CEPS INSTEAD

WORKING PAPERS

Empirical analysis of school attainment/progression in Cameroon

Michel TENIKUE

Working Paper No 2010-45
December 2010

Empirical analysis of school attainment/progression in Cameroon*

Michel Tenikue
CEPS/INSTEAD, Luxembourg
University of Namur, Belgium

December 2010

Abstract

In Cameroon, only 1/3 of children progress to secondary education. This paper estimates a sequential model of school attainment to investigate the role played by family background and individual characteristics in keeping children at school up to the end of secondary school. Using data of the 2001 Cameroon Household survey, we find that while parental wealth has no effect on the probability to enter primary school. It is however a good predictor of completing primary and secondary education. The lack of schools supply reduces school progression, particularly the lack of secondary schools hinders primary school entry. Finally, we find that male children are more likely to stay at school up to the end of secondary education.

Keywords: Schooling; Sequential

*I would like to thank Vincenzo Verardi, Philippe Van Kerm, Bertrand Verheyden and Mathias Kuepie for useful discussions and comments.

1 Introduction

Having some primary education has become less problematic while getting some secondary education is still challenging in many developing countries. Indeed, in 2007, the gross school enrolment rate in primary school is 97% in Sub-Sahara Africa, while the gross school enrolment secondary school is only 33% in the same region. In Cameroon in particular , the gross enrolment rate in primary school is 3 time higher than gross enrolment rate in secondary school [1]. In this paper, we investigate the role played by family background and child characteristics on school progression from primary school entry to the end of secondary education.

Economic theory suggests that the decision to keep a child at school up to a given level should depend on costs and benefits. Costs depend both on the opportunity cost of child's schooling time and on the direct costs of schooling. Benefits include productivity improvement and creation of future earnings opportunities. However the relative burden of costs and particularly opportunity cost, for a given level of benefits, depends on family's characteristics. For instance, poor versus non poor parents, credit constrained versus non constrained parents (Basu, 1998, Baland & Robinson, 2000), allocate child's time differently.

Most existing research on school attainment and family background effects have used static framework and have focused on single educational-transition models (Willis & Rosen, 1979, Mare, 1980) or highest grade completed (Dreze & Kingdon, 2001, Birdsall, 1985, Psacharopoulos & Patrinos, 1997). Educational-transition models consider family background characteristics as determinant of transition probability from one stage of education to another. Models of highest grade attained/completed are ordinary least squared or ordered logit. The main limitation of these studies is that they ignore "educational selectivity". A transition model from secondary to post-secondary school does not account for previous transitions, say from primary to secondary school. Models of the number of completed years of education feature the same limitation. Further, they suppose that family factors play the same role at different stages of education. Yet, schooling is a cumulative and sequencing/dynamic process. Family factors may have different influence at different level of education.

This papers models school attainment as the outcome of four decisions taken sequentially

[1] In Cameroon, the gross enrolment rate in primary school is 110% and the gross enrolment rate in secondary school is 37% .

by the household. The first is whether to let a child start primary education. The second is made by households with children already in primary schools. Should they keep children enrolled until they complete primary education. The third decision is whether to enroll primary school graduates in secondary school. The fourth and last decision is whether to keep children enrolled in secondary school until they complete secondary education. Children who enter primary school have some primary education and children who enter secondary school have some secondary education. This approach accounts explicitly for "educational selectivity" as only primary school graduates can start secondary education (Cameron & Heckman, 1998). It has been used recently by Sawada & Lokshin (2009) to study obstacle to school progression in Pakistan and by Pal (2004) to study child schooling in Peru.

We use data provided by the 2001 Cameroon Household survey to estimate a sequential model of school attainment. This 4 stages sequential probit model is estimated by simulated maximum likelihood. This estimation approach allows some household and child characteristics to affect differently different schooling decisions. The first finding of the study emerges from descriptive statistics. The retention rate in education is decreasing. Conditional on completing primary education, the probability of a child to enter secondary school is lower than its unconditional probability to enter primary school. From the estimated model, we find four other striking results. First, parental wealth has not effect on primary school entry. It is however a good predictor of completing primary and secondary education. When household wealth increases, children have higher probability to complete primary education and to move up to the end of secondary education. Second, children from agricultural households are less likely to go through the schooling process. Even when they start their education process, their probability to move to higher level of education is lower. Third, female children and in particular those with an irregular school progression rhythm, are less likely to complete primary and secondary education. Conditional on primary school entry, they are also less likely to move to higher level of education. Fourth, while the supply of primary and secondary school respectively constitute barriers to primary and secondary school entry, the lack of secondary schools has an additional negative effect on the likelihood to enter primary school.

The paper proceeds as follows. In section 2, we describe data and present the econometric framework. Section 3 provides the main estimation results and their interpretation. In Section 4, we discuss the main hypothesis made in the paper and conclude.

2 Data and econometric framework

2.1 Data

We use data of the 2001 Living Standard Measurement Survey in Cameroon. The sample has 11,000 households. The data provide information on children's current and past schooling details and on a wide range of household level characteristics. From the data, we can identify the level of education of each child in 2001.

The cameroonian education system consists of four stages: primary, secondary, post secondary and university. Primary school has 6 grades[2], secondary school has 4 grades and post secondary school has 3. Children enter primary school when they are about six years old. Supposing no failure, children complete secondary education at the age of 16. Households decide whether children will attend and complete primary education, next whether children will attend and compete secondary education and the decisions on post secondary education and on university follows. These last two decisions are not considered in this paper.

In the database, the probability of being in the last stage of the schooling decision tree becomes strictly positive at the age of 15. We therefore focus on children aged 15 to 25 years. The upper bound of 25 is chosen because some children started primary school with a delay[3] or repeated grades throughout their educational process. For some children in this age group, the decision taken by the households is not completely realized, it is right-censored. This is the case, for instance, of a child who is still enrolled in secondary school and whose parents have decided that s/he will complete secondary education. Following Sawada & Lokshin (2009), the risk set for later stage decisions, which is the set of children available at that stage of the decision, consists of children for whom current stage decision is not right-censored. In section 4.1, we check that our results are robust to this assumption.

There are 8,470 children belonging to 5,130 households in our sample. Table (2) shows

[2]There are actually two sub-systems of education in Cameroon. Differences between the two sub systems are related to the language of tuition, program content and the number of grades in primary school. Primary school has 7 grades in the English subsystem and only 6 in the French sub-system. To make children from the two sub systems comparable, we follow the Cameroonian National Institute for Statistics and consider that the first grade of primary school in the English sub-system corresponds to nursery school. Thus overall, we consider primary school has 6 six grades in the country.

[3]Late enrolment is investigated by Bommier & Lambert (2000) in Tanzania and by Psacharopoulos & Patrinos (1997) in Peru

descriptive statistics of variables used. More than half of children in the sample (54%) are female.

Table (1) shows sample conditional probabilities to move from one stage of education to the next one. The first column is the average educational survival rate. The probability of entering primary school is 85% and is higher than the conditional probability of completing primary school and entering secondary school. This decreasing feature reveals the decreasing retention rate that characterizes cameroonian education system (INS, 2001). The last column of the table shows the sample size at each decision point. As expected, the number of observations decreases as we move forward in the decision tree.

Table 1: Conditional probabilities for educational survival

	Rate (%)			
	Total	Male	Female	Sample size
Enter primary school	84,8	89,5	80,8	8470
Complete primary education	77,2	78,3	79,1	7186
Enter secondary school	76,2	77,6	75,9	5546
Complete secondary education	67,8	72,1	63,5	4256

Total is the average survival rate while male and female are gender specific survival rates.

2.2 Econometric framework

Four stages of education are considered in this paper. Accordingly the household takes four decisions sequentially. We define D_s as the indicator variable corresponding to the decision taken at the sth educational stage. The process of schooling decisions can be described as follows. Children are born with no education. When they reach the schooling age (around 6 years old), the household takes the first decision: whether to educate or not. Children who enter the education system will have at least some primary education. The uneducated do not enter primary school and are characterized by $D_1 = 0$. Next, the household decides whether children who entered primary school will stay at school up to the end of primary education or not. Children who have entered primary school but drop out before the end are represented by $D_1 = 1$ and $D_2 = 0$. The household then decides whether primary school graduates will proceed to secondary school or not. Primary school graduates who quit the education system are represented by $D_1 = 1, D_2 = 1, D_3 = 0$. Finally, some primary school graduates who enter secondary school do not complete secondary school education ($D_1 = 1, D_2 = 1, D_3 = 1, D_4 = 0$) while others do ($D_1 = 1, D_2 = 1, D_3 = 1, D_4 = 1$). Children who complete secondary education get the secondary school certificate and can proceed to post secondary school[4]. The four decisions thus generate five possible outcomes than can be ranked in ascending order. We label these outcomes by O_1 to O_5 as shown in figure (1).

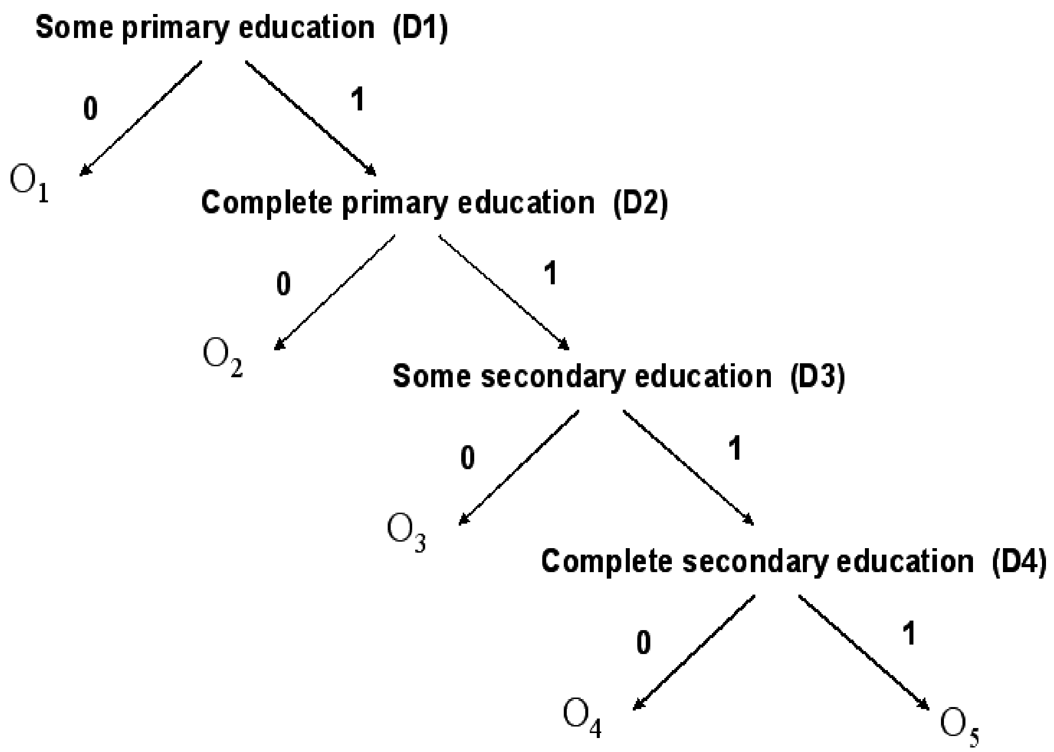

Figure 1: The sequential schooling decision tree

The outcomes are labelled O_1 to O_5. Children in the group O_1 have 0 year of education. They have never entered primary school. Children in the group O_2 have some primary education but have not completed primary education. Children in group O_3 are primary graduates who have no secondary education. Children in group O_4 have some secondary education but have not completed secondary education. Children in group O_5 are secondary school graduates.

We define the following index functions associated to the four decisions:

$$
\begin{aligned}
D^*_{is} = X_{is}\beta_s + u_{is}, \quad & D_{is} = 1 \text{ if } D^*_{is} > 0 \\
& D_{is} = 0 \text{ if } D^*_{is} \leq 0
\end{aligned}
$$

where index s= 1, 2, 3, 4 represents the decision and index i represents that child. The sets of covariates x_s include the child's and household's characteristics and also the distance to the nearest school. The vectors x_s are allowed to be different for different values of s. The terms u_s represent random unobserved variables and β_s are vectors of parameter to be estimated.

[4]This level of education is not considered here.

If we assume that household decision-making is independent across stage, or equivalently that u_{is} are independent for all i and across s, then the sets of parameters β_s will be estimated using four different simple discrete choice models. This possibility is however unrealistic. Education is a cumulative process. Success or failure in one stage of education may shape the decision made by the household at the next stage. Thus u_{is} are correlated across stages. Ignoring this correlation can create biases due to the selection rules at each stage.

We denote the joint probability density function of the error term by $f(u_{i1}, u_{i2}, u_{i3}, u_{i4})$. The probability to complete secondary education is defined by:

$$\begin{aligned} Pr(O_{i5}) &= Pr(D_{i1} > 0, D_{i2} > 0, D_{i3} > 0, D_{i4} > 0) \\ &= \int_{-X_{i1}\beta_1}^{\infty}\int_{-X_{i2}\beta_2}^{\infty}\int_{-X_{i3}\beta_3}^{\infty}\int_{-X_{i4}\beta_4}^{\infty} f(u_{i1}, u_{i2}, u_{i3}, u_{i4})du_1du_2du_3du_4 \qquad (1) \end{aligned}$$

The probability of being in an earlier stages of education can be defined similarly. It is an integral of lower dimension.

Schooling decisions are interlinked across stages and are subjected to selectivity. After the first decision, subsequent ones are relevant only for a selected group of children. The number of children is reduced for the last decision. If we denote by n_j the number of children with outcome O_j (j=1,...,5) and by N_s the number of children available at stage s (s=1,...,4), we have that: $N_4 = n_4 + n_5$ and $N_s = n_s + N_{s+1}$ for $s < 4$. The decision to enter primary school is taken for all children so that the total sample size is $N = N_1$.

The log-likelihood function of data is defined by:

$$L = \sum_{i=1}^{N} 1_{O_{i1}}log(P(O_{i1}))+1_{O_{i2}}log(P(O_{i2}))+1_{O_{i3}}log(P(O_{i3})+1_{O_{i4}}log(P(O_{i4}))+1_{O_{i5}}log(P(O_{i5})) \qquad (2)$$

where $1_{O_{ij}}$ is an indicator variable. It takes the value 1 ($1_{O_{ij}} = 1$) if the outcome of child i is O_j. The probability $P(O_{ij})$ is an integral of dimension j, j= 1,2,3,4.

Computing directly these integrals is numerically difficult. The likelihood maximization procedure is complex because integrals must be evaluated at each step of the maximization process. To simplify the computation procedure[5], we assume that the joint distribution of the four unobservable variables is multivariate normal and uses a simulated maximum likelihood

[5] There are alternative computing methods: (1)imposing a random effect variance component model with

approach. Multidimensional integrals are estimated by simulation and substituted in the log-likelihood function, which is then maximized with respect to the parameters of the model[6]. This approach is less computer intensive and straightforward. It has been emphasized recently by Cappellari & Jenkins (2006) for STATA users and its computing effectiveness is discussed in Waelbroeck (2005).

We therefore assume that $(u_{i1}, u_{i2}, u_{i3}, u_{i4})' \sim N(0, \Sigma)$ with

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_{21} & \rho_{31} & \rho_{31} \\ \rho_{21} & 1 & \rho_{32} & \rho_{42} \\ \rho_{31} & \rho_{32} & 1 & \rho_{43} \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 \end{pmatrix} \tag{3}$$

The diagonal elements of variance matrix Σ is a unity matrix. The model describe by equation (2) is numerically identified (GAO et al., 2001). The supply of schools which is stage specific reinforces the identification of the model. The probabilities of the outcomes are defined by:

$$\begin{aligned} P(O_{i1}) &= \Phi(-X_{i1}\beta_1) \\ P(O_{i2}) &= \Phi_2(X_{i1}\beta_1, -X_{i2}\beta_2 \mid \Sigma_2) \\ P(O_{i3}) &= \Phi_3(X_{i1}\beta_1, X_{i2}\beta_2, -X_{i3}\beta_3 \mid \Sigma_3) \\ P(O_{i4}) &= \Phi_4(X_{i1}\beta_1, X_{i2}\beta_2, X_{i3}\beta_3, -X_{i3}\beta_3 \mid \Sigma_4) \\ P(O_{i4}) &= \Phi_4(X_{i1}\beta_1, X_{i2}\beta_2, X_{i3}\beta_3, X_{i3}\beta_3 \mid \Sigma) \end{aligned} \tag{4}$$

where Φ is the cumulative univariate normal density and Φ_k are kth multivariate normal densities (k=2,3,4). The matrixes Σ_k, (k=2,3,4), are sub matrix of Σ where the term are multiplied by appropriate weights[7].

---

up to four replications (Lillard & Willis, 1994, Pal, 2004); (2) imposing a common factor error structure and approximating non-parametrically the joint distribution by a step function (Cameron & Heckman, 1998); (3) using the "Full Information Maximum Likelihood", based on the Gauss-Hermite quadrature, and representing the joint distribution of the four unobservable variables as a weighted sum of the products of univariate distributions (Sawada & Lokshin, 2009)

[6]The multivariate normal probabilities are calculated with the Geweke-Hajivassiliou-Keane (GHK) smooth recursive simulator that is more accurate (Greene, 2003).

[7]The terms of matrix Σ_k are defined by $\Sigma_{k_{ij}} = \Sigma_{ij} * (2Y_i - 1) * (2Y_k - 1)$.

This model is not linear and estimated parameters can not be interpreted as marginal effects. However marginal effects and estimated parameters have the same sign.

2.3 Variables

We are interested in the effects of child and household characteristics on a set of decisions taken sequentially. Table (2) summarizes the descriptive statistics of all control variables used in the sequential model described in section (2.2).

Child characteristics include the gender, age and the relationship to household's head. The gender is a dummy variable that takes the value 1 for male and 0 for female. It indicates that the share of male children is smaller at primary school entry and is increasing throughout the decision tree. By the end of secondary school, the sample is split equally among male and female children. The average age in the sample is 20.5 with a standard deviation of 3. Due to their age, some children are less likely to be at a given decision point. For example, only 22% of children aged 15 reach the fourth decision point while 47% of children aged 20 are considered for the decision to complete secondary school. This suggests the presence of age effects in the decision process. We include 4 age groups represented by a set of 4 dummy variables. The nodes used are the first three quartiles of the sample distribution age, namely 18, 20 and 23. The last child specific characteristic is whether the child is the son or the daughter of the household's head. Many studies have suggested that the type of relationship with household's head affects schooling decisions[8]. About 40% of the sample is made of sons or daughters of the head of the household were they live.

The set of household characteristics includes the level of education household head and the occupation of the household head or of his/her spouse. Four levels of parental education are considered. They are measured with 4 dummy variables: No education, primary, secondary and university. Twenty four percent of children are from household where the head has no education. Regarding occupation, we distinguish whether the main source of household's income is from wage or from self-employment. The variable included takes the value 1 if the head of the household is self-employed and 0 if s/he is a wage worker. We also distinguish if s/he is self-employed in agriculture or if s/he runs a business. While these variables indicate

[8]Case et al. (2004) and somehow Fafchamps & Wahba (2006) are two examples.

the social group, they also characterize the sources of household's income. Most wage workers are civil servants whose income is more stable compared to self-employed.

The set of household characteristics also includes an indicator of household's wealth. The wealth proxy is the predicted value of expenditure par capita. This variable reflects the permanent component of household living standard. Finally, the households size is included in the form of number of household member per age group. We consider separately the number of females and males of children in the age range of our sample[9]

The last set of variables contains details on school supply. Availability of primary and secondary school are considered separately. Also the supply of public school with lower school fees and private schools with higher school fees. On average, a public primary school is situated within 1.1 km to each child and a public secondary school is situated within 4 km. Private school are situated further away. The last variable indicates whether the household lives in a rural area or not. The dummy variable takes the value 1 if the household is in the rural area.

[9]The household size may be considered as endogenous in a household decision model(Baland & Robinson, 2000).

3 Estimation results of the sequential schooling decision model

Table (3) summarizes the results of the estimated 4-stage sequential schooling decisions model. This estimates are derived from the simulated maximum likelihood estimation of the model described by Eqs. (2) and (4). The last row of the table shows a test of the hypothesis that the matrix Σ defined by Eq. (3) is an identity matrix. We reject the null hypothesis and conclude that household decision-making is indeed related across the four stages considered. Ignoring the correlation structure of the unobserved error terms would give rise to inconsistent estimates[10].

Gender: The gender coefficient is positive and significant at all stages of education. Male children are more likely to go through the four levels of education. However the gender difference seems to be age-grade specific. Table (4) suggests that there is no gender difference among children aged 15 to 17 who have completed primary education. This is consistent with the fact that the gender difference is not really important among children with a regular school progression rhythm(TENIKUE, 2010).

The wealth indicator: The coefficient on the wealth indicator it not significant at primary school entry. It is however positive and significant for the 3 other decision points considered. This result suggests that poor households can easily send their children to primary school, but they encounter difficulties when it comes to keep children at school up to the end of primary education or to move to secondary school. This result is consistent with the hypothesis of increasing direct or indirect cost of education(Emerson & Souza, 2008). It provides a reason why we may observe high gross primary school enrolment rate and lower gross secondary school enrolment rate.

Parental education: Variables reflecting household's head education have positive and significant coefficients. This demonstrates the importance of parental education on the education of children. Educated parents value education and are more able of perceiving its benefits. Parental education may also affect child's motivation at school. The results show that, if the household's head has some primary education, it enhances the likelihood to enter and to complete primary education but it has no effect on the probability of entering secondary school.

[10]Table (5) in appendix shows the results of the independent error terms specification. This estimates are believed to be biased.

To affect significantly the likelihood to complete secondary education, the household's head needs to have a university level.

Relatedness to household's head: The variable measuring the relatedness of the child to the household's head is positive and significant at all schooling stages. Being the son or the daughter of the household's head is important for education. This results is consistent with the Hamilton's Rule[11] and has been observed in different parts of Africa (Case et al., 2004).

Occupation of the household's head: The variable self-employment is not significant at any stage of education. Whether the household relies on wage income or not seems not really important for schooling. However, when the main source of household income is self-employed in agriculture, it constitutes a threat to schooling. When the household head or her/his spouse is self-employed in agriculture, it reduces the likelihood to enter primary school. Moreover it affects negatively the probability to move toward higher educational levels. They are two possible explanations for this negative correlation. First, agricultural households may face higher opportunity cost of education due to the within household demand of child labor (Bhalotra & Heady, 2003). Second, agricultural households may be more exposed to income shock (Cogneau & Jedwab, 2008).

The supply of schools: The supply of primary and secondary schools have a significant effect on primary and secondary education. The further the distance to the school, the lower the likelihood to enter or to complete the corresponding schooling stage. The results show that the absence of a public secondary school in the village reduces the incentive to send children in primary school. This is meaningful in an environment where private return to education might be flat for children with some primary education and increases as children start secondary education (KUEPIE et al., 2006).

4 Discussion and conclusion

4.1 Discussion

Following Sawada & Lokshin (2009) we exclude children with right-censored realization of schooling decisions from our sample. Excluded observations are typically children aged 15 to

[11]The theory that the closeness of biological ties governs altruistic behavior.

25 enrolled in secondary school (15%). Taking explicitly right-censored into account requires to add one more equation to our econometric model. Under this possibility, the model becomes difficult to handle numerically. As a robustness check, we include these children in the sample under three alternative assumptions. The first is that all children still enrolled in secondary school will not complete secondary education. The second is that all children still enrolled in secondary school will complete secondary education. The third and last is that only 68% (the sample retention rate) of them will complete secondary education. We re-estimate the sequential model under the three scenarios and our results remain qualitatively unchanged.

The five possible outcomes generated by the four decisions can be ranked in ascending order. The ordered probit model can be seen as an alternative to the sequential probit model used in this paper. Unfortunately, unlike the sequential model, the ordered probit model assumes that a child can have secondary education with no primary education. Moreover, it assumes that household and child characteristics play the same role on the likelihood to be in any stage of education described.

Our data come from a cross section survey although schooling decisions are taken at different point in time. There are therefore some measurement problems as all our observations relate to the year of the survey and not when the decisions were made. We assume that the values observed during the survey reflect the values of these variables when schooling decisions were made.

4.2 Conclusion

The paper investigates the role played by family background and child characteristics on school progression from primary school entry to the end of secondary education. It models school attainment as the outcome of four decisions taken sequentially by the household. The first is whether to enroll a child in a primary school. The second decision is whether to keep children enrolled until they complete primary education. The third is whether to enroll primary school graduates in secondary schools. The fourth and last decision is whether to keep children enrolled in secondary schools until they complete secondary education. The four decisions are modeled with a 4 stages sequential probit model. This model accounts explicitly for the "educational selectivity" and is estimated by a simulated maximum likelihood method.

The first finding of this study emerges form descriptive statistics. The retention rate in education is decreasing. Conditional on completing primary education, the probability that a child enters a secondary schools is lower than its unconditional probability to enter a primary school. This is another facet of the difference observed in gross enrolment rate in primary school and gross enrolment in secondary school.

From the estimated model, we find four other striking results. First, parental wealth has not effect on primary school entry. It is however a good predictor of completing primary and secondary education. When households' wealth increases, children have higher probability to complete primary education and to move up to the end of secondary education. Second, children from agricultural household are less likely to go through the schooling process. Even when they start their education process, their probability to move to higher level of education is lower. Third, the household's head education reveals some threshold effect. While primary education enhances the probability to enter primary school, having some university education is needed to enhance the probability to complete secondary education. Fourth, while the supply of primary and secondary schools constitute barriers to respectively primary and secondary school entry, the lack of a secondary school has an additional negative effect on the likelihood to enter primary school.

Our results are richer than most existing ones. They show how some schooling determinants affect differently different levels of education. Policies to enhance education should then consider some level specific interventions.

References

Baland, J.-M. & Robinson, J. A. (2000). Is child labor inefficient? *Journal of Political Economy*, 108(4), 663–679.

Basu, K. & Van, P. H. (1998). The economics of child labor. *The American Economic Review*, 88 (3), 412–427.

Bhalotra, S. & Heady, C. (2003). Child farm labor: The wealth paradox. *World Bank Econmic Review*, 17(2), 197–227.

Birdsall, N. (1985). Public inputs and child schooling in brazil. *Journal of Development Economics*, 18(1), 67–86.

Bommier, A. & Lambert, S. (2000). Education demand and age at school enrollment in tanzania. *The Journal of Human Resources*, 35(1), 177–203.

Cameron, S. V. & Heckman, J. J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy*, 106(2), 262–333.

Cappellari, L. & Jenkins, S. P. (2006). Calculation of multivariate normal probabilities by simulation, with applications to maximum simulated likelihood estimation. *Stata Journal*, 6(2), 156–189.

Case, A., Paxson, C., & Ableidinger, J. (2004). Orphans in africa: Parental death, poverty and school enrollment. *Demography*, 41 (3), 483–508.

Cogneau, D. & Jedwab, R. (2008). *Family income and child outcomes: the 1990 cocoa price shock in Cote d'Ivoire*. Working Papers DT/2008/05, DIAL (Developpement, Institutions & Analyses de Long terme).

Dreze, J. & Kingdon, G. G. (2001). School participation in rural india. *Review of Development Economics*, 5(1), 1–24.

Emerson, P. M. & Souza, A. P. (2008). Birth order, child labor, and school attendance in brazil. *World Development*, 36(9), 1647–1664.

Fafchamps, M. & Wahba, J. (2006). Child labour, urban proximity, and household composition. *Journal of Development Economics*, 79 (2), 374–397.

GAO, C., LAHIRI, K., & WIXON, B. (2001). *Value of sample separation information in a sequential probit model:Another Look at SSA's Disability Determination Process*. mimeo, Albany-SUNY.

Greene, W. H. (2003). *Econometric Analysis*. London: Prentice-Hall Internatinal.

INS (2001). *Annuaire Statistique du Cameroon*. Rapport, Institut National de la Statistique.

KUEPIE, M., NORDMAN, C. J., & ROUBAUD, F. (2006). *Education and Labour Market Outcomes in Sub-Saharan West Africa*. DOCUMENT DE TRAVAIL DT/2006-16, Développement, Institution & Analyse de Long terme (DIAL).

Lillard, L. A. & Willis, R. J. (1994). Intergenerational educational mobility: Effects and state in malaysia. *Journal of Human Resource*, 29(4), 1126–1166.

Mare, R. D. (1980). Social background ans school continuation decisions. *Journal of the American Statistical Association*, 75(370), 295–305.

Pal, S. (2004). Child schooling in peru: Evidence from a sequential analysis of school progression. *Journal of Population Economics*, 17(4), 657–680.

Psacharopoulos, G. & Patrinos, H. A. (1997). Family size, schooling and child labor in peru - an empirical analysis. *Journal of Population Economics*, 10(4), 387–405.

Sawada, Y. & Lokshin, M. (2009). Obstacles to school progression in rural pakistan: An analysis of gender and sibling rivalry using field survey data. *Journal of Development Economics*, 88(2), 335–347.

TENIKUE, M. (2010). *Gender Gap in Current School Enrolment in Cameroon: Selection Among "Irregular" Children?* CEPS/INSTEAD Working Paper Series 2010-03, CEPS/INSTEAD.

Waelbroeck, P. (2005). Computational issues in the sequential probit model: A monte carlo study. *Computational Economics*, 26(2), 141–161.

Willis, R. J. & Rosen, S. (1979). Education and self-selection. *Journal of Political Economy*, 87(5), S7–36.

Table 2: Descriptive statistics: means standard deviations of variables

	Some primary		Complete primary		Some secondary		Complete secondary	
	mean	sd	mean	sd	mean	sd	mean	sd
Sex (Male=1)	0.46		0.49		0.50		0.50	
Dummy = 1 if age between 15 and 17	0.18		0.16		0.14		0.13	
Dummy = 1 if age between 18 and 20	0.33		0.33		0.33		0.33	
Dummy = 1 if age between 21 and 22	0.20		0.22		0.22		0.22	
Dummy = 1 if age between 23 and 25	0.29		0.29		0.30		0.31	
Son/daughter of the household head	0.41		0.43		0.44		0.46	
Head has primary level	0.34		0.35		0.31		0.26	
Head has secondary level	0.34		0.40		0.45		0.51	
Head has university level	0.08		0.10		0.12		0.14	
Head or spouse non wage worker	0.63		0.60		0.55		0.51	
Head or spouse Self-employed in agri	0.19		0.15		0.10		0.07	
Head or spouse Self-employed in business	0.22		0.22		0.22		0.22	
Estimated expenditure per capita	3.73	3.18	4.05	3.20	4.51	3.19	4.89	3.21
Rural	0.26		0.21		0.16		0.12	
Distance to private primary school	2.26	4.80	1.77	4.15	1.35	3.46	1.19	3.13
Distance to public primary school	1.10	2.07	0.89	1.27	0.81	0.91	0.77	0.89
Distance to private secondary school	3.91	6.09	3.33	5.47	2.75	4.81	2.39	4.39
Distance to public secondary school	3.82	5.27	3.07	4.42	2.58	3.77	2.23	3.23
# of 0-9 years	1.68	1.83	1.55	1.70	1.44	1.60	1.39	1.56
# of male of 10-14 years	0.43	0.74	0.43	0.73	0.42	0.73	0.43	0.73
# of female of 10-14 years	0.43	0.76	0.44	0.76	0.43	0.74	0.44	0.75
# of male of 15-25 years	1.36	1.35	1.43	1.37	1.46	1.38	1.52	1.41
# of female of 15-25 years	1.31	1.12	1.32	1.13	1.34	1.14	1.39	1.16
# of male of 26-35 years	0.36	0.62	0.36	0.62	0.37	0.64	0.37	0.65
# of female of 26-35 years	0.29	0.57	0.28	0.55	0.29	0.55	0.29	0.54
# of over 36 years	1.16	1.08	1.18	1.09	1.18	1.07	1.22	1.07
Number of observations	8470		7186		5546		4256	

The mean of a dummy variable is a proportion. Estimated expenditure per capita is in 100.000CFA.

Table 3: Estimation results of the sequential schooling decision model

	Primary	Compl. primary	Secondary	Compl. secondary
Gender (Dummy = 1 if Male)	0.355	0.091	0.098	0.267
	(6.31)***	(1.90)*	(2.02)**	(4.87)***
Dummy = 1 if age between 15 and 17		-0.308	-0.357	0.378
		(5.77)***	(6.22)***	(4.00)***
Dummy = 1 if age between 18 and 20	0.297			0.211
	(5.21)***			(3.57)***
Dummy = 1 if age between 21 and 22	0.584	0.122	0.152	
	(8.39)***	(2.39)**	(2.94)***	
Dummy = 1 if age between 23 and 25	0.457	0.212	0.239	0.088
	(7.33)***	(4.46)***	(4.88)***	(1.54)
Son/daughter of the household head	0.681	0.394	0.607	0.322
	(12.68)***	(7.61)***	(12.64)***	(4.13)***
Head has primary level	1.200	0.275	0.105	0.009
	(24.09)***	(3.14)***	(1.22)	(0.08)
Head has secondary level	1.911	0.912	1.004	0.168
	(26.05)***	(8.15)***	(10.02)***	(1.07)
Head has university level	1.858	1.056	0.987	0.743
	(12.75)***	(7.46)***	(7.74)***	(4.28)***
Head or spouse Non wage worker	0.007	-0.080	-0.047	-0.050
	(0.12)	(1.63)	(0.95)	(0.92)
Head or spouse Self-employed in agri	-0.238	-0.356	-0.353	-0.288
	(3.78)***	(6.05)***	(5.10)***	(2.74)***
Head or spouse Self-employed in business	-0.052	-0.090	0.005	-0.067
	(0.86)	(1.79)*	(0.09)	(1.11)
Estimated expenditure per capita	0.013	0.068	0.093	0.069
	(0.96)	(6.84)***	(9.49)***	(5.22)***
Rural	0.152	0.041	-0.199	0.098
	(2.05)**	(0.62)	(2.68)***	(1.05)
Distance to private primary school	-0.029	-0.027		
	(5.79)***	(4.73)***		
Distance to public primary school	-0.048	-0.054		
	(4.85)***	(3.49)***		
Distance to private secondary school	0.007	0.001	0.010	-0.011
	(1.64)	(0.15)	(1.62)	(1.69)*
Distance to public secondary school	-0.029	-0.004	-0.027	-0.027
	(5.38)***	(0.70)	(4.27)***	(2.88)***
# of 0-9 years	-0.111	-0.057	-0.035	-0.075
	(7.47)***	(3.95)***	(2.26)**	(4.31)***
# of male of 10-14 years	-0.035	0.016	0.033	0.045
	(1.12)	(0.60)	(1.17)	(1.43)
# of female of 10-14 years	0.014	-0.017	0.035	0.132
	(0.44)	(0.64)	(1.20)	(4.03)***
# of male of 15-25 years	0.009	0.017	0.053	0.050
	(0.42)	(1.00)	(3.05)***	(2.59)***
# of female of 15-25 years	0.063	0.099	0.126	0.094
	(2.67)***	(4.96)***	(6.00)***	(3.77)***
# of male of 26-35 years	-0.030	-0.013	-0.047	0.017
	(0.82)	(0.41)	(1.52)	(0.50)
# of female of 26-35 years	0.102	0.158	0.049	0.072
	(2.59)***	(4.30)***	(1.23)	(1.69)*
# of over 36 years	0.101	0.045	0.075	0.067
	(4.16)***	(2.07)**	(3.33)***	(2.53)**
Constant	-0.376	-0.160	-1.007	-0.545
	(3.53)***	(1.12)	(6.14)***	(1.93)*
Number of observations	8470			
Test: H_0: $\Sigma = I$	$\chi^2(6)$= 47.7		p= 0.000	

Primary corresponds to "have some primary education". Compl. stands for Complete. Secondary corresponds to "have some secondary education". Absolute value of z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: Estimation results: gender effect

	Some primary	Complete primary	Some secondary	Complete secondary
Age 15-17 * Female	-0.247	-0.230	-0.002	-0.064
	(3.20)***	(2.99)***	(0.02)	(0.52)
Age 18-20 * Female	-0.035	0.088	0.141	-0.390
	(0.49)	(1.29)	(1.83)*	(3.95)***
Age 21-22 * Female	0.212	0.299	0.263	-0.644
	(2.59)***	(3.97)***	(3.01)***	(5.91)***
Age 23-25 * Female	-0.006	0.295	0.254	-0.572
	(0.08)	(4.00)***	(3.04)***	(5.41)***
Age 18-20 * Male	0.280	0.215	0.280	-0.160
	(3.60)***	(3.03)***	(3.46)***	(1.49)
Age 21-22 * Male	0.562	0.285	0.325	-0.385
	(6.00)***	(3.48)***	(3.64)***	(3.41)***
Age 23-25 * Male	0.385	0.386	0.409	-0.292
	(4.78)***	(5.25)***	(4.73)***	(2.52)**
Distance to private primary school	-0.034	-0.032		
	(7.82)***	(5.51)***		
Distance to public primary school	-0.068	-0.083		
	(7.54)***	(6.04)***		
Distance to private secondary school	0.003	-0.014	-0.012	-0.022
	(0.71)	(3.02)***	(2.10)**	(3.05)***
Distance to public secondary school	-0.049	-0.033	-0.055	-0.036
	(12.40)***	(5.43)***	(6.93)***	(2.45)**
Constant	1.335	0.713	0.536	1.043
	(21.63)***	(9.09)***	(4.47)***	(5.85)***
Number of observations		8470		

The reference group to interpret the age/gender dummy variables is AGE 15-17 * Male. The model is non linear and coefficients in this table are not marginal effects. The sign of the coefficient on the age/gender dummy variables gives the sign of the difference between the marginal of the variable and the marginal effect of being in the reference group. Primary corresponds to "have some primary education". Compl. stands for Complete. Secondary corresponds to "have some secondary education". Absolute value of z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: Results under the assumption of independent error terms

	Primary	Comp. Primary	Secondary	Compl. Secondary
Sex (Male=1)	0.369	0.072	0.049	0.252
	(6.57)***	(1.57)	(0.95)	(4.75)***
Dummy = 1 if age between 15 and 17		-0.302	-0.264	0.396
		(5.80)***	(4.28)***	(5.09)***
Dummy = 1 if age between 18 and 20	0.298			0.228
	(5.21)***			(3.97)***
Dummy = 1 if age between 21 and 22	0.588	0.102	0.098	
	(8.43)***	(2.06)**	(1.77)*	
Dummy = 1 if age between 23 and 25	0.461	0.191	0.179	0.093
	(7.42)***	(4.06)***	(3.45)***	(1.64)
Son/daughter of the household head	0.643	0.363	0.502	0.332
	(12.19)***	(8.30)***	(10.11)***	(6.31)***
Head has primary level	1.195	0.208	-0.157	-0.110
	(23.97)***	(4.16)***	(2.45)**	(1.41)
Head has secondary level	1.908	0.825	0.595	0.102
	(26.15)***	(13.98)***	(8.48)***	(1.30)
Head has university level	1.859	0.959	0.567	0.697
	(12.88)***	(9.00)***	(5.41)***	(6.19)***
Head or spouse Non wage worker	0.006	-0.084	-0.029	-0.069
	(0.09)	(1.71)*	(0.55)	(1.25)
Head or spouse Self-employed in agri	-0.229	-0.344	-0.253	-0.331
	(3.67)***	(5.89)***	(3.40)***	(3.54)***
Head or spouse Self-employed in business	-0.042	-0.083	0.034	-0.065
	(0.70)	(1.65)*	(0.58)	(1.05)
Estimated expenditure per capita	0.010	0.069	0.080	0.068
	(0.80)	(7.08)***	(7.82)***	(6.69)***
Rural	0.128	0.030	-0.261	0.051
	(1.74)*	(0.46)	(3.40)***	(0.55)
Distance to private primary school	-0.029	-0.025		
	(5.80)***	(4.45)***		
Distance to public primary school	-0.050	-0.060		
	(5.19)***	(3.95)***		
Distance to private secondary school	0.007	-0.000	0.016	-0.010
	(1.62)	(0.03)	(2.59)***	(1.46)
Distance to public secondary school	-0.027	-0.000	-0.022	-0.026
	(5.08)***	(0.07)	(3.11)***	(2.98)***
# of 0-9 years	-0.116	-0.052	-0.002	-0.063
	(7.97)***	(3.89)***	(0.14)	(3.79)***
# of male of 15-25 years	0.004	0.021	0.058	0.061
	(0.17)	(1.24)	(3.11)***	(3.17)***
# of female of 15-25 years	0.066	0.097	0.105	0.101
	(2.81)***	(4.88)***	(4.64)***	(4.48)***
# of male of 26-35 years	-0.031	-0.008	-0.050	0.014
	(0.85)	(0.27)	(1.53)	(0.42)
# of female of 26-35 years	0.111	0.152	-0.001	0.088
	(2.77)***	(4.14)***	(0.03)	(2.06)**
# of over 36 years	0.111	0.046	0.070	0.098
	(4.62)***	(2.24)**	(3.00)***	(3.84)***
Constant	-0.372	-0.057	-0.276	-0.501
	(3.49)***	(0.64)	(2.71)***	(4.46)***
N	8470	7186	5546	4256

Primary corresponds to "have some primary education". Compl. stands for Complete. Secondary corresponds to "have some secondary education". Absolute value of z statistics in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

CEPS
INSTEAD

B.P. 48
L-4501 Differdange
Tél.: +352 58.58.55-801
www.ceps.lu