Measuring educational inequalities:
A method and an application to Albania [#]

Nathalie Picard [*]   **François-Charles Wolff** [**]

June 2007

Abstract: In this paper, we investigate whether educational inequalities stem rather from differences between families or within families. In a poor economy, schooling is costly for parents and education is likely to be unequally distributed among siblings. Drawing on discrete ordered choice models, we present a simple method to estimate the between and within components of both the explained and unexplained variances of education. For our empirical analysis, we use the LSMS survey conducted in 2002 in Albania. We explain about 40% of the total variance and find that inequalities in education are mainly due to differences between families. Differences within family are lower and far less easily explained.

Keywords: Education, intra-household inequality, random effects ordered Probit models, siblings

JEL classification: D13, I2, 015

[*] THEMA, Université de Cergy-Pontoise, France and INED, Paris, France.
E-mail: nathalie.picard@u-cergy.fr
[**] Corresponding author. LEN, Université de Nantes, BP 52231 Chemin de la Censive du Tertre, 44322 Nantes Cedex 3, France; CNAV and INED, Paris, France. Tel: 33240141779. Fax: 33240141743.
E-mail: wolff@sc-eco.univ-nantes.fr Homepage: http://www.sc-eco.univ-nantes.fr/~fcwolff

1/ Introduction

One of the most influential factors behind economic and demographic changes is education. Investments in human capital play a prominent role not only in developed countries, but also in developing countries. Numerous studies have empirically shown that the level of parental education is a very significant variable when explaining children's education. Interestingly, the fact that observed as well as unobserved components of family background strongly influence the outcomes for children over the life-cycle has been mainly considered from an intergenerational perspective. For instance, in numerous societies, it has been evidenced that intergenerational correlations in schooling attainment are most often comprised between 0.3 and 0.4 (see Altonji and Dunn, 1996).

Following Becker and Tomes (1986), economists have mainly focused on this issue of intergenerational mobility of education. The different models of household behaviour view educational attainment as the result of family decisions in which parental resources and children's outcomes are linked through investments in human capital.[1] Despite the well established relationship between parental and offspring education, it remains difficult to grasp the underlying mechanism. More educated parents have certainly more ability than less educated parents and these abilities are likely to be transmitted through genes or culture. But at the same time, parents with more abilities also generate more income. Another channel to understand the children's educational attainment deals with intragenerational mobility of education. Even though the same-gender sibling correlations in schooling are about 0.5 in the United States, the allocation of resources within the family often leads to unequal outcomes for the different family members.

Knowing how resources are divided within the family is further investigated in the Handbook's chapter of Behrman (1997), who provides empirical evidence on the issue of intra-household allocations. Among the most important explanatory variables for such allocations are endowments, preferences, investment prices, household resources levels, labour market opportunities and marriage markets. From an empirical perspective, problems of intra-household allocation are found in several domains including mortality, health, food, time allocation or education. Clearly, these issues are highly important from the public policy point of view since it may affect the effectiveness of redistributional policies. When inequalities within families are large, public policies transfer programs giving the same amount of money to the different members of the household are partly inefficient.

---

[1] Behrman *et alii* (1995) further examine how parents allocate human capital among their children. In a pioneering work, Sheshinski and Weiss (1982) provide a theoretical analysis of inequality within and between families.

Several empirical studies, which are further described in Behrman (1997), have suggested that there exist fairly large within-household variances. Education is certainly the most important area where it matters to measure and to understand the problems of intra-household allocations, given the long-term impact of schooling on labour markets. Intragenerational differences in education may stem from endowment differences, but they may also be explained by the size of the sibship, the birth order of children as well as the gender composition among siblings. While numerous authors have focused on the link between endowments and the allocation of schooling within the family for the US case (in particular Behrman *et alii*, 1995), there seems to be less evidence concerning the magnitude of inequalities in education in developing countries. At first sight, this is puzzling since education is more costly for parents in poorer economies, which is likely to increase the within-sibship variance.

A difficulty with the focus on intra-household allocation is the need of accurate data. As emphasized in Behrman (1997, p.127), "descriptions of the extent of variance in intra-household allocation from systematic socio-economic data sets are limited because many data sets take the household as the unit of observation for most of the information that they collect". Researchers being interested in the measurement of educational attainment of siblings are most often constrained by the lack of information on those older children who are no longer living with their parents. There are also other difficulties with the educational dependent variable. On the one hand, some children are never enrolled in school, especially when the parents have very limited resources. On the other hand, it is often difficult to know the exact number of years of schooling for children. Indeed, educational outcomes are often given by a discrete level of attainment, ranging from the decision to attend elementary school to the decision to continue beyond high school.[2] Thus, the lack of appropriate data and methodological difficulties help understanding why measures of the within-household variance of children's education remain so scarce in developing countries.

This is undoubtedly problematic since the link between economic development and the intra-household allocation of education has been recently evidenced. Yang and Zhu (2003) note that there may exist differences in the structures of production and organisation between the rural and urban sectors. The dominant form of organisation is family farms in the agricultural sector, so that allocative efficiency does not require that each family member benefits from a high level of education. Conversely, in the industrial sector, individuals are involved in paid activities,

so that there is a positive return to the schooling of each family member. In addition, structural changes are such that activities progressively transform from agricultural to industrial sectors, and parents have differentiated incentives to invest unequally in their children's education.[3]

Our purpose in this paper is to further investigate the intra-household distribution of education between siblings. Our primary aim is to give an accurate measure of the within-sibship variance, in order to know how important is the understanding of intragenerational mobility of education. This raises several methodological issues. Specifically, we rely on discrete ordered models with random effects to estimate the between and within components of both the explained and unexplained variances. We also discuss the problem of censoring for the education variable, and we choose to base the selection of the sample on the child's age. By including both parents' and children's characteristics into the regressions, we seek to better understand the factors that may explain why parents make unequal investments in the human capital of their different children.

For our empirical analysis, we use a cross sectional data collected in 2002 in Albania by the Worldbank. In that poor country, we show that completing more than the primary school level is a very discriminating variable between children. For our purpose, we select families whose children are all aged more than 16, so that each child had the opportunity to complete primary school. Then, we estimate random effects binary Probit models to perfom variance decomposition. We also assess the robustness of our binary variable approach by estimating random effects ordered Probit models. Our main result is that much of the explained variance stems from inequalities between families.

The remainder of the paper is organized as follows. In section 2, we briefly review the previous literature concerning the different hypotheses that may explain inequalities in education within families. Section 3 provides a description of the Albanian data set, with descriptive statistics on educational outcomes in that country. Section 4, which is the core of the paper, includes an econometric analysis based on random effects ordered choice models. We discuss the different factors that play a role when explaining education in Albania and we estimate the magnitude of the between and within components respectively of the explained an unexplained variances. Concluding comments are in section 5.

---

[2] It is also possible that there is some measurement error in the years of schooling variable. A simple way to check this is by seeing that the degree awarded is consistent with the level of schooling attained. For evidence on measurement error for years of schooling, see Ashenfelter and Krueger (1994).

[3] A more equal distribution for that variable is expected over time, which is found in the context of Taiwan.

2/ Previous literature

In this review of the previous literature, we attempt to show the different factors which have been proved to influence the intra-household distribution of education within the family, with a special focus on the differences between developed and developing countries. Numerous studies have attempted to provide explanations of why there may exist such differences in education among children, but very few studies have shed light on the magnitude of these differences. For instance, Ejrnaes and Pörtner (2004) indicate that differences among families measured through a simple variance analysis accounts for about half of the total variation in completed education between children.

Yang and Zhu (2003) document the changes in sibling educational structure over time in Taiwan. Their two main results are that educational inequalities decline rapidly during the course of economic development and that differences in siblings' education among rural families tend to be much higher than those for urban families.

A potential determinant of intra-household allocation concerns the birth order of children. Following findings of psychologists and sociologists, it has been argued that birth order influences children's innate abilities and endowments. When capital and labour markets are complete and perfect, basic models of human capital imply that parents educate their children up to the point at which the expected marginal return equals the marginal cost. In this context, if innate abilities and education are complements rather than substitutes for producing human capital, investments should be more profitable and therefore higher for first-born children.

This prediction seems to be supported by empirical evidence, at least for developed countries. For instance, in the US, Behrman and Taubman (1986) find that the average education significantly declines as one moves from first-born to fourth or later-born children even after controlling for family background and family size. However, poverty and capital constraints may change the conclusions, especially in developing countries. Poverty there often leads parents to send their children to the labour market (Basu, 1999). The higher innate abilities of earlier born children make their work more profitable and generate additional resources helping parents to alleviate the poverty constraint. Parents can then avoid sending the youngest children to work, and they may devote more financial resources to the financing of higher education for later-born children.

These two opposing forces imply that the relationship between birth order and education is an empirical matter. In Brazil, Emerson and Portela Souza (2004) find that male and female first-born children are less likely to attend school. While older male children are sent to the

labour market, older female children are kept out of school in order to assist the mother with housework.[4] According to Ejrnaes and Pörtner (2004), the advantage of children with higher birth orders is reinforced in the case of endogenous fertility, which is supported by empirical evidence in Philippines. They find that within family variation in education is more important when the parents are more educated and own more land.

When capital markets are incomplete, parents have to decide how to ration available resources to their children, which gives rise to the sibling rivalry theory (Garg and Morduch, 1998). A child's education then depends on the gender composition of the sibship. Under the assumption that the returns to human capital are higher for men, children are expected to fare better when a greater fraction of their siblings are female. Results from poorer countries suggest that there exists a competition for scarce resources. In Ghana, children's enrolment in secondary schooling is fifty per cent higher in all-sister households than in all-brother households. A similar finding is obtained for Tanzania, with increased years of schooling for children having more sisters, but evidence from South Africa shows very limited effects of sibling gender composition (Morduch, 2000).

When the returns to schooling differ between men and women and parents have an aversion to earnings inequality among their children, sibling gender composition effects may appear even in the absence of borrowing constraints. Indeed, Butcher and Case (1994) suggest that in the US, where liquidity constraints are less likely to be binding than in less advanced economies women raised only with brothers receive on average more education than women raised only with sisters. Nevertheless, for the same country, Kaestner (1997) and Hauser and Daphne Kuo (1998) find no support for the hypothesis that sibling gender composition affects educational achievement, and sibling inequality is independent of parental wealth (Gavaria, 2002).

Direct interactions between siblings are an additional may also result in sibling gender composition effects. As pointed out in Ono (2004) and Bommier and Lambert (2004), sibling rivalry may arise from gender-specific substituabilities or complementarities between children, either in the parental utility function or in the family production function. For instance, there may be some positive externalities between siblings, such that education becomes less expensive for latter born children. Specific goods such as books or clothes already bought for older children may be used again by younger children, or the latter may be helped by older children when doing schoolwork. In that case, the impact of an additional brother relative to an additional sister is

---

[4] A limitation of that study is that only school attendance is taken into account, while birth order may have a more sizeable impact when considering years of education.

different for a given child, although empirical evidence is rather contrasted concerning the gender and birth order of the children are favoured by different sibling compositions.

In Taïwan, having an older sister instead of an older brother increases a boy's educational attainment (Greenhalgh, 1985). In that country, remittances from unmarried daughters are used to finance higher education of sons, as male offspring are more likely to care for their parents in old age. Investing in boys rather than in girls would then be a more profitable strategy. Parish and Willis (1993) reach a rather similar result for that country, but they note that older sisters are beneficial for younger siblings of either gender. It could be that older female children are more likely to care for the younger children, as shown by Jamison and Lookheed (1987) in the context of Nepal. Lillard and Willis (1994) find that in Malaysia schooling is a decreasing function of the number of older brothers for boys and of the number of younger sisters for girls.[5] Sudha (1997) also shows that the effects of gender composition do not appear concentrated only among girls.

Last but not least, heterogeneity in preferences among household members may also explain inequality in educational attainment among siblings. According to recent collective approach or other bargaining models, the allocation of resources is expected to depend both on the level of household resources and on the distribution of control over these resources. In Brazil, maternal education has a larger impact on daughters schooling than on sons schooling in Brazil, while the influence of paternal education is more important on sons education (Thomas *et alii*, 1996). These differences result from a disagreement between spouses regarding the allocation of resources within the household rather than from gender-specific differences in the technology of human capital production. This may occur either because both spouses have different preferences for investing in the human capital of their children or because the returns to investing in gender-specific children differ for father and mother.

Overall, this literature suggests complex siblings effects. The numbers of brothers and of sisters are important determinants of a child's education, especially in developing countries. They may affect differently the education of boys and girls. Finally, (gender-specific) birth order is also an important determinant of education, although the sign of the corresponding effects is a priori ambiguous.

---

[5] But no significant differences are observed between the effects of older and younger siblings of the opposite gender.

3/ Data and descriptive statistics

3.1. The Albanian context

Over the past decade, Albania has been seeking to develop the framework for a more open society. The main challenge during the transition to a market economy was the lack of basic infrastructure in the economy, and especially in the education sector. Despite very low income levels, there has been significant progress in creating conditions to economic growth. In that country, a substantial fraction of the population remains vulnerable to poverty. Importantly, income-related poverty is mainly due to the lack of access to basic infrastructure, education and health services. Consequently, education is costly for households and expenditures related to investments in human capital are likely to be the source of strong inequalities both between and within families.

Elementary education in Albania is compulsory (grades 1-8), but most students now continue at least until a secondary education, either vocational/technical (3 to 5 years) or regular (4 years). Students must successfully pass graduation exams at the end of the 8th grade and at the end of the 12th grade in order to enter tertiary education, which is made of 4 to 5 years undergraduate, 1 to 3 years graduate and 3 years postgraduate.

The history of education in Albania has been rather eventful, in relation to the political context. Literacy rate was particularly low in the first half of the 20th century, mainly because very few schools offered education in Albanian language, and enrolment rates were very low. Until independence in 1912, education was offered mainly in Greek and Turkish. An ambitious national educational system, using Albanian language and significantly focused on professional programs, was developed during the 30s, with an important effort on teacher's formation.

However, because of political instability (Balkanian war, World Wars I and II), education was often under the control of various foreign occupation powers, and provided in Italian, French or English. Generalized education in Albanian language really began under the Communist regime (1944), but it remained oriented towards soviet ideology until the 1960 Albanian-Soviet breakdown. Children's education then progressed dramatically (from 60,000 students enrolled at all levels in 1939 to more than 750,000 by 1987), both in rural and urban areas, and adult education programs were also provided. As a result, illiteracy had virtually disappeared by the late 80's: it is today less than 7% of total population aged 9 or older.

A reorganization plan was announced in 1990 that would extend the compulsory education program from eight to ten years. The following year, however, a major economic and political crisis in Albania, and the ensuing breakdown of public order, plunged the school system

into chaos. Nearly one-third of the 2,500 schools below the university level were ransacked and fifteen school buildings razed. Many teachers relocated from rural to urban areas, leaving village schools understaffed and swelling the ranks of the unemployed in the cities and towns (about 2,000 teachers fled the country). In the late 1990s, many schools were rebuilt or reconstructed to improve learning conditions. Most of the improvements have happened in the larger cities, such as the capital Tirana which suffered from vast overcrowding of classrooms.

3.2. The LSMS data

For our empirical analysis, we use a cross-sectional survey conducted in Albania by the World Bank between April and July 2002.[6] It is part of the Living Standard Measurement Study program (LSMS thereafter), which is a very important tool in measuring and understanding poverty in developing countries. This survey has been conducted about a year after the Population and Housing Census conducted in April 2001. The Albania survey was undertaken by the living standards unit of the Albanian Institute of Statistics (INSTAT), with the assistance of the World Bank. The sample design for the survey includes 450 primary sampling units, each containing 8 households, so the sample is made of exactly 3600 households. A full description of the survey is given in the basic information document provided by the Worldbank (INSTAT, 2002). Importantly, the survey is representative for Tirana and other urban and rural areas.

Four survey instruments appear in the 2002 Albania LSMS. The core of the survey is a household questionnaire which includes all the core LSMS modules (see Grosh and Glewwe, 2000), with data related to household roster, dwelling, education, health, employment, transfers and social assistance, other income sources, consumption, and anthropometrics. A few additional modules are available, dealing with migration, fertility, subjective poverty, agriculture, and non-farm enterprises. The three other instruments are a diary recording household food consumption, a price questionnaire and a community questionnaire. In rural areas, the community is defined as a village and the inhabited area surrounding it, while it is related to administrative partitions in the urban areas.[7]

To investigate the sources of inequalities in education, either between or within families, an accurate description of educational attainment of each sibling is needed. The LSMS education module provides information for each household member on whether the person has ever attended school, the highest grade completed and the highest diploma. It also informs whether

---

[6] For further information, see the following url hhtp:\\www.worldbank.og\lsms\. Data are available on line.

the person is enrolled in school during current academic year and whether (s)he is currently attending school. However, a central difficulty arises with the Albania core education module, as in the other LSMS surveys, since it provides no information on the education of children who are no longer living in the household.

To solve this problem, we could focus on families with only young children, living with their parents, but this would lead to additional difficulties. On the one hand, these households are more likely to have incomplete fertility. On the other hand, this induces a censoring problem since younger children are more likely to have incomplete schooling. In turn, it may introduce sample selection bias since these censored observations also have a higher probability to achieve an education level higher than the reported level. Fortunately, the fertility module of the Albania LSMS allows us to overcome that difficulty. This module provides information on all the children ever born from all women in the household above 14 years.

This module includes some basic questions for those children who no longer live in the parental home. In particular, we know for each child the gender, whether the child is alive, in which country the child lives, and the highest level of completed education. However, there is no information about current enrolment in school. So, using both the education and fertility modules, we get a full description of educational attainment at the sibship level. Several variables related to the children may be constructed to explain education in Albania. For each child, we have information about gender, rank in the sibship, number of siblings, number of brothers or sisters, number of older or younger siblings, among other covariates including parental characteristics.

Our data suggest that the most discriminating educational level in Albania is the complete primary school level since about one half of the children have more than 8 years of schooling. Since children are approximately 15/16 years old once they have completed elementary education, we rely on the following strategy.

In order to ensure that there is no selection bias, we focus on children older than 16. We know whether they have more than the 8 years primary grade.[8] As pointed out in Ejrnaes and Portner (2004), selection bias is avoided because the selection is based on an exogenous covariate, i.e. age. Since children have to be older than 16 in 2002, this means that our study is

---

[7] Community information was collected from interviews with persons reputed to have superior information about each module within a community, usually an elected or appointed community leader.

[8] We assume that children aged over 16 have already either finished primary school or have dropped out and will never finish. Given that primary school lasts 8 years, it may be that some kids over 16 are still in primary school, for instance in case of multiple grade repetitions. For kids living at home and enrolled, we find that only 9 (resp. 1) children aged 17 (resp.18) are still in primary school.

based on education achieved during the last decades.[9] Although our aim is not to understand how inequalities may have been shaped by the institutional context, we can get some ideas about the evolution of education over time and we measure the consequences on the magnitude of education inequalities within and between families using the 2002 LSMS.

Hence, we rely on a binary variable to measure educational inequalities. At first glance, one could argue that this approach is not much informative, since there is not much variability in a binary variable. Nevertheless, we believe that this is a realistic approach to account for parental investments in human capital in poor countries. In Albania like in many developing countries, the elementary school system is nowadays theoretically mandatory, most often publicly provided and tuition-free. However, the tumultuous history of Albania, together with complex language and religious constraints, made it very difficult for some children to access even primary school at some periods (see Section 3.1). In such a context, the main decision for parents is just to decide, for each child, whether they will send him/her to school or not, so that some children have absolutely no education. In addition, the successive political and economic crises often induced many primary schools to close their doors, so that many children dropped out before they could finish primary school. Our binary approach therefore seems really relevant for studying education inequalities in Albania over the period concerned.

Although we mainly rely on a binary educational variable in our econometric analysis, we agree that this approach may lead to a lower bound when measuring inequalities within the family. As a consequence, we have further investigated the robustness of the binary approach by estimating other models in which education is a categorical ordered variable. We have also attempted to explicitly account for censoring when using the ordered levels of education.

The sample is restricted to families with at least one child 16 years old or more. After deleting missing values and sibships with incomplete information on education or age, we get a sample of 6959 children, in 2091 families. For these families, the household head may either live alone or in couple. In order to better understand the respective impact of father's and mother's education on the child's educational attainment, we have also considered a restricted sample in which the household head lives in couple. This new sample is made of 5612 children corresponding to 1728 families.

---

[9] In our analysis, we consider a sample of children with an average age of 30 years. We use the term « children » in this paper because our focus is on sibships and we merge the characteristics of these different children with those of

3.3. Descriptive statistics

Let us first describe education attainment in Albania. As shown in Figure 1, nearly half of the children have completed more than the primary school (48.5%). Among those who are more than the eight-school level, 26.8% have a secondary school level, 12.8% have performed a vocational program, and 8.8% are graduated or post graduated. Differences by gender are presented in Figure 2. On the whole, girls are more educated than boys according to the LSMS. The fraction of boys with at most 8 years of schooling is 52.6% instead of 50.4% for girls. Secondary and university levels are more common among girls, whereas vocational education is more frequent for boys.

Insert Figure 1 here

A preliminary approach to investigate the source of inequalities consists in decomposing the variance of the binary education variable into between and within components. In our sample, the between variance is equal to 0.182, while the within variance is significantly lower and equal to 0.087. This suggests that inequalities in education mainly occur between families in Albania, although this simple analysis clearly neglects the role of children and parental characteristics to explain educational attainment.

In Table 1, we provide a brief description of the different covariates which are included in our empirical analysis. There are slightly more boys than girls in the sample, and the average age for children is around 30 years. The mean number of siblings is 3.8, with slightly more brothers than sisters. Concerning parents, selected characteristics of the head are marital status, age, education, area of residence (either rural or urban) and religion. When comparing the distribution of education between both generations, we observe that the proportion of parents with more than 8 years of schooling is only around 30%. Many parents have completed only 4 years of primary education, or have simply no education at all.

Insert Table 1 here

As expected, both the child and parental characteristics influence the pattern of education in Albania (Table 2). Consistently with the history of education in Albania (see Section 3.1), receiving more than the primary school level is more likely for children born between 1961 and 1975. The decrease in educational level for younger children observed in Table 2 reflects both the deterioration of educational system[10] and the fact that some young children are still enrolled and

---

their parents. Clearly, we do not account for schooling attainment of young children in 2002.

[10] Children born after 1975 were 15 years old during the 90's, so they were supposed to be at primary school during the chaos of the educational system.

have not reached yet complete primary level.[11] Gender differences are also substantial. While young women are significantly more educated, the opposite result holds for women older than 45. The number of siblings reduces the fraction of children with more than 8 years of schooling. We observe a significant and continuous decline as the size of the sibship increases, especially for females. The proportion of children with more than primary education is 71.1% for those who have at most one sister or brother, but it is twice lower (32.3%) for those with at least 6 siblings.

Insert Table 2 here

In Table 3, we further investigate the role of the gender composition of the sibship. For a given number of siblings, we find that a child is less educated when (s)he has more brothers in the sibship. This clearly indicates that there exists sibling rivalry in Albania. The role of sibship gender composition seems larger as the size of the sibship increases. For instance, with 3 siblings, about 73% of the children who have only sisters have more than the primary education level, while this fraction is only 48.3% with two brothers. In addition, the impact of the gender composition of the sibship seems more important for girls than for boys.

Concerning parental characteristics, the most important covariate is the educational level of the head. Intergenerational effects of human capital transmission are large. When the parent is graduated or post graduated, nearly 90% of children have more than 8 years of schooling, while this proportion is less than 35% when the parent has at most the primary 4 years level. These differences are more pronounced for girls than for boys. Compared to Muslim and Catholic religions, Orthodox children have a higher probability to receive more then 8 years schooling. The probability to receive more than primary education is larger when parents live in an urban area.[12] Finally, the presence of a secondary school is a strong incentive for completing primary school and the presence of a secondary school has more impact on primary education than the presence of a primary school.

To better understand how these factors combine to determine the probability for a child to achieve more than the primary school level, we now turn to an econometric analysis. Drawing on random effects Probit models, we decompose the source of educational inequalities into observed/unobserved and between/within components.

---

[11] We will see, however, that our estimates are not biased by this second effect.

[12] In comparison, this proportion is more than twice lower when parents live in a rural area.

4/ Econometric analysis

4.1/ A random effects Probit formulation

We are interested in estimating the probability that a child has achieved more than the primary school level. The corresponding model is:

$$Y_{ji}^* = \beta' X_{ji} + \varepsilon_{ji} \quad j = 1,...,J \quad i = 1,...,N_j, \tag{1}$$

where the unobserved latent variable Y_{ji}^* can be interpreted as a propensity to education. Let Y_{ji} denote the variable such that $Y_{ji} = 1$ when the child has more than 8 years of schooling and $Y_{ji} = 0$ otherwise. We have $Y_{ji} = 1$ when $Y_{ji}^* > 0$, and $Y_{ji} = 0$ when $Y_{ji}^* \leq 0$. In (1), index j represents the family and index i stands for the child, N_j is the sibship size, X_{ji} is a set of explanatory variables specific either to the child or to his family, β is the vector of associated parameters, and ε_{ji} is a random term such that:

$$\varepsilon_{ji} = u_j + v_{ji}. \tag{2}$$

The random term u_j, corresponding to unobserved heterogeneity for the family j, is assumed normally distributed, and v_{ji} is a child-specific normally distributed random term. Unobservable traits inherited from parents as well as measurement errors in observable parental characteristics will be picked up in the family component u_j, while child's innate ability and measurement errors in child's covariates are part of v_{ji}. The variance of the within unexplained component is normalized to one, while the variance of the between unexplained component is denoted by σ_u^2, so that $v_{ji} \sim N(0;1)$ and $u_j \sim N(0;\sigma_u^2)$. We assume that the perturbations u_j v_{ji} are independent from each other, and from the explanatory variables X_{ji}. However, the total random perturbation ε_{ji} of the different children in one family are correlated with each other because of the common term u_j.

Let $a_{ji} = -\infty$ and $b_{ji} = -\beta' X_{ji}$ if $Y_{ji} = 0$, and $a_{ji} = -\beta' X_{ji}$ and $b_{ji} = +\infty$ if $Y_{ji} = 1$. Then, the log likelihood of the model may be expressed as $L = \sum_j \ln\left(P(Y_{j1}, Y_{j2},...,Y_{jN_j})\right)$, where the probability $P(Y_{j1}, Y_{j2},...,Y_{jN_j})$ is given by:

$$P(Y_{j1}, Y_{j2},...,Y_{jN_j}) = \int_{a_{j1}}^{b_{j1}}...\int_{a_{jN_j}}^{b_{jN_j}} \phi(\varepsilon_{j1},...,\varepsilon_{jN_j}) d\varepsilon_{jN_j}...d\varepsilon_{j1} \tag{3}$$

where ϕ is the density function of the multivariate normal distribution with positive correlations and with unity variances. The multivariate integral in (3) cannot be computed directly because of the correlations between the ε_{ji} terms. However, the conditional independence assumption implies that the multivariate normal density of the v_{ji} terms is the product of univariate standard normal densities, which allows computing the above integral. Let us denote by Φ the cumulative distribution function of the univariate standard normal distribution. Then, we get

$$P(Y_{j1}, Y_{j2}, ..., Y_{jN_j} | \varepsilon_{ji}) = \prod_{i=1}^{N_j} \left[\Phi(b_{ij} - u_j) - \Phi(a_{ij} - u_j) \right] \text{ and:}$$

$$P(Y_{j1}, Y_{j2}, ..., Y_{jN_j}) = \int_{-\infty}^{+\infty} \left\{ \prod_{i=1}^{N_j} \left[\Phi(b_{ij} - u_j) - \Phi(a_{ij} - u_j) \right] \right\} \phi(u_j) du_j \qquad (4)$$

This specification is a random effects Probit model, which can be easily estimated using Gaussian quadrature techniques (Butler and Moffitt, 1982). After estimating the random effects model, we measure educational inequalities based on the following variance analysis.

First, we compute the linear fitted value of the latent variable for each observation of the sample $\hat{Y}_{ji}^* = \hat{\beta}' X_{ji}$. This latent variable can be interpreted as a (continuous) propensity to be educated. Then, we compute the mean of these fitted values $\hat{Y}_j^*$ for each family j, and we generate a new variable $\hat{Y}_{ji}^d = \hat{Y}_{ji}^* - \hat{Y}_j^*$. So, $\hat{Y}_{ji}^d$ is a measure of intra-household heterogeneity for a given family. As a final step, we compute the variances of $\hat{Y}_j^*$ and $\hat{Y}_{ji}^d$: $V(\hat{Y}_j^*)$ is the between explained variance and $V(\hat{Y}_{ji}^d)$ is the within explained variance. Recalling that $V(u_j) = \sigma_u^2$ and $V(v_{ji}) = 1$ are the variances of respectively the unexplained between and within components, we can provide an accurate description of the pattern of educational inequalities.

4.2/ Explaining educational attainment

We use the subsample of families in which at least one child is older than 16 and estimate several random effects Probit models for the probability to have more than 8 years of schooling. Both characteristics of the child and of the parents are introduced in the various regressions. Results are reported in Table 4.

Insert Table 4 here

In model (1), the only child characteristics are gender, age, birth cohort (6 dummies), birth cohorts interacted with age, number of siblings and birth order. The null joint hypothesis that education does not depend on these covariates is strongly rejected. According to these

estimates, the probability to have more than the primary level is higher for girls (at the 5% level). Having several siblings strongly reduces schooling, since parents have fewer resources to devote to each of their children. We find a slightly positive effect, albeit not significant, for birth order. Interestingly, the magnitude of the coefficients for the child's characteristics is rather low.

Our specification means that we consider a piecewise-linear function for the child's age. To better interpret the results, we represent in Figure 3 the probability for a child to have completed more than 8 years as a function of birth year. First, for those born between 1950 and 1960, the profile is increasing. It is then slightly decreasing for those born between 1960 and 1975. Finally, we note that the probability is much lower for those children who were enrolled during the 90s (they are born after 1975). This finding is very consistent with the history of Albania, with a harmful effect of the crisis during the beginning of the 90s on schooling.

Insert Figure 3

Several variables related to the parents play a prominent role when explaining the child's educational attainment. The probability to have more than 8 years of schooling is more important when the respondent parent is a woman and lives in couple, while age is not a significant predictor. The child's education increases with the parent's education. Completing primary school is more likely when the parent has completed secondary school or more. This effect is mainly due to an intergenerational transmission of human capital, but it may also be linked to the impact of parental resources.[13]

There exist strong cultural differences when explaining education in Albania. On the one hand, a child with orthodox parents is more likely to be high educated. With respect to the Muslim reference group, the category comprising other religions also increases the schooling probability (at the 5 percent level). On the other hand, we find that children are significantly less educated when their parents live in a rural area. As parents are more likely to undertake agricultural activities in rural areas, this could prevent children from completing high education.[14] Also, it may be that schooling is more difficult in rural areas, either because of the lack of appropriate infrastructure or simply because parents live far away from secondary schools or universities. Education would then be more costly, for instance if parents have to rent a dwelling

---

[13] More educated parents have certainly a higher level of income, and it will be much easier for richer parents to finance the cost of education. Unfortunately, we only have information on the current level of income in the data. We choose to exclude this covariate, as the household income variable may be caused by the education of children and the measure of income is not related to the period of investment in education.

[14] Ideally, the regression should include a measure of the parent's socio-economic position. Unfortunately, this information is only partially available in the LSMS questionnaire. In the labour module, we know whether the respondent has worked during the past 12 months and eventually the occupation of the last job, but there is no information for those who are no longer working.

for the child. Part of these effects is captured by community variable and more schooling is observed when there is a secondary school in the community.

Our results show that parental characteristics are more influential than those of the child when explaining the child's education. The variance analysis sheds light on the source of educational inequalities in Albania. Our main conclusions are as follows (see Table 4). First, with the random effects Probit model, we explain 40% of the total variance, which is rather important. Second, much of the explained variance is due to differences between families. The weight of the within explained component is equal to 0.9%, while the between weight is 39.1%. Third, concerning the unobserved part of the total variance, we find that there is slightly less within variance than between variance (respectively 27.1% and 32.9%). These findings globally show that educational inequalities between families are more important than those within families (respectively 30% and 70%).

One could argue that the low percentage of the within observed variance is due to the lack of appropriate explanatory variables for children. We have then estimated several additional models (see models (2) and (3) in Table 4).

Model (2) tests for the presence of a sibling gender composition effect. For that purpose, we introduce into the previous model the number of sisters (assuming a linear effect). Since we control for the size of the sibship, a null coefficient for that variable means that brothers and sisters have the same effect on a child's education. According to our estimates, it is more beneficial for a child to have sisters rather than brothers. With more sisters, the probability to have more than 8 years of schooling is significantly improved. Several explanations have been given for that result.[15] Since education is more profitable for boys than for girls, parents should primarily invest in their sons if they cannot provide the efficient level of human capital for their different children.

Model (3) analyzes both the birth order and gender composition of the sibship, through the numbers of older and younger brothers and sisters. Coefficients for the numbers of brothers (either younger or older) are highly significant and negative. Both estimates are not statistically different, suggesting that all boys are treated the same way whatever their birth order. The coefficients are much lower for the number of sisters, and only the number of younger sisters is statistically significant at the 1% level. It could be that older-born sisters are more likely to help their parents in domestic tasks and have to care for the latter-born children.

---

[15] See in particular the detailed discussion in Butcher and Case (1994) and Garg and Morduch (1998).

Although these two additional models better explain the role of the child's characteristics, we do not observe much difference with respect to the previous specification in terms of variance analysis. The explained component of the variance is again about 40%, and much of the variance is due to differences between families: the within explained component represents 0.9% of the variance. When computing the unexplained variance, we still find that the between component is slightly higher than the within component (see Table 4).

Since descriptive statistics suggest substantial gender differences in educational attainment, we estimate a model where all the explanatory variables are interacted with the gender of the child. These interaction terms indicate whether or not the child and parental characteristics are more or less important for boys and girls.[16] Our results are twofold. First, the within component is now much higher, i.e. 2.1% versus 0.9%, and the unobserved within component of the variance is reduced. Of course, the within observed value is still very low, but accounting for gender differences is helpful to explain the intra-household allocation of education. Second, two important factors have a different impact on boys and girls. Birth order has a positive effect on girls' education, which is consistent with the fact that first born girls have to take care of younger children, and girls are more discriminated in rural areas than in urban areas. An explanation is that domestic tasks as well as help with farming activities prevent daughters from achieving more than the primary level.

Insert Table 5 here

We have also estimated separate regressions based on the urban-rural status. As shown in Table 5, the probability of having completed primary school is higher for girls in urban areas and it is a decreasing function of the numbers of siblings and sisters. Conversely, in rural areas, education is lower for girls and the sibship sex composition does not matter. While parental characteristics have a rather similar effect in both cases (albeit they are less significant among rural families), there are large differences in variance decomposition. Indeed, the overall explained variance is much higher in urban than in rural areas (34.3% versus 22.3%). At the same time, the within components are more important in rural areas: 2.0% instead of 1.5% for the observed variance and 34.8% instead of 29.8% for the unobserved variance.

The comparison of variance decompositions shows within families inequalities are of the same order of magnitude in rural and urban samples, whereas inequalities between families are significantly larger in urban areas. This implies that inequalities would be overestimated in urban

---

[16] With respect to the model with no interaction terms, a simple likelihood ratio test shows that the specification with interaction effects has to be preferred.

areas if they were measured at the family level (average education in the family) rather than at the child level. These larger differences in education between siblings in rural areas than in urban areas may be explained by time and budget constraints, which are probably more severe in rural areas, preventing rural parents from educating some of their children.

We also account for the role of religion and estimate regressions respectively for Muslims and non-Muslims. Again, several differences are observed between the two populations. Sibship and the siblings sex composition significantly matters for the Muslim children, whereas they do not influence educational attainment for the other religions. Muslim girls seem more likely to have completed 8 years of schooling than other girls. However, this effect is partly offset by the number of sisters effect (a girl has one sister less than her brothers). Concerning the head of the household, gender, age and marital status are also significant characteristics only for this group, while the parental education effect is almost similar for both religions. The variance decomposition is also influenced by the religion. The explained part of the variance is much lower among Muslims (37.6% versus 46%) and the within component is slightly lower among the Muslisms (28.7% versus 29.1%), suggesting that Muslim parents are slightly less unequal than other parents.

Finally, intra-family inequalities in education may be explained by parental characteristics if the father and the mother behave in a different way with respect to their children. If we suppose that mothers have a preference for investing in their daughters' education and that fathers have a preference for their sons' education, this would lead to increased inequalities among siblings. To further investigate such effects, we now consider the sample with parents living in couple. With respect to our previous estimates, we introduce into the regression the educational level of the father (5 categories) along with variables comparing education of both spouses. We construct two additional dummies, respectively when the mother is less educated and more educated than the father (the reference being 'same education' for both spouses).

Let us focus on the effect of the parental education.[17] As shown in Table 6, the child is more likely to have more than the primary level when the father is high educated, especially if graduated or post graduated. The probability to have more than 8 years is strongly reduced when the father is more educated than the mother, while the effect is positive at the 1 percent level when the mother is relatively more educated. This may be the sign that there exist gender

---

[17] According to Table 6, estimates for the other variables are unchanged with respect to our previous discussion.

differences in family resources, women being more child oriented (Thomas, 1994).[18] However, these effects do not affect the decomposition of the total variance. With interaction terms, the fraction of the explained variance is 42.7%, and much of this variance is due to differences between families. Again, the between component of the unexplained part of the variance is slightly higher than the within component.

Insert Table 6 here

4.3/ An extension to the case of ordered education

In our previous analysis, educational attainment of the child was given by a dummy variable. Unfortunately, this restriction may lead to an under-estimation of the magnitude of inequalities within families. Among those siblings who have completed more than the primary school level, parents may for instance favour vocational studies for their sons and graduate or postgraduate studies for their daughters. We now extend our method to the case where education is measured by an ordered categorical variable.

Following the previous description of education in Albania (see Figure 1), we consider that education is represented by a four categories ordered variable, the corresponding levels being primary school, secondary school, vocational studies and graduate or post graduate studies. Then, we have to estimate a random effects ordered Probit model. The underlying latent model is still given by equation (1), but the observed educational level is:

$$Y_{ji} = n \quad \text{if} \quad \mu_n < Y_{ji}^* \leq \mu_{n+1} \tag{5}$$

with $n \in \{0;1;2;3\}$, $\mu_0 = -\infty$ and $\mu_4 = +\infty$. The log likelihood of the corresponding ordered model is given by (3), but with $a_{ji} = \mu_n - \beta' X_{ji}$ and $b_{ji} = \mu_{n+1} - \beta' X_{ji}$ if $Y_{ji} = n$. Note that the binary model is an example of ordered model, except for the normalization $\mu_1 = 0$. Again, the model can be estimated using numerical approximations and Gaussian quadrature techniques.[19] The difficulty with the ordered model is still linked to censoring, since young children may have not yet completed their education. We choose two different ways to circumvent the problem.

A first solution is to consider the subsample of families amongst whom the youngest child is at least 21 years old. Indeed, a child who is above 20 and still enrolled in school will necessarily be in the upper category (graduate or postgraduate studies), so that the censoring

---

[18] Nevertheless, with additional interaction effects, we do not observe that educational differences in the allocation of resources vary with the gender of the child.

[19] Specifically, the computation of the random effects ordered Probit model is done through a program discussed in Frechette (2001), which makes use of the analytical first derivatives.

problem vanishes. There is no sample selection bias with that method since age is exogenous, but this reduces the number of children in the sample to 5615 (1701 families). A second solution is to place no restriction on the child's minimum age, but to modify the individual likelihood so as to account for censoring. The censored model is in fact a simple extension of the ordered model. The probability that Y_{ji} equals n is $\Phi(\mu_{n+1} - \beta'X_{ij} - u_j) - \Phi(\mu_n - \beta'X_{ij} - u_j)$ when the child has completed schooling, while it is equal to $\sum_{k=n}^{3} \Phi(\mu_{k+1} - \beta'X_{ij} - u_j) - \Phi(\mu_k - \beta'X_{ij} - u_j)$ when the observation is censored. This in turn increases the size of the sample, which now comprises 6956 children (2091 families).

As a benchmark, we first estimate on the sample of 5615 children the probability that a child has more than the primary level of education, using a random effects Probit model. Clearly, the first regression in Table 7 leads to very similar estimates with respect to those previously discussed, meaning that the selection is not a problem. This similarity holds both for the coefficients and for the variance decomposition. Hence, we now turn to the random effects ordered specification, with children being at least 21 years old. As shown in Table 7, we find a different pattern when decomposing the variance of the ordered educational outcome. Our results are twofold.

Insert Table 7 here

First, the decomposition into the between and within components for the explained variance still indicates that inequalities stem mainly from differences between families (31.4%) rather than within families (0.6%). However, we reach an opposite conclusion for the unobserved variance. The weight of the within component is now higher than the one of the between component (36.6% instead of 31.3%). In the binary models, the between unexplained variance fraction was almost comparable, whereas the within unexplained variance was only 25.8%. Accounting for more educational levels increases the magnitude of unexplained variations within families. It therefore seems that families tend to be more egalitarian at providing a complete primary education to all their children, but they become more unequal at the vocational, graduated and post graduated levels. Since such education is far more costly than primary education, liquidity constraints play a more important role and parents can offer higher education only to a fraction of their children.

A second result is that the fraction of the explained variance is now lower for the ordered Probit formulation than for the Probit specification, 32.1% instead of 40.8%. This suggests that the larger inequality within families for higher educational levels than for primary education can

hardly be explained by the observable characteristics. It could be argued that the comparison between the binary and ordered specifications may be inappropriate because the dependent variables are different. However, the latent variables of the binary and ordered models should theoretically be the same, except with a different normalization.

On the one side, μ_1 is equal to 0 in the binary model, so the two latent variables may differ by a constant additive term. On the other side, the share of unobserved within variance differs between the two models, so the two latent variables may also differ by a multiplicative term (because of the normalization $\sigma_u = 1$). This means that the parameters estimated in the two models should be proportional rather than equal. In Figure 4, we show that this is approximately the case. When regressing the coefficients of the ordered model on the coefficients of the binary model, the fit is very good (with an adjusted-R^2 of 97.5%) and the constant is not significant at all. This means that the estimated impacts of explanatory variables on the latent variables are very similar in the binary and ordered models.

Insert Figure 4 here

Interestingly, properly accounting for censoring in the likelihood does not really affect the above conclusions, as shown by the third regression of Table 7. On the one hand, we find that gender, number of siblings and sibship sex composition influence education attainment.[20] On the other hand, the decomposition indicates that around 35% of the variance is explained by the covariates introduced into the regression. Among the unobserved variance, the fraction of the within component substantially exceeds the one of the between component (35.5% instead of 29.6%), as in the specification without censoring. Hence, the different additional results based on a categorical outcome are close to those obtained with the binary educational variable.

5/ Conclusion and policy implications

Knowing the relative importance of between-family and within-family differences in schooling is very important from a public policy viewpoint. A central role of government policies is to affect the income distribution between individuals, but the family is another institution which is likely to affect socio-economic outcomes through investments in human capital. The family impact on the distribution of income depends on the variation in individual characteristics both between and within families, and on the rules for allocation of family resources.

---

[20] Since we account for younger children with the censored ordered Probit model with random effects, we have also estimated a model with more detailed information on birth cohorts for the children (with annual dummies for the more recent year). This specification does not influence the values of the other estimates.

Nevertheless, the role of education policies and family decisions has not been clearly identified so far when focusing on the determinants of school attainment and academic achievement in developing countries (see the discussion in Glewwe, 2002). In addition, little is known concerning the magnitude of between and within inequalities in education in those countries.

In this paper, we have presented a simple method to measure both the between and within sibship variances for the schooling variable. In a setting where educational levels rather than years of schooling are observed for the different siblings, we estimate discrete ordered choice models with random effects and focus on the latent education variable (corresponding to a propensity to be more educated) to perform a variance decomposition. Then, we provide estimates for the case of Albania. Our models explain 40% of the total variance in the propensity to be more educated and we find that inequalities in education are mainly due to differences between families. Differences within family are lower in Albania and far less easily explained. Based on an ordered education variable, the explained part of the propensity to be more educated is unchanged, but the unexplained within variations are increased. Therefore, focusing on a binary variable may lead to under-estimate within-family educational differences.

The comparison of our results obtained in rural and urban sub-samples suggests that increasing the average education level also decreases the inequality in education within sibships. On the opposite, it increases the inequality in education between families. The implications of educational policy in terms of inequalities may therefore be different at the family level and at the individual (child) level. Our findings suggest that improving education supply would both increase the average educational level and reduce education inequalities. This means that governments do not necessarily face an equity-efficiency trade-off when increasing school supply. In addition, we argue that the reduction in educational inequalities is usually underestimated when it is measured at the family level (by the average attainment in the sibship) rather than at the child level. Indeed, we believe that the child (rather than the family) is the relevant level for measuring educational inequalities.

Nevertheless, additional evidence on other developing countries for various continents is needed to extend our conclusions and to better understand the strength of inequalities within families, and we leave this task for future research. We hope that more adequate data on education in complete sibships will be collected in various countries, so that within-family educational inequalities can be analyzed all over the world. If our methodology applied to another country shows that within-family educational inequalities tend to increase when school supply is

increased, then this will be a strong motivation for targeting education towards the children who tend to be the most discriminated against (possibly eldest daughters).

References

Altonji JG, Dunn TJ (1996) The effects of family characteristics on the return to education. Review of Economics and Statistics 77:692-704.

Ashenfelter O, Krueger AB (1994) Estimates of the economic return to schooling from a new sample of twins. American Economic Review 84:1157-1174.

Basu K (1999) Child labor: Cause, consequence and cure. Journal of Economic Literature 37:1083-1119.

Becker GS, Tomes N (1986) Human capital and the rise and fall of families. Journal of Labor Economics 4:S1-S39.

Berhman J (1997) Intrahousehold distribution and the family. In Rosenzweig MR, Stark O (eds) Handbook of Population and Family Economics. Elsevier, North-Holland, 125-187.

Behrman J, Pollak RA, Taubman P (1989) Family resources, family size and access to financing for college education. Journal of Political Economy 97:398-419.

Behrman J, Pollak RA, Taubman P (1995) From parent to child. Intrahousehold allocations and intergenerational relations in the United States. University of Chicago Press, Chicago.

Behrman J, Taubman P (1986) Birth order, schooling, and earnings. Journal of Labor Economics 4: S121-S145.

Bommier A, Lambert S (2004) Human capital investments and family composition. Applied Economics Letters 11:193-196.

Butcher KF, Case A (1994) The effect of sibling sex composition on women's education and earnings. Quarterly Journal of Economics 109:531-563.

Butler LS, Moffitt R (1982) A computationally efficient quadrature procedure for the one factor multinomial Probit model. Econometrica 50:761-764.

Ejrnaes M, Pörtner C (2004) Birth order and the intrahousehold allocation of time and education. Review of Economics and Statistics 86:1008-1019.

Emerson PM, Portela Souza A (2004) Birth order, child labor and school attendance in Brazil. Mimeo, University of Colorado at Denver.

Frechette GR (2001) Random effects ordered Probit. Stata Technical Bulletin 159:23-27.

Garg A, Morduch J (1998) Sibling rivalry and the gender gap: Evidence from child health outcomes in Ghana. Journal of Population Economics 11:471-493.

Gavaria A (2002) Intergenerational mobility, sibling inequality and borrowing constraints. Economics of Education Review 21:331-340.

Glewwe P (2002) Schools and skills in developing countries: Education policies and socioeconomic outcomes. Journal of Economic Literature 40:436-482.

Greenhalg S (1985) Sexual stratification: The other side of growth with equity in East Asia. Population and Development Review 11:265-314.

Gross M, Glewwe P (2000) Designing household survey questionnaires for developing countries: Lessons from 15 years of the Living Standards Development Study. Oxford University Press, Oxford.

Hauser RM, Daphne Kuo HH (1998) Does the gender composition of sibships affect women's educational attainment ? Journal of Human Resources. 33:644-657.

INSTAT (2003) Albania Living Standard Measurement Survey 2002. Basic information document. Mimeo, Worldbank.

Jamison DT, Lookheed ME (1987) Participation in schooling: Determinants and learning outcomes in Nepal. Economic Development and Cultural Change 35:279-306.

Kaestner R (1997) Are brothers really better? Sibling sex composition and educational achievement revisited. Journal of Human Resources 32:250-284.

Lillard L, Willis RJ (1994) Intergenerational educational mobility. Effects of family and State in Malaysia. Journal of Human Resources 29:1126-1166.

Morduch J (2000) Sibling rivalry in Africa. American Economic Review, 90:405-409.

Ono H (2004) Are sons and daughters substitutable ? Allocation of family resources in contemporary Japan. Journal of the Japanese and International Economies 18:143-160.

Parish WL, Willis RJ (1993) Daughters, education, and family budgets: Taiwan experiences. Journal of Human Resources 28:863-898.

Sheshinski E, Weiss Y (1982) Inequality within and between families. Journal of Political Economy 90:105-127.

Sudha S (1997) Family size, sex composition and children' education: Ethnic differentials over development in Peninsular Malaysia. Population Studies 51:139-151.

Thomas D (1994) Like father, like son: Like mother, like daughter. Parental resources and child height. Journal of Human Resources 29:950-988.

Thomas D, Schoeni RF, Strauss J (1996) Parental investments in schooling: The roles of gender and resources in urban Brazil. Mimeo, Rand Labor and Population Program.

Yang DT, Zhu X (2003) Economic structural change and family investments in children. Mimeo, Virginia Polytechnic Institute.

Figure 1. Distribution of education for children
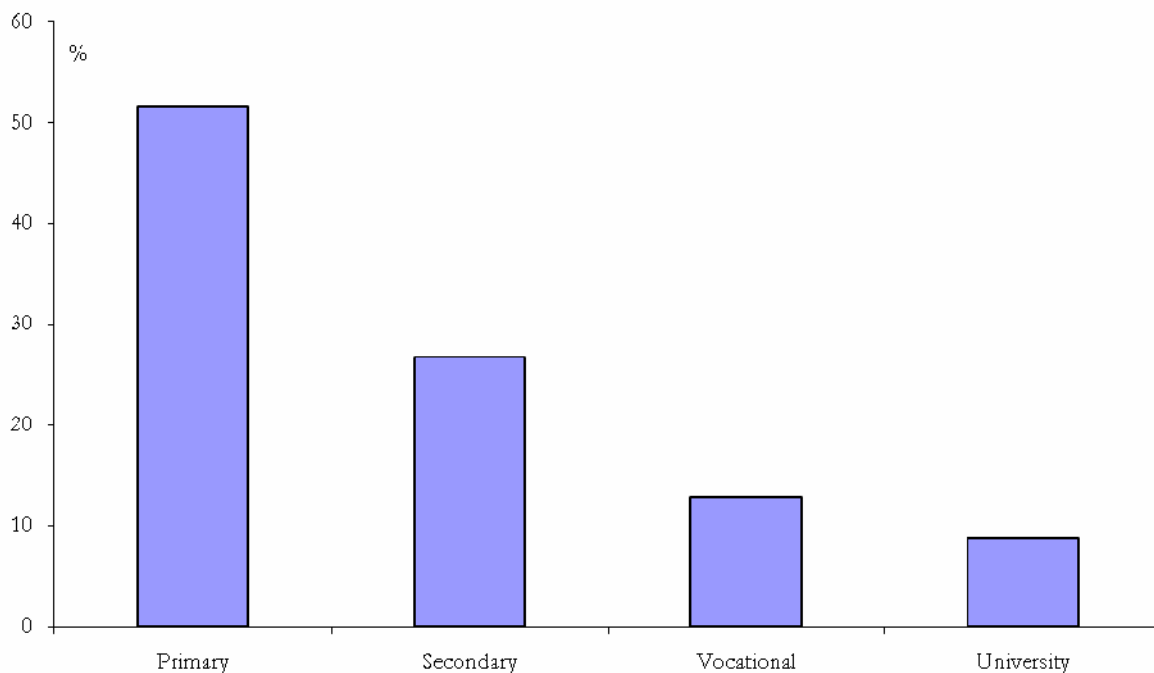
Source: LSMS Albania 2002.

Figure 2. Distribution of education for children, by gender
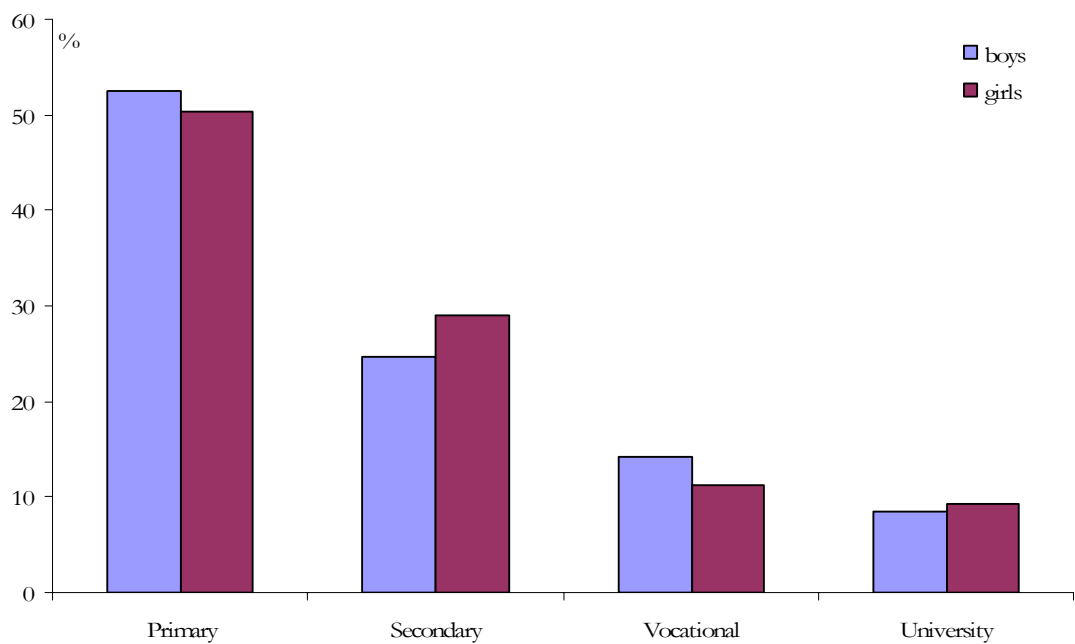
Source: LSMS Albania 2002.

Figure 3. Overall age/cohort effect on the probability to have more than 8-years schooling

Source: LSMS Albania 2002.

Figure 4. Comparing coefficients of the binary and ordered model
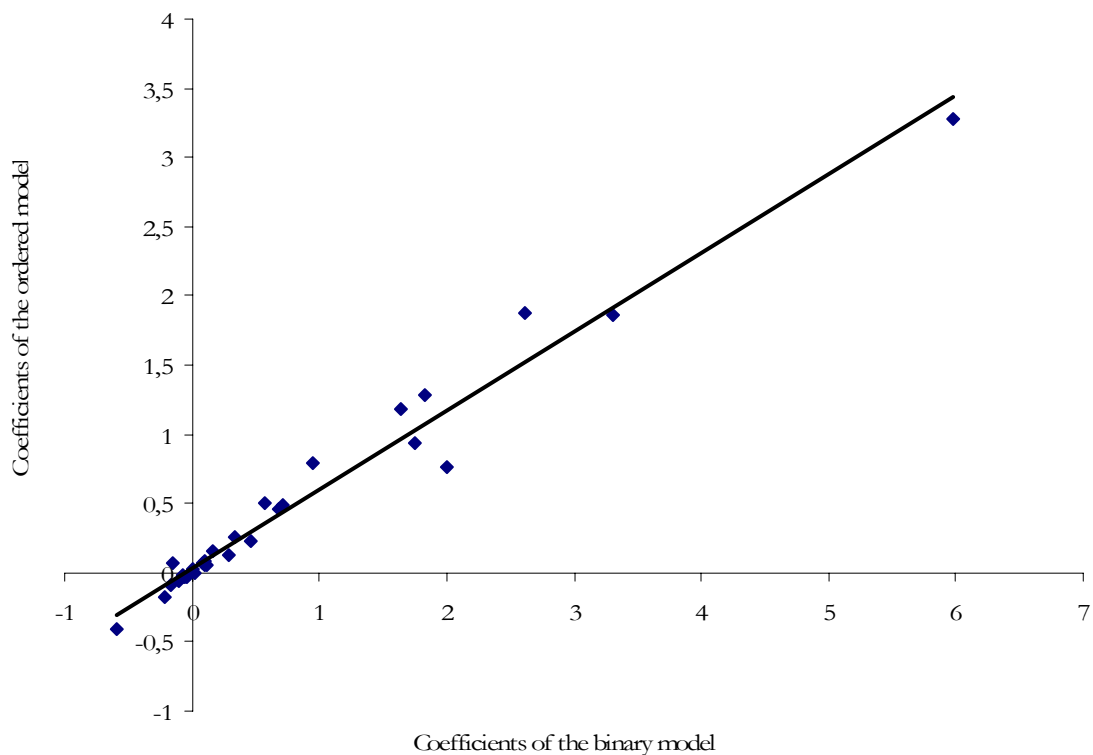
Source: LSMS Albania 2002.

Table 1. Description of the sample

Variables		Head sample		Spouse sample	
		mean	s.e.	mean	s.e.
Child's characteristics					
Education	Primary	0.515	0.500	0.519	0.500
	Secondary	0.268	0.443	0.278	0.448
	Vocational	0.128	0.335	0.115	0.319
	University	0.088	0.284	0.088	0.283
	More than 8-years school	0.485	0.500	0.481	0.500
Gender	Male	0.519	0.500	0.516	0.500
	Female	0.481	0.500	0.484	0.500
Age		30.420	9.970	28.858	9.061
Number of siblings		3.779	2.108	3.701	2.062
Number of brothers		1.921	1.394	1.873	1.372
Number of sisters		1.858	1.477	1.828	1.461
Parental characteristics		0.812	0.391	0.998	0.042
Married		61.805	10.419	60.823	9.977
Age		0.169	0.375	0.113	0.316
Education	None	0.220	0.414	0.209	0.407
	Primary 4 years	0.306	0.461	0.329	0.470
	Primary 8 years	0.229	0.420	0.260	0.439
	Secondary or vocational	0.076	0.266	0.089	0.285
	University or postgraduate	34.745	85.196	37.119	92.760
Religion	Muslim	0.791	0.407	0.782	0.413
	Orthodox	0.106	0.308	0.109	0.312
	Catholic	0.069	0.254	0.076	0.266
	Other	0.034	0.180	0.033	0.179
Region	Rural	0.481	0.500	0.463	0.499
	Urban	0.519	0.500	0.537	0.499
Number of observations		6956		5612	

Source: LSMS Albania 2002.

Table 2. Fraction of children with more than 8-years schooling (in %)

Variables		Male children		Female children		All children	
		%	N	%	N	%	N
Child's characteristics							
Birth cohort	≤1955	51.4	251	41.9	217	47.0	468
	1956-1960	51.6	345	53.4	277	52.4	622
	1961-1965	56.3	435	56.4	390	56.4	825
	1966-1970	56.1	554	53.7	492	55.0	1046
	1971-1975	49.6	613	55.9	546	52.5	1159
	1976-1980	38.2	615	45.6	614	41.9	1229
	>1980	38.9	799	43.6	808	41.3	1607
Siblings	0 or 1	67.7	579	76.5	374	71.1	953
	2	52.5	722	65.0	620	58.3	1342
	3	48.1	607	53.6	593	50.8	1200
	4	42.2	490	45.9	490	44.1	980
	5	41.2	483	37.5	514	39.3	997
	6 or more	33.4	731	31.2	753	32.3	1484
Parental characteristics							
Married	No	49.0	696	51.0	614	49.9	1310
(head)	Yes	47.1	2916	49.3	2730	48.2	5646
Education	None	37.1	631	30.9	543	34.2	1174
(head)	Primary 4 years	36.0	811	32.8	720	34.5	1531
	Primary 8 years	38.3	1094	42.7	1035	40.4	2129
	Secondary or vocational	66.2	786	73.4	804	69.8	1590
	University or postgraduate	85.5	290	92.6	242	88.7	532
Religion	Muslim	44.1	2850	46.4	2653	45.2	5503
	Orthodox	71.1	377	76.5	361	73.7	738
	Catholic	46.5	258	43.5	223	45.1	481
	Other	53.5	127	52.3	107	53.0	234
Region	Urban	62.4	1740	71.6	1603	66.8	3343
	Rural	33.5	1872	29.4	1741	31.5	3613
Community variables							
Primary school	No	30.6	458	26.5	408	28.6	866
	Yes	49.9	3154	52.9	2936	51.3	6090
Secondary school	No	32.6	1603	27.1	1491	30.0	3094
	Yes	59.2	2009	67.8	1853	63.3	3862
Mean probability		47.4	3612	49.6	3344	48.5	6956

Source: LSMS Albania 2002.

Table 3. Education and gender composition of the sibship

Variables	Male children		Female children		All children	
	%	N	%	N	%	N
Two children						
0 brother	69.5	236	78.8	113	72.5	349
1 brother	69.9	272	76.5	234	72.9	506
Three children						
0 brother	56.5	147	83.3	108	67.8	255
1 brother	48.8	402	68.4	304	57.2	706
2 brother	57.8	173	50.5	208	53.8	381
Four children						
0 brother	61.0	59	87.5	48	72.9	107
1 brother	42.5	240	54.7	214	48.2	454
2 brothers	51.6	248	45.0	249	48.3	497
3 brothers	43.3	60	57.3	82	51.4	142
Five children						
0 brother	51.7	29	80.0	15	61.4	44
1 brother	44.1	127	51.7	116	47.7	243
2 brothers	40.6	175	39.9	198	40.2	373
3 brothers	47.0	134	44.5	128	45.8	262
4 brothers	38.0	25	51.5	33	32.8	58

Source: LSMS Albania 2002.

Table 4. Random effects estimates of the probability to have more than 8-years schooling

Variables		(1)		(2)		(3)	
		coef	t-test	coef	t-test	coef	t-test
Constant		-6.885***	(5.80)	-6.869***	(5.87)	-6.874***	(5.88)
Child's characteristics							
Female		0.105**	(2.39)	0.169***	(3.41)	0.157***	(3.14)
Age		0.125***	(4.40)	0.127***	(4.43)	0.125***	(4.42)
Birth cohort: ≤1955		6.096***	(4.48)	6.086***	(4.51)	6.080***	(4.51)
Birth cohort: 1956-1960		3.557	(1.64)	3.494	(1.61)	3.654*	(1.68)
Birth cohort: 1961-1965		1.909	(1.07)	1.928	(1.08)	1.914	(1.07)
Birth cohort: 1966-1970		1.788	(1.30)	1.793	(1.30)	1.822	(1.32)
Birth cohort: 1971-1975		2.325*	(1.95)	2.304*	(1.93)	2.325*	(1.95)
Birth cohort: 1976-1980		0.275	(0.27)	0.273	(0.27)	0.293	(0.28)
Birth cohort: >1980		Ref		Ref		Ref	
Age * (Birth cohort: ≤1955)		-0.193***	(5.13)	-0.193***	(5.15)	-0.193***	(5.16)
Age * (Birth cohort: 1956-1960)		-0.139**	(2.50)	-0.138**	(2.48)	-0.142**	(2.55)
Age * (Birth cohort: 1961-1965)		-0.096*	(1.85)	-0.097*	(1.86)	-0.097*	(1.86)
Age * (Birth cohort: 1966-1970)		-0.090*	(1.92)	-0.090*	(1.93)	-0.091*	(1.95)
Age * (Birth cohort: 1971-1975)		-0.105**	(2.28)	-0.105**	(2.27)	-0.106**	(2.28)
Age * (Birth cohort: 1976-1980)		-0.037	(0.79)	-0.037	(0.79)	-0.038	(0.81)
Number of siblings		-0.146***	(6.37)	-0.206***	(6.56)		
Number of sisters				0.100***	(2.79)		
Number of older brothers						-0.190***	(5.39)
Number of younger brothers						-0.194***	(6.07)
Number of older sisters						-0.069**	(2.20)
Number of younger sisters						-0.115***	(4.13)
Birth order		0.025	(1.03)	0.027	(1.10)		
Parental characteristics							
Female		0.613**	(2.51)	0.644***	(2.64)	0.654***	(2.68)
Age		0.075**	(2.02)	0.071**	(1.98)	0.072**	(2.02)
Age squared (10^e-2)		-0.036	(1.22)	-0.033	(1.14)	-0.032	(1.14)
Married		0.300	(1.26)	0.338	(1.42)	0.338	(1.42)
Education	None	Ref		Ref		Ref	
	Primary 4 years	0.310**	(2.41)	0.311**	(2.45)	0.310**	(2.44)
	Primary 8 years	0.704***	(5.17)	0.697***	(5.21)	0.695***	(5.21)
	Secondary or vocational	1.757***	(11.48)	1.750***	(11.69)	1.747***	(11.70)
	University or postgraduate	2.441***	(12.28)	2.430***	(12.30)	2.427***	(12.30)
Religion	Muslim	Ref		Ref		Ref	
	Orthodox	0.615***	(5.09)	0.598***	(4.96)	0.599***	(4.96)
	Catholic	-0.021	(0.15)	-0.016	(0.12)	-0.016	(0.12)
	Other	0.413**	(2.06)	0.424**	(2.16)	0.427**	(2.18)
Region	Urban	Ref		Ref		Ref	
	Rural	-0.549***	(5.33)	-0.541***	(5.24)	-0.542***	(5.25)
Community variables							
Primary school in the community		0.145	(1.21)	0.154	(1.29)	0.155	(1.29)
Secondary school in the community		0.455***	(4.35)	0.453***	(4.33)	0.453***	(4.32)
Variance (B: between, W: Within, T: Total)							
B explained (% B exp. /T total)		1.442 (39.1%)		1.465 (39.5%)		1.466 (39.5%)	
W explained (% W exp. /T total)		0.034 (0.9%)		0.033 (0.9%)		0.034 (0.9%)	
B unexplained (% B unexp. /T total)		1.216 (32.9%)		1.211 (32.7%)		1.214 (32.7%)	
W unexplained (% W unexp. /T total)		1.000 (27.1%)		1.000 (27.0%)		1.000 (26.9%)	
Number of observations		6956		6956		6956	
Number of families		2091		2091		2091	
Log likelihood		-3382.2		-3378.2		-3377.5	

Source: LSMS Albania 2002.

Note: Random effects Probit models. Absolute values of t-statistics are in parentheses and levels of significance are 1% (***), 5% (**) and 10% (*).

Table 5. Gender, rural-urban and religion effects on the probability of having more than 8-years schooling

Variables	With gender crossed effects				By rural-urban status				By religion			
	Single		Crossed (* female)		Urban		Rural		Muslim		Non muslim	
	coef	t-test	coef	t-test	coef	t-test	Coef	t-test	coef	t-test	coef	t-test
Constant	-6.824***	(4.62)			-7.425***	(4.51)	-6.739***	(3.97)	-7.996***	(6.05)	-3.804	(1.57)
Child's characteristics												
Female	-0.979	(0.51)			0.553***	(7.37)	-0.158**	(2.28)	0.181***	(3.19)	0.169	(1.57)
Age	0.107***	(2.66)	0.051	(0.86)	0.169***	(4.09)	0.097**	(2.40)	0.115***	(3.61)	0.170***	(2.61)
Birth cohort dummies	YES		YES		YES		YES		YES		YES	
Age * Birth cohort dummies	YES		YES		YES		YES		YES		YES	
Number of siblings	-0.186***	(5.01)	-0.063	(1.53)	-0.297***	(6.29)	-0.150***	(3.62)	-0.230***	(6.39)	-0.158**	(2.30)
Number of sisters	0.125***	(2.90)	-0.054	(1.14)	0.147***	(2.63)	0.058	(1.17)	0.144***	(3.40)	-0.062	(0.80)
Birth order	-0.024	(0.76)	0.107**	(2.57)	0.016	(0.44)	0.051	(1.54)	0.007	(0.25)	0.133**	(2.40)
Parental characteristics	Ref		Ref		Ref		Ref		Ref		Ref	
Female	0.541*	(1.89)	0.448	(0.97)	0.819**	(2.28)	0.573	(1.60)	0.823***	(2.98)	0.169	(0.32)
Age	0.080*	(1.79)	-0.001	(0.02)	0.068	(1.31)	0.062	(1.20)	0.110***	(2.70)	-0.034	(0.47)
Age squared (10e-2)	-0.039	(1.11)	0.001	(0.03)	-0.023	(0.54)	-0.031	(0.74)	-0.058*	(1.79)	0.029	(0.50)
Married	0.293	(1.05)	0.272	(0.60)	0.324	(0.93)	0.440	(1.27)	0.467*	(1.76)	0.074	(0.14)
Education None												
Primary 4 years	0.298**	(2.05)	0.096	(0.63)	0.307	(1.49)	0.374**	(2.14)	0.270*	(1.91)	0.614**	(2.18)
Primary 8 years	0.633***	(4.11)	0.209	(1.30)	0.638***	(3.18)	0.803***	(4.14)	0.758***	(5.04)	0.496*	(1.73)
Secondary or vocational	1.688***	(9.77)	0.302*	(1.68)	1.740***	(7.89)	1.820***	(8.01)	1.785***	(10.67)	1.738***	(5.39)
University or postgraduate	2.383***	(10.50)	0.436	(1.54)	2.439***	(9.26)	2.454***	(6.89)	2.494***	(10.83)	2.439***	(6.17)
Non-muslim (Ref: Muslim)	0.320***	(3.13)	0.092	(0.80)	0.404***	(3.19)	0.318**	(2.51)				
Rural (Ref: Urban)	-0.353***	(2.91)	-0.385***	(3.00)					-0.552***	(4.57)	-0.440**	(2.24)
Community variables												
Primary school in the community	0.168	(1.21)	-0.060	(0.40)			0.135	(1.06)	0.153	(1.05)	0.091	(0.40)
Secondary school in the community	0.391***	(3.19)	0.265**	(2.04)	0.305	(1.62)	0.609***	(4.88)	0.413***	(3.44)	0.810***	(3.74)
Variance (B: between, W: Within, T: Total)												
B explained (% B exp. /T total)	1.581 (40.0%)				1.102 (32.8%)		0.582 (20.3%)		1.321 (36.5%)		1.620 (44.2%)	
W explained (% W exp. /T total)	0.082 (2.1%)				0.052 (1.5%)		0.057 (2.0%)		0.038 (1.1%)		0.066 (1.8%)	
B unexplained (% B unexp. /T total)	1.295 (32.7%)				1.206 (35.9%)		1.234 (42.9%)		1.260 (34.8%)		0.983 (26.8%)	
W unexplained (% W unexp. /T total)	1.000 (25.3%)				1.000 (29.8%)		1.000 (34.8%)		1.000 (27.6%)		1.000 (27.3%)	
Number of observations	6956				3343		3613		5503		1453	
Number of families	2091				1143		948		1608		483	
Log likelihood	-3327.2				-1579.4		-1764.9		-2689.1		-670.7	

Source: LSMS Albania 2002.

Note: Random effects Probit models. Absolute values of t-statistics are in parentheses and levels of significance are 1% (***), 5% (**) and 10% (*).

Table 6. Spousal effects on the probability of having more than 8-years schooling

Variables		(1)		(2) Single		(2) Crossed (* female)	
		coef	t-test	coef	t-test	coef	t-test
Constant		-6.231***	(5.06)	-5.181***	(3.32)		
Child's characteristics							
Female		0.175***	(3.22)	-2.481	(1.24)		
Age		0.120***	(4.07)	0.099**	(2.39)	0.056	(0.91)
Birth cohort dummies		YES		YES		YES	
Age * Birth cohort dummies		YES		YES		YES	
Number of siblings		-0.229***	(6.71)	-0.193***	(4.82)	-0.069	(1.50)
Number of sisters		0.110***	(2.78)	0.115**	(2.43)	-0.037	(0.70)
Birth order		0.038	(1.39)	-0.008	(0.21)	0.092**	(1.96)
Parental characteristics							
Age		0.061	(1.51)	0.041	(0.82)	0.047	(0.77)
Age squared (10^e-2)		-0.023	(0.70)	-0.008	(0.20)	-0.033	(0.69)
Father's None		Ref		Ref		Ref	
Education	Primary 4 years	0.183	(1.17)	0.135	(0.75)	0.067	(0.35)
	Primary 8 years	0.700***	(4.28)	0.599***	(3.17)	0.249	(1.23)
	Secondary or vocational	1.901***	(9.73)	1.802***	(7.93)	0.305	(1.24)
	University or postgraduate	2.620***	(10.79)	2.536***	(8.99)	0.389	(1.14)
Difference	Father's educ. > mother's educ.	-0.379***	(3.90)	-0.438***	(3.79)	0.119	(0.90)
in education	Father's educ. = mother's educ.	Ref		Ref		Ref	
	Father's educ. < mother's educ.	0.345***	(2.73)	0.284*	(1.91)	0.195	(1.13)
Religion	Muslim	Ref		Ref		Ref	
	Orthodox	0.588***	(4.57)	0.497***	(3.25)	0.234	(1.25)
	Catholic	-0.020	(0.15)	0.078	(0.50)	-0.221	(1.26)
	Other	0.343*	(1.65)	0.227	(0.91)	0.276	(0.98)
Region	Urban	Ref		Ref		Ref	
	Rural	-0.365***	(3.34)	-0.229*	(1.74)	-0.323**	(2.24)
Community variables							
Primary school in the community		0.222*	(1.74)	0.190	(1.27)	0.017	(0.10)
Secondary school in the community		0.444***	(4.00)	0.326**	(2.44)	0.320**	(2.21)
Variance (B: between, W: Within, T: Total)							
B explained (% B exp. /T total)		1.501 (41.8%)		1.643 (42.8%)			
W explained (% W exp. /T total)		0.034 (0.9%)		0.081 (2.1%)			
B unexplained (% B unexp. /T total)		1.052 (29.3%)		1.117 (29.1%)			
W unexplained (% W unexp. /T total)		1.000 (27.9%)		1.000 (26.0%)			
Number of observations		5612		5612			
Number of families		1728		1728			
Log likelihood		-2712.6		-2644.4			

Source: LSMS Albania 2002.

Note: Random effects Probit models. Absolute values of t-statistics are in parentheses and levels of significance are 1% (***), 5% (**) and 10% (*).

Table 7. Random effects estimates for educational attainment of the child

Variables		Random effect Probit		Random effects ordered Probit		Random effects ordered Probit with censoring	
		coef	t-test	coef	t-test	coef	t-test
Constant		-7.145***	(4.23)				
Cutoff level μ_1				5.611***	(4.36)	5.073***	(5.55)
Cutoff level μ_2				6.682***	(5.18)	6.129***	(6.69)
Cutoff level μ_3				7.603***	(5.89)	7.045***	(7.67)
Child's characteristics							
Female		0.117**	(2.09)	0.049	(1.16)	0.094**	(2.40)
Age		0.097**	(2.46)	0.059*	(1.94)	0.070***	(3.53)
Birth cohort dummies		YES		YES		YES	
Age * Birth cohort dummies		YES		YES		YES	
Number of siblings		-0.212***	(6.10)	-0.172***	(6.68)	-0.171***	(7.07)
Number of sisters		0.105***	(2.63)	0.089***	(2.82)	0.088***	(3.04)
Birth order		0.022	(0.82)	-0.000	(0.01)	-0.001	(0.04)
Parental characteristics							
Female		0.945***	(3.21)	0.789***	(3.58)	0.597***	(3.09)
Age		0.081*	(1.66)	0.071*	(1.91)	0.065**	(2.24)
Age squared (10^e-2)		-0.039	(1.05)	-0.037	(1.29)	-0.032	(1.41)
Married		0.564**	(2.00)	0.500**	(2.36)	0.345*	(1.83)
Education	None	Ref		Ref		Ref	
	Primary 4 years	0.296**	(2.20)	0.133	(1.24)	0.136	(1.30)
	Primary 8 years	0.718***	(5.00)	0.484***	(4.33)	0.436***	(4.07)
	Secondary or vocational	1.828***	(11.00)	1.277***	(10.48)	1.251***	(10.87)
	University or postgraduate	2.614***	(10.89)	1.870***	(12.57)	1.848***	(13.24)
Religion	Muslim	Ref		Ref		Ref	
	Orthodox	0.686***	(4.80)	0.456***	(4.86)	0.428***	(4.95)
	Catholic	0.005	(0.03)	0.023	(0.19)	-0.017	(0.15)
	Other	0.468**	(2.01)	0.231	(1.47)	0.228	(1.60)
Region	Urban	Ref		Ref		Ref	
	Rural	-0.592***	(4.86)	-0.412***	(4.52)	-0.423***	(5.14)
Community variables							
Primary school in the community		0.163	(1.20)	0.153	(1.34)	0.153	(1.50)
Secondary school in the community		0.330***	(2.68)	0.261***	(2.79)	0.325***	(3.84)
Variance (B: between, W: Within, T: Total)							
B explained (% B exp. /T total)		1.542 (39.9%)		0.859 (31.4%)		0.956 (33.9%)	
W explained (% W exp. /T total)		0.033 (0.9%)		0.019 (0.7%)		0.028 (1.0%)	
B unexplained (% B unexp. /T total)		1.293 (33.4%)		0.854 (31.3%)		0.835 (29.6%)	
W unexplained (% W unexp. /T total)		1.000 (25.8%)		1.000 (36.6%)		1.000 (35.5%)	
Number of observations		5615		5615		6956	
Number of families		1701		1701		2091	
Log likelihood		-2725.9		-5594.1		-6348.0	

Source: LSMS Albania 2002.

Note: Random effects Probit models. Absolute values of t-statistics are in parentheses and levels of significance are 1% (***), 5% (**) and 10% (*).