Section 4. Empirical Test

Equation (6) above applies to each child i of age a. For testing purposes this equation simplifies to

$$P_{ia} = P_a (Y_{fi}, E_{fi}, L_{fi}, N_i) \quad (7)$$

The equation estimates the probability of school attendance by all children in each group, i.e., those who completed and those who did not complete previous levels. It therefore estimates the product of p^a and p^{a-1} corresponding to level j and $j-1$.

Y = family income in thousands of baht

R_r = regions dummy, $r = 1, 2 \ldots 5$

1 is North

2 is Northeast

3 is Central

4 is Bangkok metropolis

5 is South

C_L = urban/rural dummy, L = 0, 1, 2

0 is for village

1 is for metropolis

2 is for municipal

E_e = father's education, e = 0, 1, 2, ... 5, 6-7, 8-9

0 is zero schooling

1 is Pratum 1-4

2 is completed Pratum 4

3 is completed Pratum 7

4 is completed Maw Saw 3

5 is completed Maw Saw 5

6-7 is completed college and teacher education

8-9 is technical and other training

Rel_q = relation of youth to head dummy,* q = 1, 2.... 9

1 is spouse of head

2 is unmarried son/daughter

3 is married son/daughter

4 is son/daughter in-law

5 is nephew/niece

6 is parents of head

7 is relatives

8 is other dependent

9 is servant and employee

N is number of sibling of schooling age 6-24, a continuous variable

---

*Used in the analysis are relationships coded as 2, 5, 7 and 8.

We grouped the sample population by age ranges 7-10, 11-13, 14-18 and 19-24 which correspond to Thailand's schooling stages - Pratum 1-4 (primary level), Pratum 5-7 (upper elementary), Maw-Saw 1-5 (high school) and post secondary. The latter includes teacher and technical training and university education.

Both regression and logit methods were used in testing the hypothesis. Logit fits the model better as it directly gives an estimate of the probability of attendance in school. Note that the dependent variable is dichotomous, i.e., attendance or non-attendance in school assuming a value of 1 or zero and therefore has a non-normal distribution. This violates an important assumption of least squares estimate rendering interpretation of the estimated regression coefficients and the regression statistics unclear. (Please see the technical note on this in the appendix.) Nevertheless, regression analysis was applied for two reasons. First, we wanted to compare the results of the two methods to see whether the less expensive (computer timewise) regression estimates approximate closely enough the more suitable but expensive logit model. Some recent studies involving dichotomous dependent variables relied on regression estimates (Encarnación and Canlas, 1976; Canlas and Razak, 1979). Acceptance of their results would depend in part on our findings.

A second reason is expediency. Logit analysis takes more than ten times computer time for the equivalent regression estimate. We used regression to obtain the "best" independent variable selection in terms of significance and R^2s. The "best" selection was used for the logit analysis. We did not try to investigate the validity of this approach.

We tested the function by regression including all independent variables on each sample group and for a sub-set of independent variables after eliminating those that contributed insignificant value to the R^2. (Please see Appendix table for results of the complete regression runs.) Presented below are the results that we consider to provide the best estimate of the function. Included variables are family income, fathers' education, number of sibling, region and municipal/village location category. Though we cannot interpret the coefficients in terms of probability values we take their significance, sign and movement as rough estimates of the relationship. The regression results seem to be very satisfactory. Most of the regression parameters are significant and of the expected sign. Furthermore, they moved in the expected direction as we go from the youngest age group to the oldest.

First let us look at the constants. They declined from .786 to .262 as age (or level of schooling) increases from 7-11 to 19-24. This pattern reflects in part society's attitude to the different levels of schooling. The lower levels are regarded as a basic need required of every citizen for his orderly and satisfactory participation in all types of social interaction. Hence the large constant. As we move up the schooling ladder, the role of education becomes more specialized and there is no longer a common or equal desire for each level or type. Demand partly depends on expected net monetary benefits, partly on matching of perceived ability and required qualification. For these reasons the value of the constant tends to fall as schooling level increases. On the other hand, mainly because of financial constraints, the influence of cost-related factors tended to rise with level of schooling. Recall that schooling cost increases with level. Hence, the larger absolute value of the coefficient of family income, location, number of sibling and fathers' education. Fathers' education influences more strongly decision to pursue higher levels for two other reasons. One is its positive effect on home education which enhances inherent ability and therefore scholastic performance that is recognized in schools and in the selection of applicants for college or high school. Another influence is on education and occupational identity developed in children. Since lower education levels are regarded as basic for everybody, the influence of identity factor is likely to be weak, hence father's education.

The coefficients of the independent variables obtained by logit were practically all significant at the one percent level, similar to the results obtained by OLS. However, their values and movement as we go up the education ladder were very different. The sign and value of the intercepts were also very different. The influence of the various variables as reflected in the Beta coefficients increased with age range but up to 14-18 age range only. Then their values dropped. In contrast the movement of the regression coefficient was upward throughout.

The value of the probability of attendance in school for each age range can be estimated from the logit Beta coefficients for different values or categories of the independent variables. We limited our exercise to the extreme values of the variables to obtain the range of value of the probability for each age range. We take as the extreme values for number of youth to be one and five and for income B5,000 and B100,000. The probability, P, of the dependent variable, attending or not attending school with value of 1 or 0, taking on value of 1 is

$$P/_{Y=1} = \frac{1}{1 + e^{-X'B}} = \frac{e^{X'B}}{1 + e^{X'B}}$$

X and B are vectors of independent variables and the β coefficients respectively. We obtain the following values of the probability of attending school.

Table 9

Results of Linear Regressions of
Attending or not Attending School

	6-10	7-10	11-13	14-18	19-24
Constant	.7180	.7862	.8170	.5361	.262
Family income	.0004**	.0003**	.0005**	.0007**	-.0000
Region 2 (NE)	-.0581**	.0353	-.1069**	-.0154	-.0342**
Region 3 (Central)	-.0008	-.0016**	-.0128	.0606**	-.0268*
Region 4 (Bangkok)	.0184	.0302**	.0296**	.0987**	.0815**
Region 5 (South)	-.0693**	-.0381**	.0096	.0949**	.0090
Municipal	.0806**	.0390**	.0610**	.2050**	.1521**
Non-municipal	-.0397**	-.0339**	-.0951**	-.1721**	-.0355**
Number of youth	.0060**	.0031*	-.0069**	-.0143**	-.0021
FE_2 (Father's Education)	.0152*	.0276**	.0193**	.0310**	.0258**
FE_3	.0939**	.0749**	.1365**	.2260**	.1271**
FE_4	.1358**	.0994**	.1022**	.2740**	.2006**
FE_5	.1574**	.1270**	.0758**	.2743**	.2267**
FE_{6-7}	.1899**	.1205**	.1134**	.3108**	.3156**
FE_{8-9}	.1981**	.1181**	.1170**	.2417**	.2828**
$\bar{R}^2$	.055	.036	.098	.240	.132

Table 10
Range of Values of Probability of Attending School by Age
1975

	AGE RANGE			
	7-10	11-13	14-18	19-24
1. Lowest				
$X'\beta$	.912	.137	-1.787	-1.374
$P/_{Y=1}$	.7134	.5342	.1435	.2020
2. Highest				
$X'\beta$	4.030	4.807	3.358	.257
$P/_{Y=1}$	.9825	.9919	.9664	.5639

Table 11
Value of Beta in Logit

Independent Variable	AGE RANGE			
	7-10	11-13	14-18	19-24
Constant	1.193	1.462	- .472	-1.042
Family income	.009** (39.82)	.011** (43.35)	.006** (85.12)	- .0003 (0.144)
Region 2 (Northeast)	- .229** (8.70)	- .620** (54.69)	- .130* (4.47)	- .115** (3.01)
Region 3 (Central)	- .053 (0.37)	- .132 (2.01)	.251** (15.99)	- .084** (2.18)
Region 4 (Bangkok)	.294** (9.67)	.358** (11.64)	.434** (51.69)	.221** (6.19)
Region 5 (South)	- .283** (8.34)	.099 (0.74)	.445** (32.74)	.027 (0.554)
Municipal city	.253 (11.36)	.446** (28.88)	.867** (284.90)	.454** (12.99)
Non-municipal	- .137 (3.42)	- .405** (26.14)	- .800** (178.38)	- .185** (4.74)
Number of youth	.008 (0.30)	- .071** (20.39)	- .083** (66.49)	- .006 (1.01)
Father's Education 2	.200** (13.10)	.156** (6.77)	.151** (13.67)	.796** (3.29)
Father's Education 3	.603** (13.55)	1.696** (39.18)	1.149** (121.84)	.361** (6.27)
Father's Education 4	.967** (48.78)	1.175** (41.66)	1.542** (290.20)	.563** (13.41)
Father's Education 5	2.020** (11.84)	.848** (3.90)	1.747** (63.34)	.576** (6.16)
Father's Education 6-7	1.382** (19.31)	1.512** (16.85)	2.012** (110.94)	.660** (10.29)
Father's Education 8-9	1.380** (19.33)	1.602** (14.33)	1.335** (58.06)	.672** (8.

*.05 significance level.

**.01 significance level(chi-square values with one degree of freedom) are in parenthesis except for age 19-24 where asymptotic t-values are given.

To be noted from Table 10 are the rapid drop of the value of the probability of school attendance as schooling level increases and its wide range for the different socio-economic and location classes. For the possibly worst-off children the probability drops successively from .7134 to .5342, to .1435, to .2020 for ages 7-10, 11-13, 14-18 and 19-24. The corresponding figures for the possibly "best-off" children are .9825, .9919, .9664 and .5639.

Attendance rate at each age range can be predicted for different population groups and from this the distribution of formal schooling. Attendance rate is determined by the set of characteristics of each group. The model tells us further the importance of financial variables on schooling decision and provides us with an inter-generational link in the acquisition of education capital. An increase in a group's income or a change in location would have a permanent impact on all its future generations. The link is through future fathers' education and their subsequent income. Finally, the model explains why distribution tends to be more unequal as level of schooling rises. The policy implications seem to be obvious since financial factors are not difficult to change.