The Educationally Deprived in Peninsular Malaysia

Gautam Datta
and
Jacob Meerman*

I. Introduction

This paper is concerned with those households in Peninsular Malaysia whose children, although of primary school age, are not in school. The paper seeks to isolate variables associated with such educational deprivation. (Hereinafter educational deprivation refers only to the primary level of education.)

Peninsular Malaysia provides an interesting example of universal access to primary schools with very low average payments by households to the schools. Moreover for the age group of children under consideration (ages 7 to 12), the opportunity costs of sending children to school are usually low. Yet, in spite of these favorable factors, in 1974, about 18 percent of the households in the lowest income quintile remained educationally deprived — that is had one or more children of primary school age not in school. One important reason for this may lie in the substantial out-of-pocket costs for books, fees, uniforms, school lunches, transport and supplies. In the lowest quintile, for example, these costs accounted for nine percent of household income. Household characteristics other than income, also appeared to influence household decisions regarding primary school enrollment. The paper analyzed these

* The World Bank.

determinants of educational deprivation. The conclusions are of interest since they illustrate the difficulties of inducing certain groups to consume a basic good, even in a very favourable environment. The paper is also of interest in that it juxtaposes the usual multiple regression analyses with a logit analysis, which although less common is better adapted to deal with a binominal dependent variable.

The data used in the paper were generated in a sample survey of 1462 households throughout Peninsular Malaysia in 1974. The Malaysian Department of Statistics designed the survey as an equal probability sample for the Peninsula. The survey provided information on the household distribution of benefits from government expenditure. Consequently considerable data were collected on primary school attendance. Some of these are used in the present paper. Meerman (1979) contains a detailed description of the survey.

In Section II, the data on deprivation in Malaysia are discussed. Section III continues the analyais using both regression analysis and a logit function. Section IV discusses the results of the analysis in terms of several hypotheses explaining educational deprivation. The appendix discusses the advantages of using a logit or regression analysis to analyze a dichotomous variable.

II. The Data

The analysis involves only those households—53 percent of the total—which contained children of primary school-going age, that is between 7 and 12 years of age inclusively. As indicated in Table 1, in 12 percent of these households a child, or in a few cases, more than one child, were not attending school. (However, only 7.5 percent of school age children were not in school.)

Table 1 also orders the entire sample of households by quintiles of household per-capita income. The relation between income per capita and educational deprivation is evident. There is a negative monotonic relation between the proportion of deprived households and income quintile. The proportion of deprived households in the lowest income quintile (17.7 percent) is nearly three times that of the highest income quintile (5.9 percent).

Table 1

DISTRIBUTION OF HOUSEHOLDS WITH SCHOOL AGED CHILDREN BY INCOME QUINTILE, AVERAGE SIZE AND PERCENT DEPRIVED

Income Quintile[1]	Percent of Households	Average Size of Households	Percent of Households Deprived
1 (lowest)	16.6	7.5	17.7
2	19.9	7.6	15.4
3	20.6	7.4	13.6
4	23.3	7.4	8.8
5 (highest)	19.6	6.8	5.9
Mean/Total	100.0	7.4	12.0

1/ The income concept is the par capita income of each household and the quintiles were so defined that all had equal population in the total sample of 1462 households. This is an improvement over the usual household income concept, since it "corrects" for household size. The difference between the two measures is very substantial. For example, a third of the households in the lowest quintile of household income are in the highest four quintiles of household per capita income. For a full analysis see Datta and Meerman (1980).

The distribution by region, townsize, and race largely reinforces the pattern given by the distribution by per capita income alone. (See Table 2.) The poorest region, the North, has the highest amount of deprivation; the wealthiest region, Selangor, the lowest. Similarly educational deprivation is above the mean in rural areas, where mean income is low, and below it elsewhere.

As a group Indian children are far more deprived than Malay or Chinese even though Indians have a higher per capita income than Malays. However the Indians in the Peninsula are bimodal in their income distribution with some very wealthy urban residents and many poor rural families, particularly rubber tappers. Once this is taken into account the result is more understandable. The far above average incidence of illness among Indians may also explain part of the non-attendance of Indian children.

As would be expected, in three quarters of the deprived households the head of the household is in a poverty occupation. (See Table 2.) We also considered geographic isolation, defined as being more than four miles from an all weather road as a possible determinant of educational deprivation. There was no relation between isolation defined in this manner and educational depriva-

Table 2

EDUCATIONAL DEPRIVATION BY RACE, REGION, TOWNSIZE AND OCCUPATION*

Race	Percent of Households	Percent of Households with Deprived Children
Malay	58.8	11.5
Chinese	30.7	10.4
Indian	10.0	19.2
Other	.4	
Region		
Selangor[1]	16.1	9.5
North[2]	31.5	15.8
Other[3]	52.4	10.5
Townsize		
Metropolitan	14.7	8.7
Large urban	14.2	8.1
Small urban	13.2	7.7
Rural	57.9	14.8
Occupation		
Non-poverty	43	5.9
Poverty[4]	57	16.6
Mean		12.0

* includes only households with school aged children.

1 Wealthiest state, it includes Kuala Lumpur, the capital.

2 Trengganu, Kelantan, Kedah, and Perlis

3 Pahang, Johore, Malacca, Negri Sembilan, Perak, and Penang

4 rubber tapping, landless farm labor, farming, fishing

tion. Using a three mile limit gave the same result. It may be that distance of the home from the primary school determines attendance. There was, however, no measure of such distance in the data file.

Table 3 shows that the probability of a *household* with an educationally deprived child increases monotonically with the number of school aged children. However this table should not be taken to mean that the probability of a *child* being deprived rises with the number of children. In fact, 7 percent of children are

Table 3

HOUSEHOLDS BY SCHOOL AGED CHILDREN AND EDUCATIONAL DEPRIVATION

No. of School-Aged Children	Percent of Households	Percent of Children	Percent of Households with Deprived Children
1	43	24	7
2	35	39	12
3	17	28	21
4	4	9	29
5 or more	*	1	**
Total/Mean	100	100	12

* Less than 1%.
** Partition too small to estimate.

deprived both in one child families and in families with three or four school aged children. For families with two school aged children the deprivation ratio falls to 6 percent.[1] Families with more than one deprived child are rare. They constitute only 15 percent of all deprived households, while 76 percent of all children are in families with two or more children.

The number of children per household varies somewhat across the partitions of income, race, region, townsize and occupation of head of household. The number of children of primary school age per household is significantly smaller (less than 0.8) in metropolitan towns as contrasted to all other townsizes (where the mean number exceeds unity). Consequently one would expect metropolitan towns to have a much lower percentage of deprived households relative to other urban townsizes. However this was not the case. The result is somewhat surprising since per capita household income is by far the highest in metropolitan towns for all races. However, the sample is too small to draw any strong conclusions.[2] No other result is significantly affected by the variation in

1 The hypothesis that there is a relation between the likelihood of being deprived and the total number of children (including secondary school going children) in a household was not tested because of lack of data. An interesting contribution to anlaysis along these lies is Birdsall in Simon and Da Vanzo (ed), (1980).

2 There were 114 households in the sample who lived in metropolitan areas. Of these 10 households were educationally deprived.

the number of children across partitions.

III. Econometric Analysis

Table 4 gives the simple correlation coefficients between deprivation and the independent variables. The sample consists of the 782 households with school-age children. "Femness" is defined as the ratio of female children to total children.

Table 4

CORRELATION COEFFICIENTS BETWEEN EDUCATION DEPRIVATION AND INDEPENDENT VARIABLES

	Coefficient
(1) Children per household between 7 and 12	0.21
(2) "Femness"	0.01
(3) Dummy for poverty occupations	0.17
(4) Dummy for Malay	-0.02
(5) Dummy for Chinese	-0.03
(6) Dummy for Indian	0.07
(7) Log of income	-0.12
(8) Dummy fo Metropolitan	-0.04
(9) Dummy for Large Urban	-0.05
(10) Dummy for Small Urban	-0.05

Deprivation is clearly related to the number of children, being in a poverty occupation and, inversely, to the log of income. There is also some relationship between deprivation and being Indian. The R^2 for the multiple regression is 0.08 using only number of children and poverty occupation as the independent variables. Although several of the other variables are significant their addition increases the R^2 only marginally. "Femness" was not a significant variable even with the exclusion of the largely urban and better off Chinese, who presumably do not discriminate against girls. Nor was femness a significant variable when the regressions were re-run using only Malay households.

Consequently in the multivariate analysis, femness has been dropped. The variables included are as follows:

DEPRIVED	The dependent variable, taking the vlaue 1 if a household has one or more children of school-going age not in school and the value 0 in other cases.
INCPERC	Income per capita.
NORTH	Dummy variable for the northern region (Trengganu, Kalantan, Kedah and Perlis).
SELANGOR	Dummy variable for Selangor.
METROPOLITAN	Dummy variable for Metropolitan areas (population greater than 75,000).
URBAN LARGE	Dummy variable for large urban areas (population between 10,000 and 75,000).
URBAN SMALL	Dummy variable for small urban areas (population between 1,000 and 10,000).
POVOCC	Dummy variable for the head of the household being in a poverty occupation (rubber tapping, landless farm labour, farming or fishing).
CHINESE	Dummy variable for Chinese household.
INDIAN	Dummy variable for Indian household.
OTHER	Dummy variable for the race of the household being other than Malay, Chinese or Indian.
CHILDREN	The number of children between 7 and 12 years of age in the household.

Both the regression and logit models were estimated with household income per capita and its natural log as explanatory variables. The results were very similar. Since the use of income per capita rather than its logarithm results in a more intuitive interpretation in the case of probability change, the variable was left untransformed in the presentation. There are, therefore, six underlying variables or blocks of variables: income; region (with the North and Selangor shown separately against the omitted group comprising the states of Pahang, Johore, Malacca, Negri Sembilan, Perak and Penang); size of town of residence for three sizes with the rural area being the omitted group; occupation of head of household with the omitted groups being professionals, skilled workers, students, businessmen and housewives; race with Malays as the omitted group; and the number of children of primary school-going age in the household.

Table 5 below presents the regression coefficients as well as the transformed logit coefficients. The transformation consists of multiplying the estimated logit coefficients by P(1-P), where P = probability of participation or choice.[3] This adjustment is required to make the logit and the regression coefficients comparable.[4]

Table 5

MULTIPLE REGRESSION AND LOGIT RESULTS

Explanatory Variable	Regression Coefficient	Transformed Logit Coefficient
1. INCPERC	0.00008	-0.00095*
2. NORTH	0.0796*	0.0896*
3. SELANGOR	-0.0189	-0.0229
4. METROPOLITAN	0.0095	0.0364
5. URBAN LARGE	-0.0609*	-0.0658
6. URBAN SMALL	-0.0555*	-0.0681
7. POVOCC	0.0890*	0.0923*
8. CHINESE	0.0500*	0.0693*
9. INDIAN	0.1103*	0.0977*
10. OTHER	0.1147	0.0718
11. CHILDREN	0.0794*	0.0735*
12. CONSTANT	-0.1048	-0.4241

R^2 = 0.09; Likelihood Ratio Index (for Logit Analysis) = 0.54[1]

* indicates significance at the 0.05 level for the "F" test in the case of the regression coeffcient and the "t" test in the case of the Logit coefficient.

1 This implies that through use of the logit coefficients, the prediction of whether or not a household is educationally deprived will be correct in 88.5% of the time.

The regression and Logit results are similar. The signs of the coefficients are identical except for the variable INCPERC, and their respective values are very close. But the significance tests

3 The choice in this case is the option of not sending children to school.

4 This procedure is explained in the Appendix.

produce some differences. The variables found significant according to both methods are NORTH, POVOCC, CHINESE, INDIAN and CHILDREN. In addition URBAN LARGE and URBAN SMALL are significant according to regression analysis, while INCPERC is significant in the Logit estimate.[5]

The probability interpretation of the results presented in Table 5 are interesting. The transformed Logit coefficients as well as the Multiple regression coefficients measure the increase/decrease in the probability of deprivation for a change in the value of the independent variable. Some of these shifts are quite large. According to the Logit model, being an Indian household rather than Malay raises the probability of being deprived by almost 10 percentage points. In the Logit model residence in the North or being in a poverty occupation household raises the probability by about 9 percentage points, while an increase in monthly household per capita income of $100, which is 107 percent of overall mean household per capita income, reduces the chances of being deprived by approximately 10 percentage points.[6]

IV. Explanatory Hypotheses

It is illuminating to discuss these results in terms of common causal hypotheses of educational deprivation.

Supply: In many countries children are not in school because school-places are unavilable. Or if available, they are so distant as to be inaccessable. Such short supply does not appear to be important in the Peninsula. We have discussed the lack of correlation for school attendance and geographical isolation. The survey questionnaire also included an open-ended question asking both households and community leaders which government services they desired in larger amounts. Less than five percent of the households and none of the community leaders mentioned more schools. Moreover in 1972[7] Malaysia's per capita expenditure for

5 The log of INCPERC is significant in an alternative regression exercise as well as the Logit model. Metropolitan townsize is not significant in either model, confirming earlier analysis.

6 In the case of the OLS regression estimates, these figures are 11, 8, 9, and 8 percentage points respectively.

7 The year closest to the sample data.

primary education ($US42) was, after Singapore, higher than all other developing countries.[8] The total effect of these facts strongly indicates that inadequate supply of school facilities does not explain educational deprivation in Peninsular Malaysia.

Income: Education deprivation can also arise from the demand side. Low income households may not send children to schools, even if they are accessible and have low or zero fees, as is the case in Peninsular Malaysia. This may result from the high opportunity cost of keeping children in school or because of the out-of-pocket expenses of attending school for textbooks, uniforms, school lunches, transport and supplies. The former is unlikely to be important because primary school going children have very little earning ability. But the latter is important: in response to the survey questionnaire, 22 percent of the households in the two lowest income quintiles gave educational assistance as the most desired additional public service. The inadequate income hypothesis is also supported by the Logit analysis in which per capita income remains a significant variable with negative sign.

Sex Discrimination: A frequent cause of education deprivation is discrimination against girls. In many Muslim countries, for example, the male enrollment ratio is substantially higher than that of the females. In many traditional societies such behavior may be based on the expected returns to the household, which may be higher for educating a male child rather than a female. In Malaysia such discrimination appears to be an unimportant, although not completely negligible, factor. In 1978, the female enrollment rate in primary school was 97 percent of the male enrollment rate.[9] And as noted in the analysis of Section III, "femness," the ratio of female to total children, was not a significant variable.

Low Valuation: Notwithstanding the universal and strong desire for education, households vary in the degree to which they value it. Some are willing to make far greater sacrifices to educate their children than others. Perhaps on average those with less education will desire education for their children less strongly than those with more education. In some traditional agricultural communities, other things being equal, there may be less educational

8 UNESCO (1980).
9 World Bank (1981), Table 23.

aspiration than say in urban communities where a modern life-style predominates. (In the former some parents may be suspicious of education as a force destroying traditional values and beliefs. Tradition, and low levels of parental education are both strongly associated with poverty occupation of the head of household and residence in the conservative and rural North (Kelantan, Kedah, Perlis, and Trengganu). Since both the Logit and the regression analyses account for income as a separate variable, the high significance of the North and poverty occupation—in which low levels of education predominate—in explaining educational deprivation may be due to a somewhat lower valuation of education by these groups. The stress here needs to be on the "relative." Most children in these two groups are in school. And in general the financial sacrifice by those in the North and in poverty occupations to keep their children in school indicates very high valuation of education.

Appendix

Educational deprivation — as defined here — is dichotomous: a household either has all of its children in the relevant age group in school or it does not, and is therefore educationally deprived. When the dependent variable is specified in this fashion rather than as a continuous variable, the familiar linear regression analysis can only be applied in a qualified fashion. Since this problem is not very common in economics, it merits some discussion.

A dichotomous dependent variable can be interpreted as a continuous vairable, and the regression equation can be interpreted as a linear probability function. The dependent variable, therefore, becomes the probability of an outcome (in our case education deprivation), given specific values for the explanatory variables. The coefficients of the independent variables of the regression equation can be interpreted as contributions to that probability.

However the regression approach suffers from two principal defects. The first is the failure of the probability function to be confined to the unit interval, although in the real world probabilities must lie between zero and one. A second difficulty is that the computed R^2 is invalid and likely to seriously underestimate the goodness of fit of the data. For example in case of a single independent variable, the scatter of data appears as two horizontal bands, one at zero and the other at unity. A straight line is simply not the appropriate function with which to approximate such a scatter, especially at the end points, and consequently the R^2 is liable to be very low.[10]

A simple alternative probability function is the logistic function, where P(X) the probability of choice is a function of X equal to $1/_{1+e}-(\alpha_0 + \alpha, v)$ in which α_0 is a constant and α_1 is the Logit coefficient.[11] This function is S in form and asymptotically approaches 0 and 1, thus lending itself easily to a probabilty interpretation. The Logit function does fit the data well, although the R^2 can no longer be used, since it does not have the usual interpretation because the Logit function is not linear in the parameters. The likelihood ratio index however is an appropriate measure of goodness of fit.[12]

The relation between the Logit coefficients and the change in probability is easily established. The Logit function, with a single independent variable x. has the following form:

$$\rho(x) = \frac{1}{1+e} - (\alpha_0 + \alpha, x)$$

$$\text{Then } \frac{\partial \rho}{\partial x} = \frac{\alpha_1, e^{-(\alpha_0 + \alpha, x)}}{1 + e^{-(\alpha_0 + \alpha, x)}}^2$$

$$\text{Since } \rho(x) = \frac{1}{1 + e^{-(\alpha_0 + \alpha, x)}}$$

$$1 - \rho(x) = \frac{e^{-(\alpha_0 + \alpha, x)}}{1 + e^{-(\alpha_0 + \alpha, x)}}$$

$$\text{Then } \rho(1 - \rho)\alpha_1 = \frac{\alpha_1, e^{-(\alpha_0 + \alpha, x)}}{1 + e^{-(\alpha_0 + \alpha, x)}}^2$$

10 An alternative is to approximate the scatter by a polynomianl of a sufficiently high degree. However the ready availablility of computer packages which can estimate non linear functions such as the Logit and Probit Functions, removes the need for such labored techniques.

11 Obviously this can be extended to include additional variables.

12 See Berkman et al. for an explanation.

which is seen to be identical to $\frac{\partial \rho}{\partial x}$

Therefore, the Logit coefficient, α_1, multiplied by P(I-P), where P stands for the probability of a household being deprived can be interpreted as a change in probability and is analogous to the coefficients of the linear probability function. In Table 5, for example the transformed Logit coefficient for the variable number of children is .0735. Hence each additional child increases the probability of a household having educational deprivation 7.35 percent and $P(I\text{-}P)\alpha_1 = .0735$. Hence the Logit coefficient $\alpha_1 =$ (.0735)/P(I-P).

References

Berkman, J., Brownstone, D., Duncan, G., and McFadden, D., *Quail 30 users' Manual,* September 1977 (Mimeo, Department of Economics, University of California, Berkeley).

Birdsall, Nancy, in Simon, Julian L. and DaVanzo, Julie (ed.) *Readings in Population Economics,* Vol. 2, pp. 115-150.

Datta, Gautam and Meerman, Jacob, "Household Income or Household per Capita Income in Welfare Comparisons," *Review of Income and Wealth,* Series 26, Number 4, December 1980.

Gunderson, Morley, "Probit and Logit Estimates of Labor Force Participation," *Industrial Relations,* Vol. 10, No. 2, Spring 1980.

Meerman, Jacob, *Public Expenditure in Malaysia,* Oxford, 1979.

UNESCO, *Statistical Yearbook,* 1980.

World Bank, *World Development Report,* World Development Indicators, 1981.