bivariate standard normal distribution, respectively.

If we assume that $e_1$, $e_2$ and $e_3$ are independent, then Eq.(4) can be written as follows[5].

$$\begin{aligned}
&P(y_{1,i}=1, y_{2,i}=0) = P(\varepsilon_{1,i} > -x'_{1,i}\beta_1).P(\varepsilon_{2,i} \le -x'_{2,i}\beta_2) = \Phi(x'_{1,i}\beta_1).\Phi(-x'_{2,i}\beta_2) \\
&P(y_{1,i}=1, y_{2,i}=1) = P(e_{1,i} > -x'_{1,i}\, b_1).P(e_{2,i} > -x'_{2,i}\, b_2) = \Phi(x'_{1,i}\, b_1).\Phi(x'_{2,i}\, b_2) \\
&P(y_{2,i}=1, y_{3,i}=0) = P(e_{2,i} > -x'_{2,i}\, b_2).P(e_{3,i} \le x'_{3,i}\, b_3) = \Phi(x'_{2,i}\, b_2).\Phi(-x'_{3,i}\, b_3) \\
&P(y_{2,i}=1, y_{3,i}=1) = P(\varepsilon_{2,i} > -x'_{2,i}\beta_2).P(\varepsilon_{3,i} > -x'_{3,i}\beta_3) = \Phi(x'_{2,i}\beta_2).\Phi(x'_{3,i}\beta_3)
\end{aligned} \tag{5}$$

[3] See Waelbroeck (2002)
[4] See Waelbroeck (2003)
[5] See Waelbroeck (2002)

Using the probabilities given above, likelihood function of the sequential probit model is

$$L(b_1, b_2, b_3, r) = \prod_{i=1}^{n} P_{00,i}^{(1-y_{1,i})} . P_{10,i}^{y_{1,i}(1-y_{2,i})} . P_{11,i}^{y_{1,i} . y_{2,i})} . P_{10,i}^{y_{2,i}(1-y_{3,i})} . P_{11,i}^{y_{2,i} . y_{3,i}} \tag{6}$$

Taking the natural logarithm of likelihood function $L(b_1, b_2, b_3, ?)$, we obtain.

$$\begin{aligned} \ln L(b_1, b_2, b_3, r) = \sum_{i=1}^{n} \{ & (1-y_{1,i}) \ln P_{00,i} + y_{1,i}(1-y_{2,1}) . \ln P_{10,i} + \\ & y_{1,i} . y_{2,i} . \ln P_{11,i} + y_{2,i}(1-y_{3,i}) . \ln P_{10,i} + y_{2,i} . y_{3,i} . \ln P_{11,i} \} \end{aligned} \tag{7}$$

If the error terms are independent $(? = 0)$, natural logarithm of likelihood function becomes[6].

$$\begin{aligned} \ln L b_1, b_2, b_3 = \sum_{i=1}^{n} \{ & (1-y_{1,i}) . \ln \Phi(-x'_{1,i} b_1) + y_{1,i} . (1-y_{2,i}) . \ln [\Phi(x'_{1,i} b_1) - \Phi(x'_{1,i} b_1) . \Phi(x'_{2,i} b_2)] \\ & + y_{1,i} . y_{2,i} \ln [\Phi(x'_{1,i} b_1) . \Phi(x'_{2,i} b_2)] + y_{2,i}(1-y_{3,i}) . \ln [\Phi(x'_{2,i} b_2) - \Phi(x'_{2,i} b_2) . \Phi(x'_{3,i} b_3)] \\ & + y_{2,i} . y_{3,i} . \ln [\Phi(x'_{2,i} b_2) - \Phi(x'_{3,i} b_3)] \} \end{aligned} \tag{8}$$

It is easy to numerically implement the sequential probit procedure when the error terms are uncorrelated. But ignoring the selection rules causes biases. The natural logarithm of maximum likelihood function with correlated error terms is as follows:

$$\begin{aligned} \ln L(b_1, b_2, b_3, r) = \sum_{i=1}^{n} \{ & (1-y_{1,i}) . \ln \Phi(-x'_{1,i} b_1) + y_{1,i} . (1-y_{2,i}) . \\ & \ln [\Phi(x'_{1,i} b_1) - \Phi_2(x'_{1,i} b_1, x'_2, b_2, r)] + y_{1,i} . y_{2,i} . \\ & \ln \Phi_2(x'_{1,i} b_1, x'_{2,i} b_2, r) + y_{2,i} . (1-y_{3,i}) . \\ & \ln [\Phi_3(x'_{2,i} b_2) - \Phi_3(x'_{2,i} b_2, x'_{3,i} b_3, r)] + y_{2,i} . y_{3,i} . \ln \Phi_3(x'_{2,i} b_2, x'_{3,i} b_3, r) \} \end{aligned} \tag{9}$$

## Endnotes

1. See Thompson (1998) for an early though very conservative estimate. Keeping in view the population differential between India (1 billion) and Pakistan (140 million) and the fact that India has approximately 60 million children between 9-14 years who never attended school, it would be not unreasonable to expect that Pakistan would more likely have approximately 10 million plus students who are totally illiterate. Add to this the number of children who are between 5-9 years old and the never-attended-school population increases substantially.

2. Pakistan Integrated Household Survey 2001.

3. Pakistan Integrated Household Survey 1999.

4. The linkage between human capital, income, fertility, labor force participation, number of children to have in a family and how many should be sent to school is surveyed in a theory of the family by Becker (1991).

[6] See Gao et al., (2001)

5. The evidence is mixed on this supply side issue. Alderman et al (1995; 1996) found that while the distance to primary school had a positive association with school beginners, the distance to middle school did not affect the enrollment decision. Sathar and Lloyd (1993) and Burney and Irfan (1991) found that having a public school less than one kilometer away was unrelated to primary school enrollment.

6. See Behrman and Wolfe (1987) and Schultz (1993) for opposing evidence from developing and developed countries. Ray (2000) found that in Pakistan the increased education of mothers led to reduction in child labor from which we can deduce that mother's education should be positively related to enrollment. But they also found that older girls drop out of school to join the workforce.

7. Within the South Asian context, Pal (2004) finds that household income and sibling composition are more relevant in explaining secondary school attainment than parents' education which is important in the enrollment decision at the primary school level.

8. The South Asian Association for Regional Cooperation (SAARC) countries are India, Pakistan, Bangladesh, Nepal, Sri Lanka, Maldives, Bhutan and Afghanistan.