Investment in Post-Compulsory Education in Sri Lanka

Athula Ranasinghe
Tinbergen Institute
Amsterdam
The Netherlands

Joop Hartog
Faculty of Economics and Econometrics
University of Amsterdam
The Netherlands

Abstract

In this paper, we have used the standard *Human Capital* model to describe the *post-compulsory* schooling behaviour of *Sri Lankans*. We assumed that there is *no uncertainty in the education system or in the labour market*. Therefore, in the *steady-state*, the earnings profile of one generation is a replica of the earnings of the next generation.Then, we modeled and estimated the *school enrolment* and *the length of schooling* decisions.Our results show a very clear positive association between the family background and the education decision. Children of more affluent families seem to derive more benefits from the free education policy. In particular, mother's education has a very strong effect on the education of the child. This casts doubt on the effectiveness of the free education policy as a poverty alleviation instrument and its role in social mobility.

We also observe that the ability effect on the return to education is greater than the ability effect on the cost of education. Therefore, more able children stay longer in full-time education than the less able children. We further found some negative evidence on the *Behrman and Taubman* observation of the negative *birth order effect* on ability. We observe that younger siblings are more able than older siblings of the same family.

By decomposing the total variance of the schooling length, we observed that the rate of return variation is more important in explaining the schooling behaviour than the variation in the cost of education .

Analysis of the residuals has given us an impression that there is a family fixed effect which is not explained by the model. Most probably, the poor specification of the ability and the excluded school quality variables would be the main reason for this unexplained fixed effect.

1. Introduction

There are at least two economic reasons favouring the provision of subsidised education. One can distinguish them as the arguments related with *equity* and those related with *efficiency*. Simply, the former argues that education enhances the earnings capacity of the recipients. In the absence of the subsidised education, only the children from rich families can be educated and therefore, *the vicious circle of inter-generational poverty transmission* will be continued. Children born into poor families will be poor and *vise-versa*. The free education policy is expected to break that vicious circle. The second argument is related with the efficiency. Due to the positive externalities associated with education, individual decisions will not be socially optimal. One can further argue that in the presence of capital market constraints children from poor families cannot finance their optimal length of education. Then, the subsidised education is essential for the optimal resource allocation in education[1].

[1] These are cited in almost all the related literature. see, *inter-alia*, Polachek and Siebert (1993):pp. 59 - 67.

Authors wish to thank Hessel Oosterbeek, Mirjam van Praag of the University of Amsterdam, Edwin van Gameren and Paul Frijters of the Tinbergen Institute of Amsterdam and Gamini de Silva of the Department of Census and Statistics of Sri Lanka for their valuable comments, suggestions and moral support in this endeavour.

Almost all the developed countries and also many developing countries have one or another type of subsidised education system. Among the developing countries, Sri Lanka is an interesting case where the entire education system is free at all the education levels up to the post-graduate level. Every Sri Lankan can enjoy free education from the primary to the university level. A literacy rate of over *80* percent and the school enrolment rate of over *70* percent, which have been cited repeatedly, are some of the direct outcomes of this policy. However, the income of many Sri Lankan families is still below the poverty line[2].

In contrast to the standard theories of human capital, this evidences that something has gone wrong in Sri Lanka. In theory, we assume that education enhances the earnings capacity of the recipients. Then, in the presence of the free education and provided that people are rational, a poor child can acquire more education and can easily get away from poverty. However, this is conditional on two factors. *First*, the labour market must be favourable for the educated people both in terms of the employability and the payments. *Second*, one's optimal education decision should not be positively related with his or her family background.

As a number of studies show, the rate of return to education is quite high in Sri Lanka, particularly in the higher education. For example, *Kelly's (1993)* provisional estimates show that the marginal rate of return to the primary education is nearly *6* percent and that increases to nearly *18* percent in the *G.C.E. (O.L)* (grade *11* completers). The marginal rate of return to the university education is also around *18* percent. These statistics show that the rate of return to education is high and increasing with the schooling level. The employability of the trainees is also crucially important. Otherwise, free education merely equips people with useless skills. Some recent evidence shows that the relationship is *inverted-U* shaped ; with lowest unemployment rates for uneducated and highly educated people. For example, *The Quarterly Labour Force Survey -1995* shows that the unemployment rate of the category with no school education is 3.1 percent and that increases to 4.9 percent for the category with reported education between 1 year to 5 years. It further raises to 21.3 percent for the category with G.C.E.(O.L) and that of the category with G.C.E.(A.L.) certificate was 28.3 percent. The unemployment rate of the category with a degree or post-graduate qualification was 3.2 percent. This shows a very clear parabolic association between the unemployment rate and the level of education[3].

---

[2] For example, according to the World Bank statistics, 22 percent of Sri Lankans earn less than *the upper poverty line*. See, World Bank Annual Report- 1995

[3] In part, this inverted-U shaped relationship may reveal the economic crisis taking place in Sri Lanka. In the recent history, the economy was able to create new jobs only in defence sector (military) and in garment industries. Most of these jobs were for less educated people. Due to the political pressure, governments also were having special

Further, there is a third possibility for continuing the vicious circle of the inter-generational poverty transmission. If people themselves are selective in the schooling decision, the free education policy will not be effective simply because there are some people who do not want to be in school due to reasons other than the direct cost of education. Therefore, understanding of the individual level motivations in schooling decisions is vital both to the policy makers as well as to the education economists.

From the policy makers' point of view, for example, if the children with affluent family background are more likely to stay longer in school than the others, the free education policy will be counter-productive. Then, the free education policy strengthens the vicious circle rather than breaking it. For an economist, this is related with the appropriateness of the *Mincerian type rate of return estimate* which assumes that education is exogenous. If the education is an outcome of the maximisation behaviour of individuals, the *OLS* estimates of the rate of return will be biased *[Wills and Rosen (1979)]*.

In this paper, we re-state the *Beckerian* theory of investment in human capital *[Becker (1967-Woytinsky Lecture)]* in order to accommodate our empirical analysis within the standard theoretical framework. The theory assumes that the child in the decision making age knows for sure his or her ability to complete the stipulated length of education and the labour market is in a *steady state*. Therefore, the earnings profiles of the older generation are a replica of the earnings potentials of the younger generation. Further, the decision making child will stick to the decision made at the end of the post-compulsory stage because it is the optimal decision at any point in the life span.

Uncertainty in both the schooling system and in the labour market and the dynamic behaviour of the rational individual (*i.e.*; optimise the expected life-time earnings at the end of each period and change the optimal plan if necessary) *[Levhari and Weiss (1974), Altonji (1993)]* are beyond the focus of the present paper.

2. Theoretical Foundation

We assume that at the end of the compulsory education *(age 10 in Sri Lanka)* a child faces a choice problem. He(she) has to decide on the optimal length of post-compulsory education *(d)*. In order to determine it, he(she) maximises the present value of the life-time income to the length of post-compulsory education, *d*.

schemes for graduate employment.

Thus, he(she) solves the following optimisation problem[4].

$$\underset{d}{Max}\ V_d = -\int_{x}^{x+d} K_d\, e^{-rt} dt + \int_{x+d}^{x+d+T} W_d\, e^{-rt}\, dt \tag{1}$$

where V_d is the present value of the net lifetime income due to d years of post-compulsory education. x is the age at which the compulsory education is completed, or the age at which the post-compulsory education decision is made (in our particular case $x=10$). Then we can also write $x+d$ as the age at which the child expects to leave school. K_d is the annual direct cost of education which is assumed to be a constant for each schooling year. r is the personal discount rate that is used to discount the net income flows from a given length of education. T is the length of working life. Or, $x+d+T$ is the retirement age. W_d is the wage rate for d length of education, and e is the natural logarithmic base.

Now, assuming that T is very large and $\partial K_d / \partial d = 0$, we can write the first order condition as follows:

$$r(K_d + W_d) = \frac{\partial W_d}{\partial d} \tag{2}$$

where the left hand side is the *marginal cost* and the right hand side is the *marginal rate of return* (benefits) to education.

Now, following *Card (1994)* we assume that for a given individual both the marginal cost and the marginal return to education are linear functions of the level of education[5]. The marginal cost increases and the marginal return decreases with the education. We further assume that the intercepts of the above two relationships vary over individuals, the effect of the level of education on both is constant. Then, we can write the following two equations to represent them.

$$r(K_d + W_d) = c_{o_i} + c_1 d_i \tag{3}$$

$$\frac{\partial W_d}{\partial d} = b_{o_i} - b_1 d_i \tag{4}$$

a linear equation for the joint effect of discount rate and intercept of the earnings function, and a linear equation

4 This is the standard human capital model as developed by Becker and Mincer. There are many applications of this approach in the literature. See, inter-alia, Hartog (1993), Card (1994) and Oosterbeek and Van Ophem (1995).

5 In fact, this is a simplified version of the implied equilibrium condition. According to the theoretical notion that we have developed so far, the equilibrium is essentially a quadratic equation. This makes the model more complicated. Card (1994) assumed that the direct cost is zero and that the left hand side of the equilibrium condition represents only the discount rate. Card (1994) has ignored the opportunity costs component.

An alternative justification for this simplified version is to assume that the Wd component on the left hand side of the equilibrium condition represents only the intercept of the earnings function, i.e. W_0.

for the slope of the earnings function, where the subscript *i* indicates that the intercept varies over individuals. Now, we can solve this equation system to obtain the equilibrium length of education for the representative individual *i*. This yields,

$$\bar{d}_i = \frac{b_{o_i} - c_{o_i}}{b_1 + c_1} \tag{5}$$

Now, we assume that the individual specific component of the marginal return to education, b_{0i} is a function of ability (*ABL)*, gender *(G)* [G=1 if Male] , the place of residence (U) [U=1 if Urban] and a random component; ϵ_1 . The individual specific component of the marginal cost curve, c_{0i} is a function of ability (*ABL*), parents' income (*I*), gender (*G*) , the size of the household (*N*) and a random component (ϵ_2) . Both are assumed to be linear functions.

$$b_{o_i} = \beta_o + \beta_1 \,(ABL)_i + \beta_2 \,U_i + \beta_3 \,G_i + \epsilon_{1_i} \tag{6}$$

$$c_{o_i} = \gamma_o + \gamma_1 \,(ABL)_i + \gamma_2 \,I_i + \gamma_3 \,N_i + \gamma_4 \,G_i + \epsilon_{2_i} \tag{7}$$

We expect that children with greater ability have a higher rate of return to education : $\beta_1 > 0$ and that they also have a higher opportunity cost because they can earn higher wages at all education levels, ($\gamma_1 > 0$). The clear implication of this is that the ability effect on the schooling decision is unpredictable. Whether more able children stay longer in school depends on the relative effect of the ability on the slope (*return to education*) and the intercept of the earnings function (*cost of education*). If the ability effect on the slope is greater than that on the intercept, more able children will enjoy more education. In other words, *the ability effect on schooling is positive if the earnings functions with different ability levels diverge*[6].

We assume that both the intercept of the earnings function and the marginal rate of return to education for men are not smaller than that for women with the same characteristics. [$\gamma_4 \geq 0$], [$\beta_3 \geq 0$]. his assumption is consistent with the common belief that women are discriminated in the labour market.

Parents' income is assumed to have a negative effect on the cost of education. Children born into rich families can afford the education cost from their own funds for which they can apply a lower discount rate

---

[6] See, Hartog (1993) and Oosterbeek and Van Ophem (1995)

($\gamma_2 \leq 0$).

Children belonging to large families have to share the given family resources and therefore more likely have to rely on outside funds for which the rate of interest is higher than the rate of interest to own funds. Furthermore, because of the same reason those children would have to find jobs to compensate for the insufficient family income and therefore they have higher opportunity cost. All these arguments imply that the family size *(N)* has a positive effect on the cost of education ($\gamma_3 > 0$).

Because of the unequal economic development between the urban and the rural sectors in many developing countries, most of the employment opportunities are generated within the urban sector. Therefore, we expect that the rate of return to education is higher in the urban sector than that in the rural sector, ($\beta_2 > 0$)[7]. Another reason to expect a positive sign for this coefficient is that the quality of education varies among sectors. Most of the highly facilitated schools in Sri Lanka are located within the urban sector.

The ability variable in this model is assumed to cover both *the cognitive and the affective abilities*. Here, we assume that ability can be divided into four additive components; *genetic ability*, ability accrued from *education*, ability accrued from the *socialisation process at home* and an *unmeasurable component* of the ability. With regard to the post- compulsory education decision, ability accrued from compulsory schooling is assumed a constant for everyone. Everyone makes the decision at the same age and with the same level of education. Then, we assume that the *genetic ability* and the ability accrued from the *socialisation process* at home is a function of the *birth order* of the child and the *education background of the family which is assumed to be represented by the mother's education*. In the economic literature, *birth order* is used first by *Behrman and Taubman (1986)*, as an indicator of the *in-born ability*. As they have presumed, higher order births (*younger children*) have disadvantages both in genes and parental care. The *Behrman and Taubman (1986) hypothesis* claims that *ceteris-paribus*, younger members of a family have lower genetic endowment than the older siblings of the same family and hence are less able. However, as we have already discussed, when the ambiguity of the ability effect on the schooling decision is concerned, the *birth order effect* on schooling is not predictable. In this model, we have used an indirect method to test the *Behrman and Taubman (1986) hypothesis* on the effect of birth order

[7] This may be partially eliminated by the internal migration of the workers. In fact, the willingness of most Sri Lankans to live in the urban sector may be explained by the higher rate of return. In the presence of practical difficulties such as the financial and psychological costs of migration and high concentration of public and private investments in the urban sector, the internal movements of workers is not sufficient to equalise the rate of return over regions.

on ability. Children of educated parents receive more attention from the parents and these parents are more aware of the development of the child both physically and psychologically. In this model we use the education level of the mother as an indicator of the family educational background[8].

Then, we can write the following equation to represent the ability.

$$ABL_i = \mu_0 + \mu_1 M_i + \mu_2 B_i + \epsilon_{3_i} \tag{8}$$

where μ_o is the constant representing the ability accrued from compulsory school. M is the mother's education which is assumed to represent the education background of the family. μ_1 is the marginal effect of the education background of the family as measured by mother's education. B is the *birth order* of the respondent. This is a serial number assigned to the siblings in each family arranged in descending order of their age (i.e. the oldest child gets birth order 1) and μ_2 is its marginal effect on the ability. Finally ϵ_3 is a random error component which represents the un-observable ability variations. All coefficients are assumed to be positive .

Substituting the equations *(6)* , *(7)* and *(8)* into *(5)* we can write the reduced form of the equilibrium length of education as follows,

$$\bar{d}_i = \theta_o + \theta_1 M_i + \theta_2 B_i + \theta_3 U_i + \theta_4 G_i + \theta_5 I_i + \theta_6 N_i + \epsilon_i \tag{9}$$

where, $\theta_o = \dfrac{\beta_o - \gamma_o + \mu_o(\beta_1 - \gamma_1)}{b_1 + c_1}$, $\theta_1 = \dfrac{\mu_1(\beta_1 - \gamma_1)}{b_1 + c_1}$, $\theta_2 = \dfrac{\mu_2(\beta_1 - \gamma_1)}{b_1 + c_1}$

$\theta_3 = \dfrac{\beta_2}{b_1 + c_1}$, $\theta_4 = \dfrac{\beta_3 - \gamma_4}{b_1 + c_1}$, $\theta_5 = -\dfrac{\gamma_2}{b_1 + c_1}$ $\theta_6 = -\dfrac{\gamma_3}{b_1 + c_1}$ and

. $\epsilon_i = \dfrac{\epsilon_1 - \epsilon_2}{b_1 + c_1} + \dfrac{\beta_1 - \gamma_1}{b_1 + c_1}\epsilon_3$

The expected values of ϵ_1 , ϵ_2 and ϵ_3 are zeros, any pair of error terms is not correlated and the variance of each error component is a constant by assumption[9].

---

8 The reason for choosing the mother's education rather than that of the father or the average education level of the family is conventional. As the mother has a dominant role in child rearing, we assume that mother's education is more important than the alternative measures.

9 In fact, these are rather unrealistic assumptions. There are many possibilities which jointly determine the error terms of these equations. For example, the unmeasurable component of the ability variable contributes to the error term in equation 8. At the same time the unmeasured part of the ability would have some effect on both the intercept and the slope of the earnings function. However, we neglect such possibilities in order to avoid further complication of the analysis.

None of the structural parameter is identified in this model. However, since b_1+c_1 is a constant and positive by assumption, we can identify some of the structural parameters at least up to a scalar of (b_1+c_1).

Now, we can use this specification to test whether the ability effect on the rate of return is greater (or less) than the ability effect on the opportunity cost. [i.e. $\beta_1 \geq \gamma_1$][10]. However, it is worth noting that this is a weak test in the sense that the ultimate conclusion depends on the assumptions concerning the sign of the structural parameters of the ability equation which are not empirically verifiable. We have two reduced form coefficients in the model facilitating this test, θ_1 and θ_2 . However, the sign of μ_2 in θ_2 is not predictable even at the theoretical level. It is more acceptable to assume that the mother's education has a positive effect on the ability of the child ($\mu_1 \geq 0$). Therefore, we choose θ_1 for this test. Then, the result of that test and estimated θ_2 can be used to test the so-called *Behrman and Taubman (1986) hypothesis* that younger children of a family have disadvantages in terms of genetic endowment and parental care. If that hypothesis is true, μ_2 in our specification must be negative. Since the sign of b_1+c_1 is assumed to be positive and the sign of $\beta_1 - \gamma_1$ is revealed by the test on θ_1, the sign of μ_2 can be detected from θ_2. Apart from that we can also have a rough idea about the relative importance of changes in the rate of return and the changes in marginal cost as determinants of the optimal length of education.

However, it should be kept in mind that all our results are conditional on our assumptions concerning some of the unidentified structural parameters. Despite this limitation, this model gives us very important information concerning the policy related questions such as the question that we posed at the very beginning of this paper: *does free education indiscriminately help everybody to acquire marketable skills in terms of years of schooling ?*

---

[10] This is the sufficient and necessary condition for the positive ability effect on education. See, Hartog (1993) and Oosterbeek and Van Ophem (1995)

3. Education System in Sri Lanka

In Sri Lanka, school going starts at the age of five. From then on, children undergo compulsory education for five years. At grade five (*age 10*) a child (*or, parents*) can decide whether to continue her education or to leave school. However, for many, this is not considered to be a terminal step. Main focus in these five years of compulsory education is on *literacy and numeracy*. If a child decides to continue his(her) education further, he(she) will face the first terminal at grade eleven, (*at age 16*). At grade eleven, the child is required to get through the first selection criterion. He(she) sits for a government examination, *General Certificate of Education, Ordinary Level (O.L)*, in which the child is tested on eight subjects including *literacy, numeracy, science, religion and cultural affairs etc*. However, so far no specialisation is available . Only the students with high records in this exam are entitled to continue further education. A child can sit a maximum of three times this exam and unless successful in one of these trials, he(she) has to leave the school. Among those who succeed, only the pupils with highest scores are eligible to continue education. They are required to study for another two years. At this level students have to choose one of the specialisations depending on their *O.L* results. There are four such specialisations: *Arts, Commerce, Biological science and Physical science*. After two years, they sit for another government examination, *G.C.E. Advanced Level, (A.L)*. This is the last exam in the school system and it is also the university qualifying examination.

Full time education beyond the secondary level is rather complicated. A child can attend an alternative education institute at different levels. For example, there are technical and poly-technical programmes for the children with different education qualifications. Depending on the programme offered the entry requirements vary from grade eight to G.C.E (A.L).

The university education which is considered to be the best alternative is limited only for the Advance Level qualified students with the highest overall scores. There are three types of university degrees in Sri Lanka; *special degrees* with a four year duration (*eg. special degree in Economics)* , *general degrees* with a three years duration (an example is the general degree with *Economics, Sociology* and *Political Science*) and *medical degrees* take five years.

Those who are qualified at the A.L but have not sufficient marks to enter universities have several alternatives such as the law college, the open university and the external degree programmes and many other professional exams conducted by the private sector are few of them.

In this paper, we will only analyse the school enrolment decision up to the G.C.E.(A.L). Decisions concerning post-schooling like university is left out. One of the main reasons for this is that we had only few observations in our sample belonging to this category in the specified age group.

Another important feature of the education system in Sri Lanka is the quality disparity in school facilities. The quality disparity between urban and rural schools is highlighted by many researchers. For example, the *Presidential Committee on Youth Unrest (1993)* highlights that one of the main causes for the youth unrest in Sri Lanka in the 1990's is the *unfair distribution of schooling facilities between urban and rural sectors*. The following example will indicate the degree of disparity. In 1990, nearly 19 percent of the total university admissions were from *Colombo* district (*the administrative and commercial capital*) whereas it was only one percent from the *Moneragala* district which is considered as one of the most under-privileged districts. Further, the percentages which were selected to *medical, dental* and *architecture* faculties on merit basis[11] in 1990 were 41 percent from *Colombo* and zero from *Moneragala*. The population shares for these two districts are 11.4 and 1.8 percents respectively.

4. The Model

In a cross-section data set, like the one we have at our disposal, we observe only whether any age eligible child is in school or not and how many years (or levels) that particular child has already completed when the survey is conducted. For those who have left the school, the optimal length of education is observed. It is the number of years accumulated in their school career. However, for those who are in school, the distribution of optimal length of education is right censored.

4.1 Modelling the Schooling Decision

The conventional approach to model the schooling decision is to consider schooling decision as a sequential decision making process [*Hartog et.al (1989)*]. Within this framework, we assume that the decision making child decides whether to continue or to stop the education at the end of each period (year) in school. Then, the decision making criterion is the maximisation of the net-present value of the life-time wealth as we have

[11] Due to the regional disparity in facilities, university selection is done on three bases: the Merit Basis, the District Basis and the Under-privileged Basis. 40 percent is selected on Merit basis, 55 percent is on District basis and the remaining 5 percent is on Under-privileged basis. See: *Statistical Hand Book* the University Grants Commission.

already described in section 2.

In this paper, we use a cross-section data set to analyse the schooling decision. Further, the level of education of the respondents is not reported in years but in levels. Therefore, with regard to the schooling decision, what we observe in our data set is whether individuals are still in school or not and their schooling as an ordered variable. One can use this information to model the division of age eligible individuals into two categories as *current enrollees* and *school leavers.* An alternative method is to model the length of schooling directly considering the fact that some respondents from the age eligible population have not left school. Subsequent sections of this paper develop both these models and test them empirically for Sri Lanka. We use the same theoretical structure in both cases.

4.2 Enrolment Decision: *The Enrolment Model*

The enrolment decision can be modeled easily by defining the criterion according to our theoretical structure, a child in the post-compulsory schooling age must be in school if the following condition is satisfied.

$$\bar{d}_i \geq Age_i - x \tag{10}$$

where $\bar{d}_i$ is the optimal length of post-compulsory education as defined by the model. *Age* is the current age of the child and *x* is the age of compulsory education. In our case this is 10. We can define a binary variable E_i which is equal to one if the above condition is satisfied (and hence the child is in school) and zero otherwise.

Following the conventional practice, we can express the above decision making criterion in probability terms, namely we can write the probability of school enrolment.

$$P(E_i = 1) = P(\bar{d}_i \geq Age_i - 10) \tag{11}$$

Substituting equation (9) for d_i into the above probability expression and re-arranging the terms we can write:

$$P(E_i = 1) = P[\epsilon_i \geq -(\theta_o + \theta_1 M_i + \theta_2 B_i + \theta_3 U_i + \theta_4 G_i + \theta_5 I_i + \theta_6 N_i) + A_i] \tag{12}$$

Where, *A=Age - 10*. Now, the problem of defining a suitable probability distribution for ϵ remains. We assume that ϵ is normally distributed. This yields a *probit* binary choice model[12].

The likelihood function can be written as follows.

$$L = \prod_{i=1}^{n_1} [1-\Phi_i]^E \prod_{i=n_1+1}^{n} [\Phi_i]^{1-E} \tag{13}$$

where, $\Phi(.)$ is the CDF of ϵ which is assumed to be distributed normally. We assume that the data series is arranged in such a way that first n_1 observations contain all school enrollees and the remaining $n - n_1$ respondents have left the school. The total number of observations is n. Probit binary choice models identify the regression parameters only up to a fraction of the standard error ,

$(\frac{\Theta_k}{\sigma})$. We call this model *the enrolment model*.

4.3 Length of Schooling Life: *The Schooling Level Model*

An alternative approach is to model the realised length of schooling. This yields a *Tobit-like* equation of which the dependent variable is fully observed for school leavers, whereas for the current enrollees, the dependent variable is right censored. For this model, the dependent variable is the realised length of education and the censoring variable is a dummy indicating whether the child is currently in school or not. However, the estimation of this model involves further practical problems. The education level variable in this data set is not the actual years of schooling but levels. Education is reported in *six levels* as *no education (0) ,Grade one to five(1), Grade six to eight (2), Grade nine to ten (3), Passed G.C.E.(O.L) (4) and Passed G.C.E.(A.L) (5)*[13]. We define a *Tobit-like Ordered Probit (Ordered-Tobit)* model to deal with the problem.

In this section, we summarise *equation 9*, with the following notation for easy presentation. $\bar{d}_i = X^{/}\beta + \epsilon_i$. Where X is a column vector of all the variables on the right hand side and β is also a

---

12 We also estimated the *Logit* counterpart of the model. The results were reasonably similar to the probit results. However, the schooling length equation (second model) is *probit*. Therefore, in order to maintain the consistency in estimation procedures, we used the *probit* model for this as well.

13 In the entire labour force survey, we have eight levels: The six listed in the text plus two additional levels for degree and post-graduates. Due to insufficient information we excluded the degree and post-graduate categories from our analysis.Reason to leave out the respondents with zero education is justified by the model. In this model we consider only the post-compulsory schooling decision.

column vector of all regression parameters including the intercept. However, the survey data do not provide us the $\bar{d}$, the optimal length of education, of the respondents. Instead, we observe which category a given respondent belongs to. We assume a set of unknown thresholds, α_j. A particular respondent belongs to the lowest category (i.e. compulsory education) if his (her) optimal length of schooling, $\bar{d}_i$ is less than α_1. If $\alpha_1 < \bar{d}_i \leq \alpha_2$ respondent's optimal length of schooling is located in the second schooling category and so on.

By substituting the summarised form of *equation 9* and rearranging the terms we can write the probability that the individual i belongs to category j, $P(\bar{d}_{i_j})$.

$$P(\bar{d}_{i_j}) = \Phi(\alpha_j - X'\beta) - \Phi(\alpha_{j-1} - X'\beta) \tag{14}$$

where, $\Phi(.)$ represents the cumulative normal distribution function and α_j , α_{j-1} and β are regression parameters to be estimated. By assumption , all $\alpha's \geq 0$ and $\alpha_j \geq \alpha_{j-1}$ for all j's.

However, the reported level of education of the school enrollees treats a different story. For example, if a current enrollee reports that he(she) belongs to the jth schooling category, it indicates only that his(her) optimal length of schooling does not belong to a category less than the jth category. Therefore, we can write the following probability expression for the current enrollees.

$$P(\bar{d}_{i_j}) = 1 - \Phi(\alpha_{j-1} - X'\beta) \tag{15}$$

Now let *E=1* if the respondent is in school and *E=0* otherwise. Note that this is the observable counterpart of the dependent variable of the *enrolment model* described above. We can write the following likelihood function.

$$L = \prod_{j=1}^{4} \{\prod_{i=1}^{n_1} [1 - \Phi_i(\alpha_{j-1} - X'\beta)]^E \prod_{n_1+1}^{n} [\Phi_i(\alpha_j - X'\beta) - \Phi_i(\alpha_{j-1} - X'\beta)]^{1-E}\} \tag{16}$$

where we assume that the first n_1 observations contain school enrollees and the total number of observations is n. Further, $\alpha_0 = -\infty$ by definition. We call this model *the schooling level model*.

5. Data, Data source and Data Transformation

5.1 Data and Source of Data

The only survey which provides the current information concerning the variables of our model is the *Quarterly Labour Force Survey* conducted by the *Census and Statistics Department of Sri Lanka*. The current study is based on the information collected in *four quarters in 1993.*

However, the use of this survey data has several drawbacks. *First*, the information is not available regarding children below 10 years of age. Taking this into consideration, we develop our models assuming that the child makes her decision of the optimal length of schooling only after completion of the compulsory education, though there is a significant percentage of Sri Lankans with either zero level of education or who have dropped-out before the end of compulsory level.

Second, the information regarding the school leavers is not properly dated. For this study, we need the information regarding the school leavers when they decided to do so. However, this survey provides only the information at the time when the survey was conducted. Therefore, we have to assume that the information available to us provides good proxies for what is actually required.

5.1.1 The Quarterly Labour Force Survey

The survey is conducted quarterly to produce estimates of employment, unemployment, labour force participation, and basic demographic characteristics. The scope of coverage includes all households of Sri Lanka, though the *North* and *Eastern* provinces were excluded from the sample since 1990 due to the civil war.

Sampling method used for the quarterly labour force survey is named as the *Stratified two-stage probability sample design*. In the first stage, all 9 provinces of the country are divided into three sections as the *Urban* and *Rural* sectors, and the *Greater Colombo*[14] area. This consists of 19 domains. From among these 19 domains, 1,008 *Census blocks* were selected in the survey. The selection is done according to the population distribution in *1981 Population Census* so that the selection probability is proportional to the size of population in the block. Then, from each block, 10 housing units were selected at random for the survey. Everyone in the selected housing unit, of the relevant age group (+10) were included in the sample. Further, by the definition of

[14] Greater Colombo area is one part of the western province of Sri Lanka. This consists of the *Colombo M.C, the Dehiwela-Mt.Lavinia M.C and the Kotte U.C.*

a housing unit, (...*a house, an apartment, a group of rooms, or a single room when occupied as separate living quarters,...*) households with 5 or more lodgers and all *institutional units* such as hospitals and military camps etc. were excluded from the survey. Then, the entire sample was divided into four random groups for each quarter[15].

5.1.2 Background Information Regarding Data Generation

The *quarterly labour force survey* in 1993 has reported labour force information about 7,516 households which include 36,693 individuals. For the present study, we have to choose the households with at least one son or a daughter of the household head in age group 10 to 20 . This reduces the sample to 2,194 households. There were 8,811 children in the 10 to 20 age group. Then we choose the children from the 10 to 20 years age group whose reported education is below university degrees and above zero. This reduces the sample to 8,665. Out of them, 1,976 cases were not living with their parents and therefore *parents' information is not reported.* For them, we used the average of the missing family variables where averages are calculated using the reported observations.

5.2 Data Transformation

The questionnaire used in this survey provides comprehensive information about the labour force characteristics of the Sri Lankans. For the purpose of this study, some questions gave directly useful information whereas for the others, we had to impute the hypothetical variables with some reasonable assumptions. The following section provides a brief introduction for all the variables used in this study.

5.2.1 Level of Education

This is the dependent variable of one of our models. This question was asked from all the members in the sample unit. This reports the highest educational attainment of the respondent. This is reported in intervals.

---

15 This is the method used in general. However, due to some practical difficulties, the number of provinces and hence the number of census blocks have been reduced since 1991. The sampling method is published in the quarterly labour force reports published by the *Department of Census and Statistics*, Sri Lanka.

5.2.2 School Enrolment

Whether an age eligible child is in school or not is the dependent variable of the school enrolment model and it is the censoring variable of the second. Information was asked from all members in the household who are above 10 years of age and who reported that they are *currently unemployed and are not seeking work*. The question asks for the reasons, if the respondent was not actively working or not available or not looking for work. It is a pre-coded question and one of the answers provided to the respondent is *Studies*. We define *all who are in the relevant age group (10-20) and reported that they are studying as enrollees*. All the others in the same age group, including those who were not required to answer to this question[16] were considered *not enrolled*. Then the binary variable E_i is defined as zero for *non-enrollees* and one for *enrollees*.

5.2.3 Demographic Characteristics

Demographic characteristics such as *Age, Gender, Ethnicity, Residence, Family size, Birth Order mother's education* and *whether father is employer or self-employed* are either directly available or easily generated using the given information.

5.2.4 Imputed Variables

As the labour force survey covers only the income received as wages and salaries, income from other activities is not reported. Therefore, for those who earn income other than wage[17], the earnings variable is reported as missing. Further, also the reported income is truncated at the upper tail (*Rs.10,000 and above*). Finally, there are some respondents who had reported that the monthly income is less than *Rs.100*, which is not reliable. Therefore, we use the following procedure to impute their income[18].

We impute the income of the following categories. a.) people with income of *Rs.10,000* or more, b.)

---

16 Those who were currently employed or unemployed but reported as seeking work were not required to answer this question. Then, in fact, they were not attending school when the survey was conducted.

17 According to the survey definitions, income must be reported only for employees. If the labourforce status is either employer, self-employed, unpaid family worker, unemployed or not in the labourforce, income is not expected to be reported. However, we found that some employees had not reported the income whereas some respondents from non-employee categories have reported the income.

18 Note that we had one group of respondents for whom no information was available about their parents. For them, we simply impute the average of the manipulated income series. The following section of the text describes the manipulation procedure.

non-wage earners[19] and, c.) income reported as below *Rs.100.*

In order to estimate the income of above *Rs.10,000*, we assume that income is *log-normally* distributed. Then, we use the *TOBIT* estimation procedure to estimate the mean and standard deviation of the *truncated log-wage distribution*, which is then used to estimate the average earnings of the above *Rs. 10,000* group[20]. According to our calculation, the average income of the group with more than *Rs. 10,000* is about *Rs. 12,484* (the mean and the standard deviation of the truncated log income distribution are 7.5 and 0.7 respectively).

Then, the standard *Mincerian earnings function* is used to estimate the earnings of other missing cases. Regression results are presented in *Table II.* Regression results of the earnings equation are discussed in the next section.

Zero wage is assumed for those who have explicitly reported that they are unemployed or that they are not in the labour force.

The income series of the mother and father of each household[21] were aggregated to measure the *parents' income*.

However, in addition to the ambiguity in estimating the income of un-reported cases, it does not represent the income of the parents of school leavers. Actually, what we need is the parents' income when the decision was made. As far as the current enrollees are concerned we have the correct variable. For those who have left school however, unless we assume that the current income represents the income at the time they decided to leave school, current income is not a good proxy. Therefore, we also estimated an alternative specification replacing the parents' income variable with father's occupation. In that we divided all the reported occupations into four groups.

The first group includes *Legislators, Senior Officials and Managers* (*ISI Category I*) and *Professionals* (*Category II*). The second group includes the following *ISI categories. Technicians and Associate Professionals and Clerks. Service workers, shop and market sales workers, skilled agricultural workers Fishery wokers and*

---

19 Employers, self-employed and unpaid family workers. And also for the employees who had not reported income.

20 See Madala, pp. 365-367

21 We use the parents' income instead of the family income. This is simply because of the unreliability of the available information. With regard to the current enrollees, the information is accurate. However, for the school leavers, we do not know whether the other siblings were earning when the respondent decided to leave school. In fact, this is equally true also for the parents. However, assuming that parents were working at the time is more reasonable than assuming that the other siblings were working.

craft and related workers are included into group three. Then, *plant and machine operators, elementary occupations and Armed forces* are in group four. The control group consists of all others including unemployed, retired and not reported cases etc.

In addition to the variables defined by the model, we also included three quarterly dummies, a dummy to distinguish families with more than one age eligible member, ethnicity, whether father is an employer, whether father is self-employed or an un-paid family worker, whether the child is living with their parents and the location of household .

5.2.5 Descriptive Statistics

Means and Standard deviations of the continuous variables appearing in the model are presented in *Table I* below. In calculating means of the level of education of mothers and children, we assumed that the respondents have completed the highest level of education in the education category to which they belonged. For example, for those who reported zero length we assumed zero years, those who reported that they are belonged to category 1 (grade 1 to 5), we assumed 5 years of education and so on. The second part of *Table I* presents the related statistics for categorical variables in the model. According to that 75 percent of respondents belonged to *Sinhala* ethnic group, nearly 75 percent were school enrollees, 83 percent were with more than one age-eligible respondent in the sample and so on.

Table I: Descriptive Statistics

	Name of the Variable	Mean	Standard Deviation
	Continuous Variables		
a.)	Age of the Respondents (Years)	14.92	3.14
b.)	Birth Order	2.36	1.21
c.)	Length of Education (Years)	8.62	2.00
d.)	Parental Income (Rs.)	2289.80	1735.65
e.)	Mother's Education (Years)	7.98	3.01
f.)	Family Size	5.96	1.57
	Categorical Variables	Proportion of "Yes" Responses	Standard Deviation
a.)	Current Enrollees	0.75	0.43
b.)	Sinhalees	0.78	0.41
c.)	Father is an Employer	0.04	0.17
d.)	Father is self-employed	0.33	0.41
e.)	More than one Member in the Sample	0.83	0.33
f.)	Male Respondents	0.51	0.50
g.)	Urban Respondents	0.54	0.50
h.)	Respondents from Quarter I Survey	0.24	0.42
i.)	Respondents from Quarter I Survey	0.25	0.43
j.)	Respondents from Quarter II Survey	0.26	0.44
k.)	Respondents from Quarter III Survey	0.25	0.43
l.)	Respondents Living with Their Parents	0.77	0.42

Source: Selected Sub-sample from *Quarterly Labour Force Survey- 1993:* Census and Statistics Sri Lanka

6. Model Estimation and Empirical Results

6.1 Earnings Function

Table II presents the regression results of the earnings function. As indicated by the *Adjusted* R^2 *38* percent of the total variation of the log wage is explained by our specification of the earnings function. All the regression coefficients are highly significant and with the expected signs. On average males earn 20 percent more than females do. Our result is consistent with other estimates available for Sri Lanka. For example, according to *Kelly's (1993)* provisional estimates, the same coefficient is *0.28.* This is also consistent with our assumption concerning the earnings differential between male and female.

The earnings profile is maximised at age *47*. On average Sinhalese earn less than the others do. However, this should not be interpreted in terms of the racial distribution of income in the country because the *North* and the *Eastern* provinces where the predominant ethnic group is the ethnic minorities in the country are excluded from this sample survey.

Vocational training which is vaguely defined in the survey shows a highly significant effect on earnings.

On average, vocational training increases income by *20* percent.

The marginal rate of return to education shows an interesting pattern. If we set the marginal duration for O. Level., A. Level and university at 1, 2 and 4 years respectively, we can obtain an indication by simply dividing the differences in coefficients by additional school years. For example, per-year marginal rate of return to the primary education (grade 1 to 5) can be obtained as 0.1/5 and the same for the education level 6 to 8 can be calculated as (0.29 - 0.1)/2 and so on. The marginal rate of return to education per-year increases from 2 percent for the primary education to 9 percent for people with 6 to 8 years of education. The marginal rate of return is around 9 to 11 percent for all the reported education levels except for the G.C.E.(O.L). For the G.C.E (O.L), the marginal rate of return is 35 percent. On average the return to an extra year of education is about 9 percent, but there is a big jump at the O. Level. One of the explanations for this pattern is the *effect of certification*. According to the prevailing education system, the O.Level is the first exam in Sri Lanka which issues a valid certificate that can be used on the labour market[22]. Therefore, this can be considered to be the effect of the certification. Another explanation is that the O.Level is the minimum educational qualification required to be recruited for many white collar jobs in Sri Lanka. As the wage differential between the blue collar and the white collar jobs is substantial, this can also be considered as the wage premium for white colour occupations.

However, the estimation of the rate of return to education is not the objective of the present paper. The estimated earnings function is only a supplementary part of the paper. Therefore, we used the simplest specification of the earnings function. These results must be considered as the preliminary findings and one has to be careful in interpreting the results.

[22] In 1992, nearly 30 percent of the school candidates were able to get through the O.L. exam.

Table II : **Estimates of Earnings Equation**

Name of the Variable	Coefficient (t-values)
Intercept	5.92 (101.41)
Age	0.04 (14.07)
Square of Age	-0.0004 -(11.77)
Education (Primary)	0.10 (3.14)
Education (6 to 8)	0.29 (8.83)
Education (9 to 10)	0.51 (16.11)
Education (O.L)	0.87 (27.48)
Education (A.L)	1.08 (31.76)
Education (Degree & Above)	1.41 (33.07)
Vocationaly trained	0.20 (11.64)
Ethnicity (Sinhala=1)	-0.04 -(2.62)
Gender (Male=1)	0.20 (13.85)
Adjusted R^2	0.38
F statistic	418.82
Standard Error of Regression	0.55
Number of Observations	7,378

Note Figures given in parentheses are *t-statistics*

Source: Selected Sub-sample from Quarterly Labour Force Survey- 1993: Census and Statistics, Sri Lanka

Table III:Models of Enrollment and School Level Selection
(All Cases included- Model with Parents' Income)

	Enrollment Model		School Level Model	
Name of the Variable	Coefficient	t-stat	Coefficient	t-stat
Intercept	5.32	33.44	-	-
Gender of the child (Male = 1)	-0.01	--0.33	-0.02	-0.59
Age - 10 of the respondent	-0.35	-45.00	-	-
Ethnicity (sinhala = 1)	0.28	6.29	0.46	8.48
Quarter II	-0.06	-1.05	0.02	0.34
Quarter III	0.03	0.56	0.04	0.57
Quarter IV	-0.02	-0.31	0.04	0.53
Urban Residence	0.16	3.60	0.32	6.04
Birth Order of the Child	0.10	5.39	0.12	1.64
Mother's Education	0.07	9.26	0.12	12.40
Family with more than 1 Child in Sample	0.35	4.84	0.49	4.71
Family Size	-0.05	-3.65	-0.02	-1.10
Parents' Income	0.08	5.68	0.08	4.13
Father is an Employer	0.20	1.60	0.20	1.13
Father is a self-employed	0.06	1.21	0.11	1.67
Parent's information not reported	-0.49	-11.38	-0.56	-10.47
$\hat{\alpha}_1$	-	-	0.00	0.00
$\hat{\alpha}_2$	-	-	0.72	54.05
$\hat{\alpha}_3$	-	-	1.79	70.03
$\hat{\alpha}_4$	-	-	3.14	47.95
Mc Fadden R^2	0.38	-	0.11	-
$\hat{\rho}$ Error Correlation of Two Models	-0.33	-	-	-
Number of Observations	8,665	-	8,665	-

Source: Selected Sub-sample from Quarterly Labour Force Survey- 1993: Census and Statistics, Sri Lanka

Table IV: Models of Enrollment and School Level Selection
(Excluded the Cases Not Reported Parents' Information: Model with Parents' Income)

	Enrollment Model		School Level Model	
Name of the Variable	Coefficient	t-stat	Coefficient	t-stat
Intercept	5.37	29.99	-	-
Gender of the child (Male = 1)	-0.09	-1.97	-0.09	-2.10
Age - 10 of the respondent	-0.34	-37.46	-	-
Ethnicity (sinhala = 1)	0.21	3.85	0.44	9.56
Quarter II	-0.03	-0.57	0.11	1.97
Quarter III	0.06	0.98	0.17	2.94
Quarter IV	0.001	0.12	0.11	1.96
Urban Residence	0.14	3.00	0.32	7.09
Birth Order of the Child	0.11	5.47	0.12	6.55
Mother's Education	0.07	9.32	0.13	18.13
Family with more than 1 Child in Sample	0.35	4.83	0.52	6.94
Family Size	-0.06	-3.90	-0.02	-1.95
Parents' Income	0.08	5.66	0.08	5.30
Father is an Employer	0.20	1.59	0.22	1.77
Father is a self-employed	0.06	1.23	0.12	2.46
$\hat{\alpha}_1$	-	-	0.00	0.00
$\hat{\alpha}_2$	-	-	0.74	45.31
$\hat{\alpha}_3$	-	-	1.87	62.92
$\hat{\alpha}_4$	-	-	3.39	41.16
Mc Fadden R^2	0.38	-	0.13	-
$\hat{\rho}$ Error Correlation between Two Models	-0.30	-	-	-
Number of Observations	6,689	-	6,689	-

Source: Selected Sub-sample from Quarterly Labour Force Survey- 1993: Census and Statistics, Sri Lanka

Table V: Models of Enrollment and School Level Selection
(All Cases included: Model with Father's Occupation)

	Enrollment Model		School Level Model	
Name of the Variable	Coefficient	t-stat	Coefficient	t-stat
Intercept	5.32	31.55	-	-
Gender of the child (Male = 1)	-0.03	-0.67	-0.03	-1.02
Age - 10 of the respondent	-0.35	-44.52	-	-
Ethnicity (sinhala = 1)	0.30	6.50	0.47	12.15
Quarter II	-0.05	-0.95	0.02	0.50
Quarter III	0.04	0.77	0.04	0.93
Quarter IV	-0.001	-0.02	0.04	0.92
Urban Residence	0.13	3.19	0.30	7.95
Birth Order of the Child	0.10	5.21	0.12	6.45
Mother's Education	0.09	12.91	0.14	23.37
Family with more than 1 Child in Sample	0.10	1.67	0.15	2.78
Family Size	-0.06	-4.17	-0.04	-2.84
Father has a Category I Job	0.85	6.72	0.93	6.66
Father has a Category II Job	0.63	6.01	0.53	5.08
Father has a Category III Job	0.12	2.16	.014	2.88
Father has a Category IV Job	-0.01	-0.16	-0.05	-0.89
Parent's information not reported	-0.32	-5.70	-0.39	-7.74
$\hat{\alpha}_1$	-	-	0.00	0.00
$\hat{\alpha}_2$	-	-	0.73	55.99
$\hat{\alpha}_3$	-	-	1.79	72.90
$\hat{\alpha}_4$	-	-	3.11	45.52
Mc Fadden R^2	0.35	-	0.10	-
$\hat{\rho}$ Error Correlation Between Two Models	-0.03	-	-	-
Number of Observations	8,665	-	8,665	-

Source: Selected Sub-sample from Quarterly Labour Force Survey- 1993: Census and Statistics, Sri Lanka

Table VI: Models of Enrollment and School Level Selection

(Excluded the Cases Not Reported Parents' Information: Model with Father's Occupation)

	Enrollment Model		School Level Model	
Name of the Variable	Coefficient	t-stat	Coefficient	t-stat
Intercept	5.38	28.24	-	-
Gender of the child (Male = 1)	-0.10	-2.37	-0.11	-2.63
Age - 10 of the respondent	-0.34	-36.91	-	-
Ethnicity (sinhala = 1)	0.23	4.11	0.45	9.76
Quarter II	-0.03	-0.41	0.11	2.01
Quarter III	0.08	1.23	0.18	3.07
Quarter IV	0.03	0.49	0.13	2.17
Urban Residence	0.10	2.16	0.28	6.34
Birth Order of the Child	0.11	5.29	0.12	6.47
Mother's Education	0.09	12.93	0.14	23.14
Family with more than 1 Child in Sample	0.10	1.72	0.15	2.69
Family Size	-0.07	-4.38	-0.04	-2.84
Father has Category I Job	0.84	6.66	0.95	6.69
Father has Category II Job	0.64	6.08	0.55	5.14
Father has Category III Job	0.12	2.14	0.13	2.71
Father has Category IV Job	-0.02	-0.27	-0.08	-1.26
$\hat{\alpha}_1$	-	-	0.00	0.00
$\hat{\alpha}_2$	-	-	0.75	45.39
$\hat{\alpha}_3$	-	-	1.87	62.62
$\hat{\alpha}_4$	-	-	3.34	41.00
Mc Fadden R^2	0.38	-	0.13	-
$\hat{\rho}$ Error Correlation Between Two Models	-0.05	-	-	-
Number of Observations	6,689	-	6,689	-

Source: Selected Sub-sample from Quarterly Labour Force Survey- 1993: Census and Statistics, Sri Lanka

6.2 School Enrolment and Schooling Length

Depending on the variables and the number of observations that we used in the estimation, we estimate four different versions of each model. The results are presented in *Tables III to VI.* Regression estimates differ first by the type of the variables we used. As we have already explained in a previous section of this paper, the family income variable is imputed for some households. Therefore, in addition to the models with family income variable, we also estimate the two models replacing the family income variable with the father's occupation. Table III and IV give the models with family income while V and VI give the models with father's occupation.

Again, due to various reasons, some respondents in our sample are not living with their parents and therefore family information is not reported. Taking this into account, first we estimate each model (i.e., with two sets of variables) for the entire sample. In this case, we replaced the missing parents' characteristics by the average parents' characteristics of the reported cases and included a dummy to differentiate that group from the others. (*Parents' Information Not Reported = 1*). Then, we also estimate the models only with observations that include parents' information. This reduces the total number of observations from 8,665 to 6,689 cases. Table III and V give the regression results with all the observations (with mean imputation of the missing parents' characteristics) while table IV and VI give the results when we used only the *parents' information reported cases*.

Cross comparison of table III with table IV and table V with table VI shows the effect of the exclusion of the respondents who are not living with their parents. The exclusion of these respondents has affected mainly the coefficients attached to the gender in *both models*, quarterly dummies in the *schooling level model* and the ethnicity in the *enrolment model*. Though the signs are not affected, all these coefficients, except the ethnicity dummy have increased in magnitude and become nearly significant when those respondents living away from the parents were excluded.

In the school level model with parental income, excluding incomplete- information cases reduces all standard errors. The ethnicity effect is significant in all the models. When the *parents' information not reported cases* were excluded, the magnitude of the ethnicity effect dropped slightly. The dummy representing the *parents' information not reported* respondents appears to be highly significant. The sign of this coefficient is negative in all the models indicating that the majority of the respondents living away from the parents' homes have left school and they have a shorter schooling length when compared to the schooling length of their counterparts.

The difference between male and female enrolment disappears if incomplete- information cases are dropped. The difference is mainly visible for children still living with their parents.

Using father's occupation instead of parents' income leaves most results unaffected. Only one coefficient really changes. That is the dummy which distinguishes single respondent families from multi-respondent families (if more than one child is in the sample this dummy equals one). When the parents' income is replaced by the father's occupation, the effect of this dummy variable is reduced substantially. The effect of the family size on the schooling length (ordered probit) is increased. However, it is hard to find any explanation for this change. Probably the ambiguity of the reference category in the occupational dummy model would have some effect on that. Children of the unemployed fathers, retired fathers and those for whom the information with regard to father is missing were lumped into the reference category. This makes the reference category rather heterogeneous.

The overall significance as indicated by *Mc-Fadden* R^2 is quite low in the ordered probit equation whereas it is satisfactory in the enrolment model. At best, only 38 percent of the variation in the enrolment decision and 13 percent of the schooling levels is explained by our model.

Let's now turn to a discussion of the substance of our results.

If our assumptions regarding the labour market and the education system are correct, the regression coefficient attached with the age in the enrolment equation must be one *(cf. equation 12)*. However, as the probit specification of our model helps us to identify all the regression parameters only up to a proportionate of σ (standard deviation) the regression coefficient attached to the age variable can be interpreted as $1/\sigma$, the inverse of the residual variance. This implies that $\sigma = 2.86$ [23].

However, the accuracy of this estimate depends on the reliability of our assumptions.

As we have already described in the theoretical framework, whether the ability effect is positive or not can be tested indirectly. By definition the regression coefficient attached to the mother's education is $\hat{\theta}_1 = [\ \mu_1(\beta_1 - \gamma_1)\ ]\ /\ (\ b_1 + c_1\ )$. Our results show that this is positive and statistically significant in all the models that we have estimated. In the enrolment model, the coefficient is nearly 0.1 and it is 0.14 in the ordered probit model. As we have presumed at the beginning of this paper, if our assumptions concerning μ_1

[23] The reported standard deviation of the schooling length (d) in table I is 2.This is smaller than σ=2.86. Since the length of schooling is reported only in intervals, to calculate the standard deviation we assume that children in each educational category have reached to the highest level within the category. [see, page 14]. However, when we replace that with the mid-point of each education category, [i.e. (1+5)/2=3 for the lowest and (6+8)/2=7 for the second and so on], the standard deviation of the schooling length turns out to be nearly 3 years.

and $b_1 + c_1$ are true (i.e., they are positive), $\hat{\theta}_1$ can be positive only if. $\beta_1 > \gamma_1$. This indicates that the *ability effect* on the slope of the earnings function (i.e., the effect of the ability on the return to education) is greater than that on the intercept of the earnings function (i.e., the effect of the ability on the cost of education). In other words, *the earnings functions for a person with different abilities diverge and therefore more able children stay longer in school than the less able children*.

The *Birth Order effect* is positive and is significant at *5%* . The *Behrman- Taubman hypothesis* claims *ceteris-paribus* the younger members of a family must be less able than the older members of the same family. In fact, this hypothesis itself is quite controversial. There are many counter arguments. By definition, $\hat{\theta}_2 = [\ \mu_2(\beta_1 - \gamma_1)\ ]\ /\ (\ b_1 + c_1\ )$. By assumption $b_1 + c_1 > 0$ and we have also shown that $\beta_1 > \gamma_1$. Then, the sign of $\hat{\theta}_2$ is entirely dependent on the sign of μ_2. As the birth order effect is positive in all the regressions that we have estimated, we can conclude that $\mu_2 > 0$. In contrast to the *Behrman and Taubman* hypothesis, in our sample, the younger members are more able than the older members in a given family. *Behrman and Taubman's* claim that the younger are less able depends on two conditions. *First*, it is the presumed negative relationship between the genetic endowment and the birth order. *Second*, the marginal utility of child rearing is subject to decreasing returns. Therefore, younger children receive less attention from the parents. The first argument is essentially a non-economic one. Despite the accuracy of that one can find many reasons to reject the second assumption. Probably the younger siblings receive better parental care because parents are now more experienced in child rearing[24] [25].

θ_3 the place of living effect on education is positive and statistically significant in all the models. This implies that the market rate of return to education is higher for the urban residence than that for their rural counterparts. This is congruous with our theoretical expectation.

Gender, θ_4 (*Male=1*) has negative effect on optimal schooling levels. However, the effect is only significant if incomplete-information cases are excluded. This suggests that the effect is significant among children still living with their family and insignificant among children who have left their home. This suggests

---

24 For further criticisms of Behrman-Taubman hypothesis see, Zvi Griliches, (1986)

25 Athurupana (1992): (Unpublished Ph.D Thesis) has reported the same results for Sri Lanka.

a particular type of selectivity: boys and girls who leave the parental household before age 20 are not different in their demand for schooling, boys who remain in the household have a lower schooling demand than girls. *Athurupana (1992)* estimates separate models for *Urban, Rural and Estate* sectors of the country. He found that urban males are more likely to stay longer in schools whereas in the rural sector the opposite is true. In the estate sector, according to his findings, there is no consistent association between these two variables[26].

According to the specification of the model in this paper, the weakly negative association between schooling decision and gender can be expected if the gender effect on the return to education is weakly less than the gender effect on the cost of education. The insignificant coefficient in our estimate shows that the return and the opportunity cost of education for males differ by about the same magnitude from return and cost for females. $[\beta_3 \approx \gamma_4]$.

According to our specification, *Athurupana's* finding that in the urban sector males have longer schooling career than females have and in the rural sector the reverse is true imply that $\beta_3 > \gamma_4$ is true for the urban sector and $\beta_3 < \gamma_4$ is true for the rural sector. Since we have controlled for the rate of return variation over place of living in our model, we can safely assume that β_3 is same for both urban and rural sectors. Then the above inequality implies that γ_4 is varying over urban and the rural sectors such that cost of education for rural male is higher than the cost of education for an urban male. Probably high incidence of rural poverty may be an explanation for this. As the role of male in economic activities in rural sector is more important than that of a women, families rely more on the boys' assistance to overcome the economic hardship and therefore rural boys may quit school sooner than their girl counterparts.

The estimated coefficient θ_5 reveals the sign of γ_2 to be negative, indicating that a higher parental income reduces the cost of obtaining funding for school attendance, as we anticipated. We find the same result if we use father's occupation instead of parental income. Despite the heterogeneity in the reference category, comparison of the regression coefficients attached to the occupation dummies shows that *ceteris-paribus*, children of the fathers with *professional or administrative (Managerial*) level occupations have much higher probability to stay in school and complete higher levels of education than the children of fathers with *minor occupations*. There is a clear positive association between the father's job level and the education decision of the

[26] When we apply our model for the three sectors separately we find that our results are consistent with *Athurupana.*

child so that the children of the fathers with higher positions in the occupation status are more successful in their educational attainment.

The estimated coefficient of θ_6 ,negative in all models, reveals the sign of γ_3 to be positive: in larger families ,the cost of investment funds is higher. This also in confirmity with our expectations. The effect of living in a family with more than one age-eligible member on the demand for schooling is positive. In a structural interpretation, this effect should come about through cost rather than returns. Hence, the result suggests that cost of investment funds are lower in multi-respondent families than in single-respondent families. This result is rather puzzling, in particular if we recall the finding that the effect is reduced substantially if we use father's occupation rather than parental income[27]. In other words, with family income constant, there is a strong effect , with occupation constant, there is a weak effect. This suggests that it is not a straight cost effect, because one would expect that to be smaller if income is held constant. An alternative explanation for this apparently controversial finding is however that there is an e*conomies of scale* effect for the respondents with more than one member in the eligible age.

Table VII: Simulation of Probability Distribution of Highest Schooling Level Completed: A

	Reference Category (a)	Sinhala Ethnicity (b)	Urban Residence (c)	Mother has a Degree (d)	(b) & (c)	(b), (c), (d) &Family Income=Rs. 10,000
Below 5	9.68	3.92	5.26	1.16	2.84	0.20
Grade 6 to 8	18.42	11.00	13.15	4.90	5.85	1.34
Grade 9 to 10	40.69	36.28	38.34	25.50	29.90	12.25
G.C.E. (O.L)	27.92	40.42	36.82	49.22	46.95	46.47
G.C.E. (A.L)	3.29	8.38	6.43	19.22	14.46	39.74

Note: (a): Reference Group in *Regression equation with Parents' Income.*
(b): Effect of Ethnicity (*all other variables are same as the Reference group)*
(c): Effect of Place of Residence (*all other variables are same as the Reference group)*
(d): Effect of Mother's Education (*Mother's Education is Degree: 16 years: all other variables are same as the Reference group)*
(b) & (c): Sinhala Urban Residence (all other variables are same as the Reference group)
Last column: Sinhala, Urban Residence with Mother's Education is degree *(all other variables are same as the Reference group)*
Source :Selected Sub-sample from Quarterly Labour Force Survey- 1993: Census and Statistics, Sri Lanka

27 When we droped the family size and the dummy for more than one member in the sample, we found that the coefficient attached to birth order variable is unaffected. We also estimated the models excluding only the dummy for more than one member in the sample and still the results were unaffected.

Table VIII
Simulation of Probability Distribution of Highest Schooling Level Completed: B

	Reference Category	Father has a Level I Job	Father has a Level II Job	Father has a Level III Job	Reference Category with Uneducated Mothers	Job I Category with Graduate Mothers
Below Grade 5	12.30	1.83	4.55	9.68	48.01	0.06
Grade 6 to 8	28.51	6.86	28.81	18.75	27.16	0.58
Grade 9 to 10	32.76	29.52	40.21	40.36	20.74	7.00
G.C.E. (O.L)	23.87	46.40	23.87	27.70	3.98	37.98
G.C.E. (A.L)	2.56	15.39	2.56	3.51	0.11	54.38

Note: These percentages are calculated using regression equation with father's job levels instead of family income.
Source: Selected Sub-sample from Quarterly Labour Force Survey- 1993: Census and Statistics, Sri Lanka

7. Variation in Schooling Length

Using the regression results presented in the previous sections, we perform a simulation to understand the sensitivity of the schooling length to changes in different variables. As the tables VI and VII show the relative frequency of ultimate school levels is highly sensitive to different socio-economic background variables.

Table VII and VIII present several predictions of the highest school level completion rates. These figures were calculated using the *schooling level models* in tables III and V. In the case of continuous variables, average values are used. Table VII reports the sensitivity of school level completion rates to *ethnicity, place of living, mother's educational attainments and family income.* Column (a) reports the frequencies of the reference group. The reference category in the model reported in table III is *Woman*, *minority ethnic groups (tamil, moor etc.), people living in rural sector, quarter I of the year, families with only one child in the sample and cases for which the information with regard to parents is reported*. We use the averages of the birth order, mother's education, family size and parents' income. Comparison of column (a) with (b) shows the ethnic differences in school enrolment. Our results show that the Sinhalese are more likely to stay longer in school than the other minority ethnic groups. However, we must be very cautious in interpreting this result because the labour force survey has not covered the North and Eastern Provinces where the minority ethnic groups are highly concentrated. Almost all the minority groups covered by our sample are from the estate sector where the schooling quality is extremely poor and due to various reasons, education is not given high priority in that society.

The movement from column (a) to (c) shows the effect of the place of residence on the schooling levels. As depicted in column (c), the change in place of residence has dropped the percentage of the below grade 5 category from 9.68 to 5.26 percent while the last category (A.L or above) has increased from 3.29 to 6.43 percent. In part this jump may be attributed to the changing schooling quality over urban and rural sectors. Although there is a quality variation even within sectors, still the average quality in urban schools is much higher than the average quality of education in rural sector. If we assume that the main difference between the urban and rural sectors is the difference in schooling quality, we can argue that children living in rural sector are less likely to be trained in a high quality schools. As *Card and Krueger (1993)* argued, the quality of school affects the productivity and hence the marginal rate of return. A child trained in a high quality school has a better training and hence is more productive. This shifts the rate of return curve upward. At the same time, high school quality means better teaching, more academic oriented peer groups etc. All these have a negative effect on the discount rate because they increase the *Ability*. Therefore, high school quality can have a positive effect on the length of schooling.

Reading the frequencies in column (d) shows that the mother's education has an enormous effect on the schooling decision. *Ceteris-paribus*, increase of mother's education from the average level (8 years) to university degree level (16 years) has increased the proportion of people in the *A.L. and Above* category from 3.29 for the reference category to 19.22 percent. The last two columns of the table show two extreme cases. Column [(b) & (c)] gives the frequency of school levels for the Sinhalese living in the urban areas and the last column gives the frequency for Sinhalese, living in urban sector, mother's education is a university degree and the monthly income of parents is Rs.10,000. The comparison of this column with the frequencies reported in column (a) shows the difference.

Table VIII reproduces the same information with job level breakdowns. These frequencies were calculated using the ordered probit results reported in table V. Since the *occupation category IV (plant and machine operators, elementary occupations and Armed forces)* is insignificant in our model, we ignored it in the analysis. The reference category is *occupation level I (Legislators, Senior Officials and Managers and Professionals).* Again we note a strong effect of family background. The last two columns compare the schooling decision of the respondents from two extreme categories. The last but one column gives the estimated education composition of the reference category when the mother has no education and the last column gives the estimated education composition of the respondents with fathers in *occupation level I (Legislators, Senior Officials and Managers and Professionals)* and mothers with *a university degree*. Almost 50 percent of the category with *uneducated mothers* have left school just after the compulsory education while little above 50 percent of the respondents with fathers in *occupation level I (Legislators, Senior Officials and the Managers and Professionals)* and mothers with *a university degree* have reached the highest steps of the education ladder in Sri Lanka.

7.1 Decomposition of the Variation in Schooling

Right hand side of the specification of schooling model can be divided into five sets of variables as a.) common variables which appear in both the rate of return and cost of education equations (ability and gender) , b.) variables appearing only in the rate of return equation (place of residence) , c.) variables appearing only in the cost equation (parents' income and family size), d.) control variables that are neither in the cost curve nor in the rate of return curve and e.) an unexplained component.

In this section, we will decompose the total variation of the schooling length into the four components described above. The objective of this exercise is mainly to understand the relative contribution by each component to the variation in schooling length. First, we describe the methodology of the decomposition and then give the results.

Let,

$$Y_i = X_{1_i}\lambda_1 + X_{2_i}\lambda_2 + X_{3_i}\lambda_3 + X_{4_i}\lambda_4 + \nu_i \tag{17}$$

where, *Y* denotes the dependent variable, X_1 , X_2 , X_3 and X_4 represent n by k_j matrices of the common variables, variables which appear only in the rate of return equation, variables which appear only in the cost equation and the controlling variables respectively. λ's are column vectors of their coefficients. ν is the unexplained component. In relation to equation 9, X_1 represents mother's education (M), birth order (B) and gender (G) and $\lambda_1' = [\theta_1, \theta_2, \theta_4]$. X_2 is the locational dummy (U) and $\lambda_2 = \theta_3$.X_3 is parents' income (I) and family size (N) while $\lambda_3' = [\theta_5, \theta_6]$. X_4 represents all other explanatory variables in the regression model. Then, assuming that these components are independent from each other, we can write the variance of *Y*, *V(Y)*:

$$V(Y) = \lambda_1' V(X_1)\lambda_1 + \lambda_2' V(X_2)\lambda_2 + \lambda_3' V(X_3)\lambda_3 + \lambda_4' V(X_4)\lambda_4 + V(\nu) \tag{18}$$

where, $V(X_1)$, $V(X_2)$ and $V(X_3)$ are diagonal matrices (by the assumption of linear independence of all X variables) of each set of variables represented by X_1 , X_2 and X_3 respectively. *V(Y)* and the first four components of the right hand side can be estimated directly from the data and the $V(\nu)$ can be derived from it.

Then, dividing both sides of equation (18) by V(Y), we can define a simple test to identify the relative significance of each component to the total variation. We performed this test using the results of *School Level Model* (Ordered probit) reported in *Table III.*

According to our results, 2.9 percent of the total variation of the length of education is due to the variation in the variables appearing in both the rate of return and the cost of education curves: mother's education, birth order and gender are the common variables. The percentage of the total variation due to the variables appeared only in the rate of return equation (place of residence) is 0.5. Only 0.4 percent is due to the variables appearing only in the cost equation. 1.4 percent is due to the control variables. Then, the remaining unexplained component has contributed 94.8 percent of the total variation of the length of education. This essentially indicates the need for an extension of the model with more information.

Nearly one percent of the total variation of the schooling length is explained by the rate of return and the cost of education variations in our model. Therefore, it is hardly possible to draw any valuable policy conclusion. However, we observe that out of the very small proportion of the schooling variation explained by our regression model, the rate of return variation is greater than that of the cost of education. Though this is negligible, one would argue that this together with our earlier observation that the ability effect on the rate of return is greater than the ability effect on the cost of education (i.e. $\beta_1 > \gamma_1$) shows that one of the main reasons for the respondents in our sample to have different length of schooling is the variation in the rate of return, not the variations in the cost of education. In other words, the variation in the educational attainment across individuals are determined more by the fluctuations of the marginal return function than that of the marginal cost function.

7.2 Analysis of Residuals

Since we have information on siblings, sharing the same family background , we can search for family effects in the error terms. Using the estimated residuals of the *enrolment model* (*table III)* we correlate the error components of the children from the same family with each other. *Table IX* reports the results. The correlation between the error components of the first and the second child is 0.46 and it is 0.43 between the second and the third child. It is nearly 0.4 between the third and the fourth child. All these correlations are significant at one percent These indicate that there is an unexplained fixed family effect. *Table IX* also shows that the correlation declines as the distance between any pair of children increases. For example, the correlation between the first and the third one is 0.22 and that between the first and the fourth one is 0.09. Probably, this may be due to the changes in the family environment in the long-run. The family environment for a subsequent pair of children may be more or less the same. However, as the distance between any pair of children increases we can assume that the family environment for the two births would be different. Since coefficients estimated are still unbiased under neglected heteroscedasticity (though not efficient), we did not adjust our econometric specification.

Table IX: **The Correlation Matrix**
Error Correlation Coefficient Between the Respondents from the Same Family

	First Child	Second Child	Third Child	Fourth Child
First Child	1.00	0.4637	0.2262	0.0897
Second Child	-	1.00	0.4338	0.2150
Third Child	-	-	1.00	0.3687
Fourth Child	-	-	-	1.00

Note: Fifth and higher order children were excluded because we have only few observations
Errors of the School Enrollment Model, Table III
All correlation coefficients are significant at 1 percent

We also estimate the correlation coefficient between the error terms in *enrolment model* and that in the corresponding *schooling level model.* (see *tables III to VI*). One would expect that common omitted variables imply a positive correlation between the errors in the two models. But, remarkably, the unexplained part of the two regressions are negatively correlated. The correlation coefficient is very low when we use the father's occupation instead of the parents' income. However, it is rather high as -0.3 when we use the father's occupation instead of the parents' income. Perhaps, this negative association may be due to the fact that the excluded schooling quality would have affected both the error terms in opposite ways. One possible explanation is that children from *disadvantaged schools* would have to retain at school longer (for a given age) due to high repetition rates[28]. At the same time they hardly complete the successive schooling levels. This implies that when we control all other variations (except for school quality), children from *disadvantaged schools* are more likely to stay at school longer due to high repetition rates but as far as the school achievements are concerned they have

[28] For example, in the year 1991, 5 percent of the students in government schools in Colombo repeated their grades more than once. Whereas the repetition rate in Moneragala was 12 percent. (see, Educational Statistics of Sri Lanka (1992); table 40.

poor performance. Thus the excluded school quality makes an upward bias in *enrolment decision* while the same makes a downward bias in *the schooling length decision* or *vice-versa.*

7.3 Policy Implications

If the education decision is heavily dependent on family characteristics, policy makers would be interested to find ways to eliminate the family background effect on the schooling decision. In our model, the only variable that can be affected by policy makers is parental income. Therefore, in this section we estimate the per-child compensation that is needed to eliminate the family background effect on schooling decisions. Our estimates are presented in *Table X.* In this endeavour, first we divide the whole sample into four income groups by income quartiles. *Table X* consists of two sections. In the upper part of the table we have presented the number of children from each income quartile, means of the continuous variables in the model and estimated relative frequencies of school completion for each income bracket. These estimates are done on the basis of the ordered probit equation in table III. Further, the estimates are done only for the reference category of that equation [*See page 27*]. Column two of the lower part of table X presents the estimated number of children from each income bracket. This estimate was done by multiplying the total school population in 1992 (as reported in the school census:1992) by the relative frequencies of the children in each income bracket in our sample (as in column two of the upper part of table X).

Then, we consider two policy regimes: regime A and regime B. Policy regime A is to equalise the school completion frequencies of the three lower income quartiles to the school completion frequency of the top income quartile. For example, our desire is to reduce the drop-out after compulsory education to 5.16 percent and to increase the percentage of A.L. completion to 6.55 percent. In this case, compensation is required only for the first three income quartiles. Setting the school completion probabilities to our desired levels, we can solve the ordered probit equation for the level of income that each category should receive if they want to report the expected sachool completion probabilities by the policy regime A. Then, the difference between the estimated required income and the mean income of a given income bracket (column three of the top part of table X) gives the per-family cost for the implementation of policy regime A. Then, dividing that by the estimated average number of children in each income quartile (average family size - 2) we obtain the cost per-child. This is reported in column three of the lower part of table X. Then multiplying the per-child cost by the number of children in each income bracket, we have the total cost of the policy regime A. In fact, the cost to implement policy regime A is quite high. Our estimate shows that it is nearly 4,200 million Sri Lankan rupees [nearly 85 million US dollars] which is nearly 5 percent of the total government revenue. If this policy regime were implemented, our estimate shows that the average length of education of the society will increase by one year. However, with the already existing budget deficit in the country this seems to be a hardly achievable target.

Policy regime B is to change the school completion probabilities of the lowest income quartile to the school completion probabilities of the second quartile. As far as its effect on the average education level of the society is concerned the effect is trivial. It remains at 9 years. It will rise the average schooling of the lowest group by one year. The cost to implement this regime is also calculated. The last three columns of the lower part of table X present the results. Nearly Rs. 570 millions is required for the implementation of this policy. This is 0.6 percent of the government revenu.

Table X: Estimation of Monetory compensation to Eliminate the Family Background Effect from Schooling Decision

Income Group	No. of Cases	Means of Continuous Variables				Predicted School Completion Frequencies %				
		Inco. (Rs.)	Birth Order	Mother's Ed. (Years)	Family Size	Grade 1 to 5	Grade 6 to 8	Grade 9 to 10	G.C.E. (O.L.)	G.C.E. (A.L.)
Lowest 25 %	2166	617	2.65	6.73	6.13	14.69	22.38	39.96	21.14	1.83
Second 25 %	1722	1756	2.34	7.78	6.09	10.93	19.57	40.73	25.96	2.81
Third 25 %	2613	2336	2.31	7.87	5.91	9.85	18.58	40.72	27.63	3.22
Top 25 %	2164	4326	2.15	9.53	5.73	5.16	12.98	38.22	37.09	6.55
Total	8665	2290	2.36	7.98	5.96					
Income Group	Est. of Number of Children[a]	Cost per-child (Regime A) (Rs.)	Total cost Required (Regime A) (Rs. thousands)	Mean Education Before Policy (years)	Mean Education After policy (Regime A)	Cost per-child (Regime B) (Rs.)	Total cost Required (Regime B) (Rs. thousands)	Mean Education After policy (Regime B)		
Lowest 25 %	1039320	1757	1,26485	8	10	547	568480	9		
Second 25 %	826274	1222	1009913	9	10	-	-	9		
Third 25 %	1249967	1096	1369529	9	10	-	-	9		
Top 25 %	1042199	-	-	10	10	-	-	10		
Total	4157760		4205927	9	10		268480	9		

Note: [(a)] Total number of pupils in four income classes are estimated by multiplying the total school population in 1993 by the relative frequency of pupils in our sample (ie. as in the first part of the column 2 in this table)

Note that the second column of the top part of table shows number of cases in each income percentile. By definition, frequency of each percentile must equal to each other (ie. 25 % of total number of respondents must be in each class). However, in the second percentile number is slightly higher than the number in third percentile. This is due to the fact that we have lot of cases in the third percentile for whom the parental income is imputed with mean income of the distribution.

Source: Selected Sub-sample from Quarterly Labour Force Survey- 1993: Census and Statistics, Sri Lanka

8. Summary, Conclusions and Recommendations

8.1 Summary and Conclusions

This paper developed and applied a human capital theoretic model to describe the enrolment decision in the post-compulsory education. The model assumes that the child with compulsory education designs a plan for his(her) optimal length of post-compulsory education. His(her) decision criterion is the optimisation of the present value of the life time income. We assume that the decision maker is sure of the ability to complete the stipulated length of education and labour market is in *steady-state.*

Using the same theoretical structure, we have looked at our data from two perspectives: the *enrolment decision* of age eligible children and the *length of schooling*. The school enrolment decision is modeled in terms of a probit enrolment probability model. We construct *a tobit-like ordered-probit model* to analyse the schooling length.

Then, we estimate four different versions of each model to see the sensitivity of coefficients for the changes in sample size and changes in variables. Although some coefficients are affected with those changes, we consistently observed in all the models that the choice of education is to a great extent related with the family characteristics. Children from more affluent families stay longer in school than their counterparts do. This has a very strong implication for the free education policy of Sri Lanka. Clearly, free education is insufficient to erase the family background effect on schooling participation. It seems that the children from more affluent families get more benefits from the free education than the poor children do. This means that by subsidising the education of the affluent children the free education system would worsen the income distribution pattern of the country.

According to our estimates, if government is prepared to spend additional 4,000 million Sri Lankan rupees to pay for the children belong to poor families, family background effect on schooling decision can easily be eliminated.

Apart form the policy issues, we also tested some theoretical notions regarding the schooling decision. Our empirical results support the hypothesis that the ability effect on the return to education is greater than the ability effect on the cost of education. This explains why more able people stay longer in full-time education. Further, our results reject the *Bahrman and Taubman's* hypothesis which argues that the birth order has a negative effect on ability. We have found a positive birth order effect.

Still a significant percentage of the total variation of the education is not explained by our model. Therefore, we analysed the residuals of the regressions. We estimate the correlation coefficient of the unexplained components of the children from the same family. We found that there is a family fixed effect which is not explained by our model.

Poor measures of ability, missing family characteristics and the excluded school quality variables would be the reason for this unexplained family fixed effect.

8.2 Recommendations

As we have highlighted several times in this paper, the conclusions of the paper are subject to a number of limitations. *First*, it is the data limitations. For example, due to the unavailability of information about school leavers at the time when they made the decision to leave school, we had to use the current information. This necessarily leads to the recommendation of a new survey which collects all this information.

In addition to the data limitations, our model assumes that the decision concerning the optimal length of schooling career is made under a perfectly certain situation. In reality, these assumptions over-simplify the complexity of the problem. Therefore, a theoretical structure which is able to deal with these aspects is called for.

In terms of policy analysis, our results show that the free education policy alone does not help the poor much. Therefore, assessing the free education policy in depth and the effectiveness of alternative *individual specific subsidy policies* such as provision of financial incentives to poor families in order to send their children to school for a longer period of time and more effective compulsory education policies are timely and important.

Bibliography

Altonji, J. G. (1993), The Demand for and Return to Education When Education Outcomes are Uncertain, *Journal of Labor Economics,* Vol. 11, pp. 48-83

Amemiya, T (1981) , Qualitative Response Models: A Survey , *Journal of Economic Literature*, vol.XIX , 1483-1536

Athurupana, H, (1993) Is Education Beneficial ? A Microeconomic Analysis of the Impact of Education on Economic Welfare in Sri Lanka, unpublished Ph.D. Thesis, Cambridge University

Becker G.S, (1967), Human Capital and the Personal Distribution of Income, *W.S. Woitinsky lecture no. 1*, University of Michigan, U.S.A.

Becker, W. E and P. E. Kennedy (1992), A Graphical Exposition of the Ordered Probit, *Econometric Theory*, 8, pp. 127-131

Behrman, J. R , P. Taubman (1986), Birth Order, Schooling, and Earnings, *Journal of Labour Economics*, Vol. 4, pp. s121-s145

Ben-Porath, Y (1967), The Production of Human Capital and the Life cycle of Earnings, *Journal of Political Economy*, 75, pp. 352-365

Card, D and A. Krueger (1992), Does Schooling Quality Matter ? Returns to Eduction and the Characteristics of Public Schools in the United States , *Journal of Political Economy*, 100 pp. 1-40

Card , D (1994), Schooling and Ability Revisited, *NBER Working Paper 4832*

Greene, W. (1993), *Econometric Analysis*, Macmillan Publishing Company

Griliches Zvi (1986), Comment on Behrman and Taubman, *Journal of Labour Economics*, vol. 4, no. 3 pp. s146-s150

Hartog, J, G. Pfann and G. Ridder (1989), (Non-) graduation and the Earnings Function: an inquiry on self-selection, *European Economic Review*, 33, pp. 1373-1395

Hartog, J. (1993), On Human Capital and Individual Capabilities, *Invited Lecture for the European Association of Labour Economists Meeting,* Maastricht

Judge, G . G, W.E. Griffiths, R. Carter Hill, Helmut Lutkepohl and Tsoung- Chao Lee (1984), *The Theory and Practice of Econometrics*, John Wiley and Sons publication

Kelly, T. (1993), *Does Vocational Education Enhance Competitivity in the Labour Market ? : Recent Evidence from Sri Lanka* (A Draft Discussion Paper)

Kodde, D. (1985), *Microeconomic Analysis of Demand for Education,* Unpublished Ph.D thesis, Erasmus University, Rotterdam, The Netherlands

Lazear, E. (1977) , Education: Consumption or Production ? *Journal of Political Economy* Vol. 85, no.3, pp. 569-597

Levhari, D and Yoram Weiss (1974), The Effect of Risk on the Investment in Human Capital, The *American Economic Review* Vol. 64, No. 6, pp. 950-963

Madala, G.S (1983), *Limited Dependent and Qualitative Variables in Econometrics,* Cambridge University Press, London

Mincer J, (1970), Distribution of Labor Incomes: A Survey With Special Reference to the Human Capital Approach, *Journal of Economic Literature* vol. 8, pp. 1-26

Oosterbeek, H. (1992), *Essays on Human Capital Theory* Unpublished Ph.D Thesis, University of Amsterdam

Oosterbeek, H.and H. V. Ophem (1995), *Human Capital Technology and Schooling Choices*, Discussion Paper Presented at the Tinbergen Institute, Amsterdam

Peoples' Bank *(1994,May/June) Economic Review*

Pissarides, C. (1981) , Staying-on at school in England and Wales, *Economica*, 48, pp. 3 45-363

Polachek, S.W and W. Stanley Siebert (1993), *The Economics of Earnings*, Cambridge University Press

Schultz, T. W (1989), Investing in People: Schooling in Low Income Countries, Economics of Education Review, vol. 8, pp. 219-223

Solon, G. (1992), Intergenerational Income Mobility in the United States, *The American Economic Review*, Vol. 82, pp. 393-408

University Grants Commission: Sri Lanka (various issues), *Statistical Hand Book*

Ven, W. P.M.M. Van De and Bernard M.S.Van Praag (1981), The Demand for Deuctibles in Private Health Insurance: A Probit Model with Sample Selection , *Journal of Econometrics*, 17 pp. 229-252

Wills R.J. and Rosen Sherwin (1979), Education and Self-Selection, *Journal of Political Economy*, Vol. 87, no. 5, pt. 2 pp. s7-s36

Zimmerman, D. J. (1992), Regression Toward Mediocrity in Economic Stature, *The American Economic Review*, Vol. 82, pp. 409-429