Educational Gap and Family Structure in Uruguay

Cid, Alejandro

Universidad de Montevideo

October 2007

Educational Gap and Family Structure in Uruguay

Alejandro Cid[1]

October 2007.

Abstract

In this paper we use household survey data to study the determinants of the children's educational achievement in Uruguay. As an indicator of this educational achievement, we built the "educational gap" which is the difference between expected years of schooling of a child and actual years of schooling. Among the determinants, we introduced indicators of family environment, focusing on the impact of the parents' marital status on their children educational attainment. In particular, the results suggest positive influence of having married parents on daughter's educational outcomes, after controlling for household background variables such as parents' education, income per capita, wealth and number of children.

JEL classification: J12, J13, C14, C34, I21

Keywords: censored data, treatment evaluation, education, family instability, cohabitation.

1. Introduction

Previous investigations analyse the possible determinants of schooling gap[2] -a censored variable- but not few methodological problems arise. When data are censored, OLS regression can provide misleading estimates. But also using traditional maximum likelihood

[1]Alejandro Cid is Professor of Economics, Universidad de Montevideo.

I thank Fernando Borraz for his helpful comments.

[2]Schooling gap indicates the relative lag behind the age-appropriate schooling level. In other words, schooling gap is defined as the difference between expected years of schooling (number of years of schooling a child would have under assumption of an initial enrollment age of 6 and completing one grade per year without grade repetition) and actual years of schooling, as a proportion of expected years of education.

methods for censored models, like Tobit one, could not be appropriate because of the lack of normality of the error terms or due to the existence of heteroskedasticity. In order to overcome these requirements of strong distributional assumptions, many semiparametric alternatives were developed. To examine these methodological issues, this paper applies and compares different methods to estimate the determinants of the educational gap.

There are several papers that examine the possible relationship between family structure and child well-being (Axinn and Thornton, 1993; Brown, 2004; Manning and Lichter, 1996; McLanahan, 1985; Raley and Wildsmith, 2004). Brown (2004) provides an extensive summary on the emerging literature on the effects of cohabiting families on children residing with them, and suggests children's academic performance is negatively associated with cohabitation. Brown sets the hypotheses that both the impermanence of cohabiting unions and their incomplete institutionalization (unclear family roles, rights, and obligations) set the stage for a family environment that may undermine child development.

We use the schooling gap as one proxy of child well-being. In the case of two biological-parent houses in Uruguay using the Continuous Households Survey (Encuesta Continua de Hogares, ECH), the fact of having married parents seems to be a significant determinant of the daughters' schooling outcomes. In fact, having married parents contributes to the decrease of the girls' educational gap in the year 2001. For example, taking into account the results of the Symmetrically Censored Least Squares Estimator, the fact of having married parents reduces the educational gap of the daughters by 0.123. The ECH for year 2001 was selected because of two reasons: it contains recent data and it is previous to the year 2002 in which Uruguay experienced one of the greatest adverse economic shocks of its history.

2. Background

It could be observed some main trends in estimating the determinants of school achievement. One trend is to estimate the educational gap using OLS. Take for example the case of Behrman, Birdsall and Székely (2000) who estimate the effects of parental education and household income on the schooling gap of their children. With respect to Brazil, they find that maternal education has a slightly stronger effect than paternal education in 1995. But the authors use OLS, despite the fact that the gap measure is censored. Also Andersen (2001) estimates the schooling gap by OLS. He finds that schooling gap is negatively related with the adult household income per capita, with the maximum of father's and mother's education, with the age of the head of household at the birth of the teenager and with the fact of living in urban areas. On the contrary, the schooling gap is positively related with the presence of a younger sister or brother, and with the fact of being a teenager that is not the son or the daughter of the head of the household.

Other approach is to use Probit models. For instance, Fishback and Baskin (1991) compare the level of educational achievement between black and white children in the first part of the twentieth century. They assume that child's literacy (a variable which was valued at one if household members said the child could write and zero if (s)he could not write) is a function of school inputs, household educational inputs, the time devoted to learning, and characteristics like de child's age and sex. Lacking direct measures of income and wealth, they include a variety of indirect measures: the age of the head and the spouse to measure their position in the life cycle; home ownership, and an index of occupational status. They estimate the effect of each determinant using maximum likelihood probit analysis. The analysis shows that the largest contributor to the black-white literacy gap was the difference between the educations of black and white parents. The estimation results also show that the length of the school term was a key school input for developing basic literacy, and the higher parents occupational status contributed to the child's literacy too.

Another trend in previous literature is to use a Tobit procedure. For example, Psacharopoulos and Arriagada (1989) estimate educational attainment among 7 to 14 years old employing a Tobit model. They find that maternal education has a stronger effect than paternal education on boys and girls taken together. However, boys and girls are pooled, so it is not possible to make any further gender comparisons. Margo (1987) specifies a model of school attendance and constructs an equation as the outcome of a household utility maximization. Parents derive utility from consumption of markets goods and home production by household members and from their children's schooling. How frequently a child attends school depends on the characteristics of the parents and the child; on the availability of schooling, quantity and quality; and on the returns to schooling compared with other uses of the child's time, which may vary with the household location. The dependent variable is the number of months of school attended in the census year. Because many children did not attend school the dependent variable is censored at zero, and Tobit analysis is used. Margo's results show that the presence of a child under age 5 in black families lowered school attendance among older siblings, presumably by increasing parental demands on the sibling's home time. Margo also finds that longer school terms and smaller class sizes encourage children of both races to attend school more frequently, but the effect was larger among blacks. And better-trained teachers also increase attendance in the black schools. Finally Margo observes that urban children of both races attend school significantly longer than rural children, and the effect is larger among blacks. Saha (2005) also use a Tobit model, focus on one age cohort and restricts the analysis to children who turn 15 during the survey years. The explanatory variables used by Saha are: number of younger siblings, household size, parental age, rural dummy, income, mother's education, father's education. Saha found that maternal and parental education, and the household income were positively correlated with the educational attainment, while the household size, the fact of living in rural areas and the number of younger siblings were negatively correlated with the school outcomes. Also, Saha found that the maternal effects differed by household type: maternal education in two-parents household widened the gap between son and daughters educational attainment, and, in sharp contrast, maternal education in female-headed households contributed to the decrease in the gender gap.

Finally, we could consider the treatment of endogeneity in previous literature. Case and Deaton (1999) examine the relationship between school inputs, particularly the pupil-teacher ratio, and various measures of educational outcomes, including educational attainment, enrolment rates, the reason for not being in school, educational expenditures, and test scores. They present the results of a series of regressions in which the pupil-teacher ratio (or the presence of other facilities) is an explanatory variable. Among the other controls are age, urbanization, sex, and various measures of family background, such as whether the household is headed by a woman, household size, the educational attainment of the head, and the logarithm of total expenditure per capita. They think of head's education as both a direct input into the educational process and a measure of household resources. In one of the regressions (estimated using OLS and Two Stage Least Squares with robust standard errors), the dependent variable is educational attainment measured as years completed. The reason to introduce Two Stage Least Squares is this: the pupil-teacher ratio for Blacks may be affected by household characteristics. Thus the estimated effects of the pupil-teacher ratio may be coming from the influence of unmeasured household characteristics. They consider possible instrumentation using the racial composition of magisterial districts (they checked that pupil-teacher ratios can be predicted by racial composition). They found that the TSLS results were very similar to the OLS results. Case and Deaton show that gender and household characteristics have important effects in the regressions. Black female students have on average about half a year of educational attainment more than the Black male students, and among Black students there are the expected positive effects of household resources and of education of the household head. Head's education is a strong predictor of educational attainment among both Blacks and Whites. Controlling for household background variables, they find strong and significant effects of pupil-teacher ratios on enrollment, on educational achievement, and on test scores for numeracy. Bjorklund et al (2005) focus on children who live with both biological parents and analyze whether marriage confers any educational advantages to children that cohabitation does not, for the case of Sweden. They use a natural experiment, namely the marriage boom in Sweden in the last two months of 1989, created by the reform of the widow's pension system (those who were married by the end of 1989, would be entitled to widow's pension if their husband died), to identify the causal effect of marriage on child outcomes. This experiment enables the authors to compare educational outcomes for children whose parents married in November and December 1989 to those of children whose parents were already married and to those of children whose parents continued to cohabit. They find that children whose parents married in the end of 1989 had similar educational outcomes than children of cohabiting parents which suggest some doubts on the direct causation of legal marriage on children educational outcomes. For comprehensive handling of the problem of endogeneity, we should refer to Francesconi et al. (2006). They analyse the impact on schooling outcomes of growing up in a family headed by a single mother. They test the hypothesis that a non-intact family in Germany is associated with worse educational outcomes, and employ propensity score matching models, mother fixed effects and quasi experimental models, and models based on comparisons between individuals whose fathers died, divorced, or remained married. The principal schooling outcome analysed is whether an individual has educational qualifications to university entrance level. They find that although almost all the point estimates indicate that non-intactness of family structure has an adverse effect on schooling outcomes, confidence intervals for estimated effects are wide so the data are consistent with the impact of family structure being zero as well as adverse. About the possible presence of endogeneity, they argue that there's disagreement about whether the family structure is causal: lone parenthood may be correlated with other socioeconomic disadvantages, and so inferior outcomes may arise from (potentially unobserved) factors other than a parent's absence.

The following section gives details about the data used in the empirical application. In section 4 different methods for censored regression models are defined, while Section 5 presents the results and makes comparisons between the different methods. The final section discusses conclusions, limitations of the approach adopted here and points out some issues for further research.

3. Data

We use cross-sectional data of the year 2001 from Continuous Households Survey (Encuesta Continua de Hogares - ECH) which includes socio-economic information of people and households. ECH is conducted by the National Institute of Statistics (Instituto Nacional de Estadística - INE) of Uruguay and is an urban representative sample with a total sample size of 57394 observations. We take into account only sons and daughters with ages which fall in the interval [8,14] and who live with both biological parents (a sample size of 4067 observations). We focus in the interval [8,14] because -as it is observed in table 1- the proportion of children with positive schooling gap is nearly zero for children of 6 and 7 years old (the initial enrolment age in Uruguay is usually 6), and children with 14 years old or above are considered to be part of the labor force by the ECH.

Table 1 -

age	percentage of children with education-gap = 0
6	99.50
7	99.17
8	67.77
9	63.49
10	64.76
11	64.44
12	59.64
13	60.67
14	54.96

3.1 The dependent variable

The dependent variable, educational gap of the sons and daughters, indicates the relative lag behind the age-appropriate schooling level. It is computed as (under the assumption of an initial enrolment age of 6):

$$educ\_gap = \left(\frac{age - 6 - years\_of\_schooling}{age - 6} \right)$$

In other words, educational gap is defined as the difference between expected years of schooling (number of years of schooling[3] a child would have under assumption of an initial

[3] The variable "years of schooling" is measured as years completed both in primary and secondary school plus one. The reason to add the value "one" is that the survey (ECH) used does not provide information about the child's birthday and this is a problem in order to estimate the "schooling gap". In our country, a child is able to start primary school if (s)he is at least 6 years old before the 10th. of May. Take for example that one child with age 7 could claim in the survey that she has 0 year completed of schooling (thus

enrolment age of 6 and completing one grade per year without grade repetition) and actual years of schooling, as a proportion of expected years of education.

3.2 Family structure as a regressor

As it is stated in Section 1, previous research for other countries suggests some linkage between family structure and children school engagement. For this reason, this empirical application introduces -as a regressor for children educational gap- parents' marital status: a binary indicator variable which takes the value one if the parents are married, and zero in the case of cohabitation. This paper concentrates in these two types of family structure because of the increasing rate of cohabitation during the last thirty years (Brown, 2004; Raley and Wildsmith, 2004; Cid, Presno and Viana, 2004; Manning and Lichter, 1996). As an example of this trend, consider that, in Uruguay, the proportion of informal unions in the total of couples rose from 7.65 percent in 1963 census to 16.45 percent in 1996 census and this augmentation occurred basically in the younger age groups. For example, for the 15-19 age group this ratio is multiplied by more than three times in this period.

Introducing family explanatory variables pretends to stimulate further research on this topic which could be fruitful to improve our knowledge of the causes of the low educational achievements in our country. Filgueira, Filgueira and Fuentes (2003) states that Latin American countries have invested considerable economic resources in order to improve their educational supply, particularly in terms of school infrastructure, human and material resources, and innovative strategies to make schools more appealing to students. However, children academic performance remains a daunting challenge because of great drop-out rates, low grade completion and low schooling rates. Filgueira et al. observe that the key to this failure seems to be not on the supply side but on the demand side: little is known regarding

a schooling gap of 100 percent). But her birthday is the 20th. of May so she started primary school at 6 years old (as early as she was able to) but the survey was executed on August when she is 7 and ECH says that she has 0 year completed and in fact her educational gap is zero. To sum up, adding the value "one" to the years completed at school, we are able to guarantee that every child with an educational gap greater than zero has really a gap. It is important because, precisely, we wish to analyse the determinants of this gap.

how and why the targeted population behaves as it does, and thus, the primary focus of diagnosis and policy should go from supply to demand. And, precisely, in the demand side of education, the family could play a crucial role.

3.3 Other regressors for the educational gap

Seeking for the determinants of the educational gap, the explanatory variables also included in this paper are:

log household income per capita, entered linearly and quadratically: Brown (2004) states that poverty is closely linked to different features of child well-being like school outcome. Saha (2005) believes that, in the presence of credit constraints, poorer families are less able to pay for the direct costs of education, such as books and transportation; and poorer families are also more likely to send children into the work force to supplement family income;

subjob: indicates if the principal job of the mother and/or the father is an informal one (e.g., without social security in case of being ill or unemployed). An informal employment could imply job instability and thus could create worse household environment to children school engagement.

mother inactive and father inactive: these dummy regressors indicate if the mother or the father are not employed and not seeking for an employment (for example, the mother spends her time studying in order to complete her undergraduate degree, and looking for the children and the house). With these variables we intend to measure the closeness of the parent-child relationship. Datcher-Loury (1988) observes that greater child care time of highly educated but not of less well-educated mothers significantly raises offspring years of study.

mother education and father education, entered linearly and quadratically: each one shows the number of completed and approved years of education since Primary School. It is expected that, for example, children whose parents have a university degree are more engaged with school than those whose parents only have few completed years of Primary School (see Brown, 2004).

quantity of children with age below 15: socioeconomic literature (see, for instance, Becker [1988] or Saha [2005]) suggests a negative relationship in the short run[4] between number of children and parents resources per capita which could imply worse school engagement.

quantity of people with age above 59: the presence of grandparents in household composition could have a positive effect on children's school outcomes because of the greater guidance and supervision or the spill over effects of more contact with the adults. In the same sense, this research included home-aid: a binary regressor coded one for the presence of an additional adult at home which helps with homecare (laundry and meal preparation, etc.).

private children education: a proxy of education quality. Heckman and Rubinstein (2001) quote the conjectures that the decline in discipline in some public schools could be a major source of their failure on children's school engagement, and that the greater effectiveness of some private schools could come in producing more motivated and self-discipline students.

scholarship: a binary regressor with the value one in the case of a child with income from a scholarship. It could be expected that someone with a grant should show better academic performance.

public job: a dummy variable coded one if the mother and/or the father have a public job, and zero otherwise. The hypothesis is that parents public job could be an indicator of economic stability, thus it can influence positively children education.

[4]"Ironically, when I began to work on population studies, I assumed that the accepted view was sound [that a higher population growth implies lower standard of living]. I aimed to help the world contain its "exploding" population, which I believed to be one of the two main threats to humankind (war being the other). But my reading and research led me into confusion. I arrived at a theory implying that population growth has positive effects in the long run, although there are costs in the short run" (Simon, 1998)

remittances: this regressor pretends to capture the conjecture that child human capital decisions could be positively related with the fact of having a family member working abroad. McKenzie and Rapoport (2005) observe that previous research has suggested the potential of remittance income to improve access to education to the poor. They also state that a new literature has emphasized a possible link between expectation of future migration and current schooling decisions: education is needed to migrate, and since income abroad is much larger than at home, this raises the potential returns to schooling.

number of people with income at home: the hypothesis is that the larger number of individuals at home with a personal income (salary, profits, pensions, etc.), the greater the closeness of children to real world: the offspring experiment the need of being educated to cope with the market.

absolute wealth: the ECH provides information about thirteen comfort goods that each household could have: hot water heater, electric tea kettle, refrigerator, color television, cable TV service, VCR player/recorder, washing machine, dishwasher, microwave, computer, internet connection, automobile for personal use, telephone service. These goods could show different levels of wealth. For each comfort good i, we have constructed a dummy variable d_i which takes value 1 if the house has this good or service, and 0 otherwise. Then we have developed the index "wealth" $= \frac{1}{13}\sum_{i=1}^{i=13} d_i$

relative wealth: besides the previous wealth index which is an absolute indicator of wellbeing, we have built also an index of relative wealth using the comfort goods information of the ECH. For each comfort good i, we have constructed a dummy variable d_i which takes value 1 if the house has this good or service, and 0 otherwise. Thus, we have developed this indicator in two steps:

1st) the sample mean of each d_i is calculated;

2nd) "relative wealth index" $= \dfrac{\sum_{i=1}^{i=13}[1 - \text{mean}(d_i)]\, d_i}{\sum_{i=1}^{i=13}[1 - \text{mean}(d_i)]}$

(therefore, as an indicator of relative welfare, it can be seen in the formula above that greater average of people in the sample having a comfort good implies less relative welfare).

Besides quadratic and interactive forms of these explanatory variables, we also included among the regressors dummies with the purpose of controlling potential e¤ects of population density and economic situation of the region of residence, or the possible incidence of the sector of the economy in which the parents are employed.

3.4 Summary Statistics

The Continuous Household Survey (ECH) of the year 2001 provides information of 6.384 children in the interval of age [8,14]. Among them, 4.067 are children living with both biological parents (so they represent a 64 percent of the children of this interval). Other 1.479 children live with his/her biological father or mother (alone or with a step-father/mother). Other 665 children claim to live in a household where the grandfather/grandmother is the person with more authority in the house (the "chief" in terms of the ECH). Other 114 children claim to be only "other relatives" while 59 children describe themselves as no relatives at all.

Table 2 - Descriptive statistics for daughters living with both biological parents in the interval of age [8,14]

*** means are statistically di¤erent at 1%; ** at 5%; * at 10%

	Cohabit	Married		
	(258 obs)	(1797 o.)	Di¤erence	p-value
father age	43.09	43.13	-0.04	0.936
mother age	38.45	39.75	-1.3***	0.008
child age	10.60	11.09	-0.49***	0.000
child educational gap	0.147	0.072	0.075***	0.000
father education	6.86	10.02	-3.16***	0.000
mother education	6.98	10.37	-3.39***	0.000
people living at home	5.85	5.13	0.72***	0.000
n. of children age < 15	2.77	2.09	0.68***	0.000
n. of people age > 59	0.08	0.09	0.01	0.779

Note: This table includes the results of t-tests on the equality of means allowing the variances to be unequal. "Cohabit" column contains the daughters who live with cohabiting parents; "Married" column contains the daughters who live with married parents;

Table 3- Descriptive statistics for sons living with both biological parents in the interval of age [8,14]

*** means are statistically di¤erent at 1%; ** at 5%; * at 10%

	Cohabit	Married		
	(235 obs)	(1777 o.)	Di¤erence	p-value
father age	41.74	43.23	-1.48**	0.023
mother age	37.04	39.95	-2.91***	0.000
child age	10.49	11.19	-0.70***	0.000
child educational gap	0.155	0.094	0.06***	0.000
father education	7.01	9.92	-2.91***	0.000
mother education	7.44	10.22	-2.78***	0.000
people living at home	6.01	5.08	0.93***	0.000
n. of children age < 15	2.94	2.05	0.89***	0.000
n. of people age > 59	0.12	0.09	0.03	0.265

Note: This table includes the results of t-tests on the equality of means allowing the variances to be unequal. "Cohabit" column contains the sons who live with cohabiting parents; "Married" column contains the sons who live with married parents;

The tables 2 and 3 show the means of individual and household characteristics by parental marital status and by child gender. The cause of presenting di¤erent tables for boys and girls is that in developing countries (Saha, 2005), older children, usually girls, are often responsible for home production and care of younger siblings. And these tasks could mean less time to devote to school work and, then, worse academic performance.

Descriptive factors to note are the statistically signi…cant di¤erences between two-biological cohabiting parents and married parents. Cohabiting parents are younger and have less completed years of schooling. Their children are younger but have greater schooling gap. Another feature to mark is that cohabiting households have bigger family sizes and a larger number of younger siblings. In spite of these di¤erences between the children who live with married parents and those who live with cohabiting parents, we have to bear in mind that in order to asses properly the determinants of the di¤erent educational gap, we ought to execute econometric analysis (as we do in the next section).

Table 4 - Educational gap - children with age among [8,14]

	Bio Paren. Cohab		Bio Paren. Marr	
	Girls (258)	Boys (235)	Girls (1797)	Boys (1777)
Median	0.111	0.125	0	0
Mean	0.147	0.155	0.072	0.094
Std. Dev.	0.166	0.172	0.127	0.145
Variance	0.028	0.029	0.016	0.021
Skewness	0.799	0.810	2.121	1.855
Kurtosis	2.718	2.816	8.543	6.925

As it can be observed in the graphics below, the educational gap is skewed right for all the children with age among [8,14] and it is more marked for the children who live with married parents. Also, in reference to the Kurtosis analysis, it can be perceived in the graphics below that the peakedness is more pronounced for the children who live with married parents because the proportion of children with a educational gap near 0 is greater among the children living with married parents.

Distribution of the educational gap - Normal density overlaid for comparison

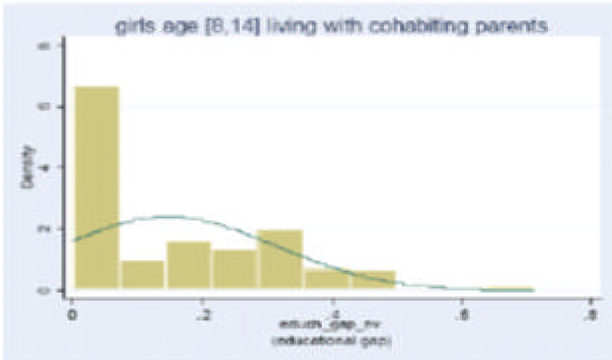

Figure 1:

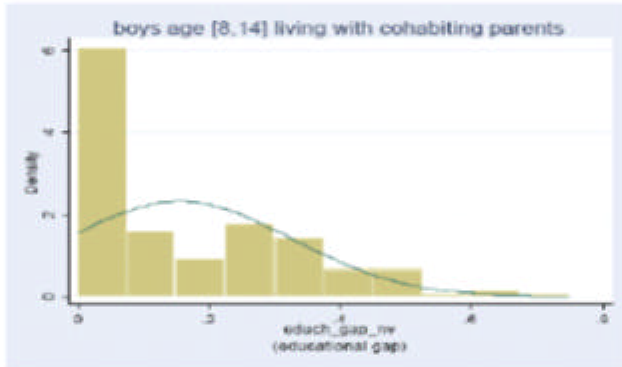

Figure 2:

Figure 3:

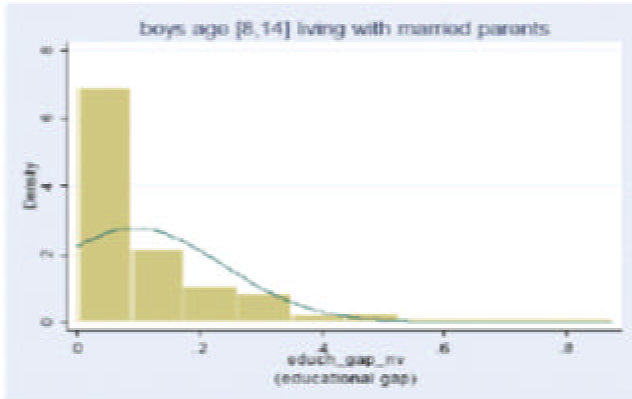

Figure 4:

4. Methods of Estimation

Binary Probit Model

One possibility is to define the educational gap, y_i, as a binary response variable, taking on the values zero when the actual grade attainment does not lag behind the age-appropriate schooling level, and one otherwise. But allowing only binary response, we lose information about the relative lags and their possible determinants. In other words, y_i would take the value one both for a child of eleven years old and no grade attainment, and for a child of eleven years old and three grades completed: these children are actually different but our explained binary variable would give them the same weight.

Multinomial Ordered Models

Also it could be studied to apply Multinomial Ordered Models to test the determinants of the absolute schooling gap (take into account that considering only the absolute gap, that dependent variable could take only the integer values from 0 to 8 because children ages are in the interval [8, 14]).

In general for an m-alternative ordered model we could define (Cameron, 2005):

$$y_i = j \qquad \text{if } \alpha_{j-1} < y_i^* \leq \alpha_j$$

where $\alpha_0 = -\infty$ and $\alpha_m = \infty$. Let y be an ordered response taking on the values $\{1,2,...,J\}$ for some known integer J.

However, an m-choice Probit model requires numerical evaluation of an (m-1)-variate integral: and this is a problem since in this application m=9, while a trivariate normal integral is the limit for numerical methods. To cope with this problem, we use in this paper an Ordered Logit Model, which contains the limitation of the IIA (independence from irrelevant alternatives) assumption.

Also, the empirical application section of this paper includes a Binary Probit estimation: it could be useful in the comparison of the signs of the partial effects of each explanatory regressor within the different models results.

Tobit Model

The educational gap, y_i, is a doubly censored variable which takes on the value zero and one with positive probability. In other words, the dependent variable suffers from interval censoring: the values of the true dependent variable, y_i^*, are observed only if they fall within the interval [0,1].

Algebraically,

$$y_i^* = x_i'\beta + u_i; \qquad u_i|x_i \sim \text{Normal}(0; \sigma^2)$$

$$y_i = 0 \quad \text{if} \quad y_i^* \leq 0$$

$$y_i = y_i^* \quad \text{if} \quad 0 < y_i^* < 1$$

$$y_i = 1 \quad \text{if} \quad y_i^* \geq 1$$

where x_i is a K x 1 vector of observed regressors, β is a K x 1 vector of unknown regression coefficients to be estimated, u_i is an unobserved error.

Tobit assumptions

Heteroskedasticity and nonnormality result in the Tobit estimator $\hat{\beta}$ being inconsistent for β, and entirely changes the functional forms for $E(y|x; 0 < y_i^* < 1)$ and $E(y|x)$. Wooldrige (2002) observes that y_i^* should have a homoskedastic normal distribution and the variable y should be (roughly) continuous when $y > 0$. Thus the Tobit model is not appropriate for ordered responses. In the empirical application of this paper, we do a Tobit analysis with robust standard errors to cope with the possible existence of heteroskedasticity.

Normality was also tested using various procedures. Kernel density estimators were used to approximate the density f (residuals of robust TOBIT)[5] and a Normal density was overlaid for comparison.

[5] In the Tobit model, the dependent variable is educational gap and the regressors are the household characteristics (family structure, income, wealth, parents' education, quantity of children at home,...) and dummies controlling population density, the economic situation of the region of residence and the sector of the economy in which the parents are employed.

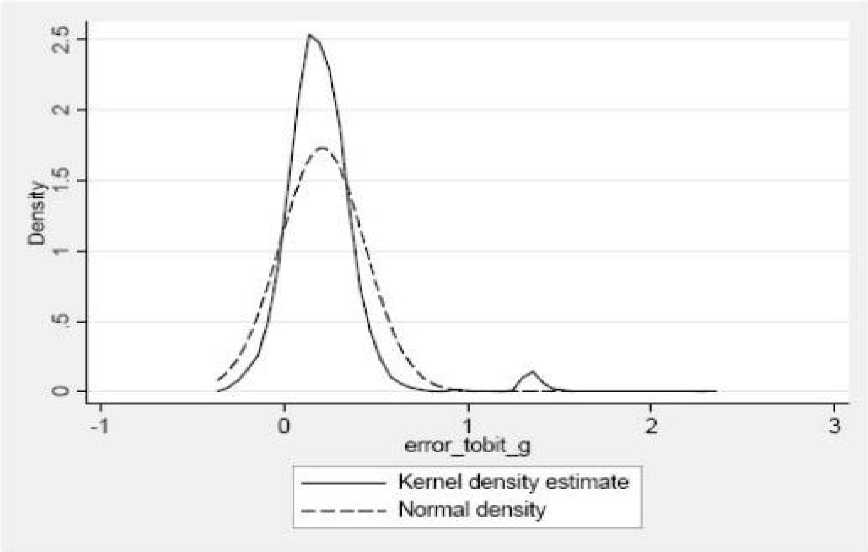

Figure 5:

Kernel density estimation of the residuals of the TOBIT model for education-gap (variables and results in Table 7) - Normal density overlaid for comparison - Only daughters with age among [8,14]

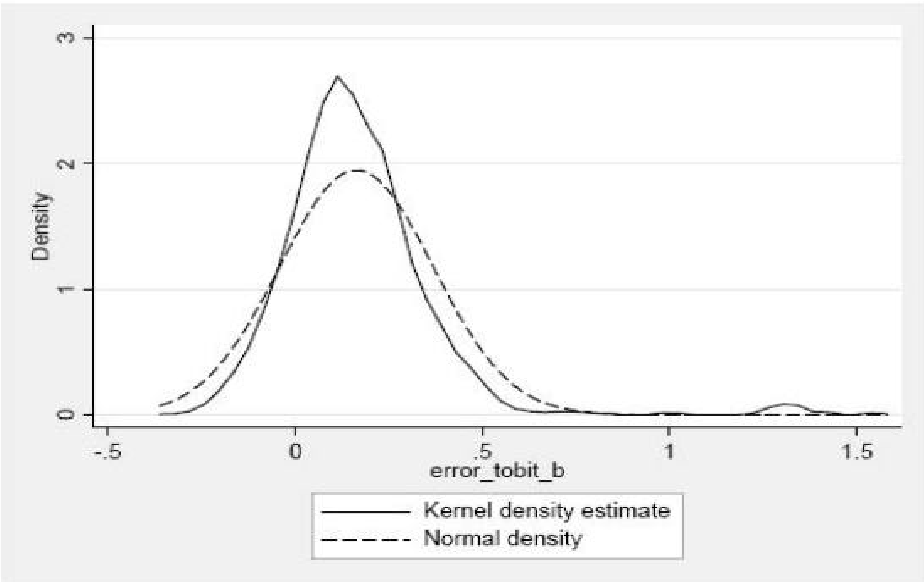

Figure 6:

Kernel density estimation of the residuals of the TOBIT model for education-gap (variables and results in Table 8)- Normal density overlaid for comparison - Only sons with age among [8,14]

Also, we tested normality in two ways: (1) a test based on a combination of a test on skewness and a test on kurtosis[6]; (2) the Shapiro-Francia test[7].

Table 5 - Skewness/Kurtosis tests for Normality -Children with age among [8,14]

	Variable	Pr(Skewness)	Pr(Kurtosis)	Prob>chi2
Girls	error_tobit	0.000	0.000	0.0000
Boys	error_tobit	0.000	0.000	0.0000

Both in case of girls and boys, we can reject the hypothesis that the error term is normally distributed. The source of the problem is both in skewness and kurtosis.

Table 6 - Shapiro-Francia W' test for normal data - Children with age among [8,14]

	Variable	Obs	W'	V'	z	Prob>z
Girls	error_tobit	2055	0.77732	239.489	8.277	0.00001
Boys	error_tobit	2012	0.82780	183.551	8.130	0.00001

The value reported under the W' is the Shapiro-Francia test statistics. The test also reports V' which is a more appealing index for departure from normality. The median value of V' is 1 for samples from normal populations. Large values indicate nonnormality. The 95% critical values of V', which depend on the sample size, are between 2.0 and 2.8. (There is no additional information in V' than in W' - one is just the transform of the other). Thus, we can reject that the error term is normally distributed.

So the distribution of the residuals (the estimation analogous to the error term) could be subject to nonnormality. If so, the Tobit estimators will not provide a consistent estimate. Therefore the common practice of employing Tobit estimators for estimating educational attainment as it can be seen in previous literature should be checked through to avoid inappropriate conclusions. Thus, relaxing distributional assumptions on the error terms and

[6]Tested with "sktest" command of STATA

[7]Tested with "sfrancia" command of STATA.

seeking for models which succeed with those weaker distributional assumptions is mandatory to obtain more proper results.

Semiparametric Censored Regression Models

As we have seen in the previous sections, Tobit models require some specifications of the error distribution: normality and homoskedasticity. In order to relax these requirements, the semiparametric approach has been proposed in the recent economic literature to provide consistent estimates for censored data. Thus one of the advantages of the semiparametric approaches for censored models is that estimators are consistent under weaker distributional assumptions. The attribute "semiparametric" in this model comes from the fact that the distribution of the errors u_i given the explanatory variables does not have a known parametric form.

This paper uses two semiparametric estimators for censored regression models: the censored least absolute deviations (CLAD) and the symmetrically censored least squares (SCLS) (for a summary, see Chay and Powell, 2001, or Cameron and Trivedi, 2005).

Censored Least Absolute Deviations Estimator

The censored least absolute deviations (CLAD) approach was developed by Powell (1984). The key distributional assumption of CLAD estimator is that $u|x$ has median zero, and this means weaker distributional assumptions than the Tobit model which need normal errors. CLAD estimator is a generalization of least absolute deviations estimation for the standard linear model. Thus, the CLAD estimator minimizes the sum of absolute deviations of y_i over all β:

$$S_T(\beta) = -\frac{1}{T}\cdot\sum_{t=1} |y_t - y_t^*|$$

where

$$y_t^* = 1 \quad \text{if} \quad x_i'\beta \geq 1$$

$$y_t^* = x_i'\beta \qquad \text{if} \qquad 0 < x_i'\beta < \infty$$

$$y_t^* = 0 \qquad \text{if} \qquad x_i'\beta \leq 0$$

Powell (1984) shows that CLAD $\hat{\beta}$ estimation is consistent, asymptotically normal and its asymptotic covariance matrix can be consistently estimated. Thus, tests of hypotheses concerning the unknown regression coefficient can be constructed, which are valid in large samples (precisely, in this paper we work with more than 4.000 observations: it could be seen as a "large sample"). Unlike estimation methods based on the assumption of Gaussian distributed errors terms, the CLAD estimator is consistent and asymptotically normal for a wide class of error distributions, and is also robust to heteroskedasticity.

Symmetrically Censored Least Squares Estimator

The symmetrically censored least squares (SCLS) approach was proposed by Powell (1986). This estimator is based on the assumption that errors are symmetrically (and independently) distributed around zero, so is less restrictive than Tobit requirements (normally distributed and homoskedastic errors). The SCLS estimators are consistent and asymptotically normal for a wide class of symmetric error distributions with heteroskedasticity of unknown form. But the assumption of SCLS that errors are symmetrically and independently distributed around zero is stronger than the zero median restriction of the CLAD estimator.

Powell (1986) states that if the underlying error terms were symmetrically distributed about zero, and if the latent dependent variables were observable, classical least squares estimation would yield consistent estimates of the parameter vector β. But due to the censoring, the observed dependent variable y has an asymmetric distribution. Powell's approach consists in symmetrically censoring the dependent variable y (it is usually known as a "symmetric trimmed" method) so that symmetry can be restored, and then the regression coefficients can be estimated by least squares. Symmetric censoring of the dependent variable implies that observations with values above the censoring point are dropped, and this means that there could be a loss of efficiency due to the information dropped in those observations. However this problem is reduced in the present paper because a relative large sample is used.

Treatment Evaluation and Parents' Marital Status

The typical dilemma in treatment evaluation involves the inference of a causal association between the treatment and the outcome. In this paper, we pay particular attention to the effects of parent's marital status on the educational attainment of their children. Thus, we observe $(y_i; x_i; D_i)$, $i = 1; ...; N$, where y_i is the educational gap, x_i represents the regressors, and D_i is the treatment variable and takes the value 1 if the treatment is applied (married parents) and is 0 otherwise (cohabitating union). The impact of a hypothetical change in D on y, holding x constant, is of interest. But no individual is simultaneously observed in both states: with the data available, it is not possible to view the same child both with married parents and with cohabitating ones. Moreover, the sample does not come from a randomized social experiment: it comes from observational data and the assignment of individuals to the treatment and control groups is not random. Hence, we estimate the treatment effects based on propensity score: this approach is a way to reduce the bias performing comparisons of outcomes using treated and control individuals who are as similar as possible (Becker and Ichino 2002). The propensity score is defined as the conditional probability of receiving a treatment given pre-treatment characteristics:

$$p(X) \equiv \Pr\{D = 1|X\} = E\{D|X\}$$

where $D = \{0; 1\}$ is the indicator of exposure to treatment and X is the vector of pre-treatment characteristics.

The propensity score was estimated in this application using a logit model [8]. Due to the probability of observing two units with exactly the same value of the propensity score is in principle zero since $p(X)$ is a continuous variable, various methods have been developed (for a summary, see Cameron et alt. 2005) to match comparison units sufficiently close to the treated units. So, after estimating $p(X)$ we employed Kernel Matching method[9].

[8] Applied with the Stata ado file "pscore" developed by Becker and Ichino (2002).

[9] This matching method was applied using the Stata ado files psmatch2 developed by E. Leuven and B.

5. Empirical Results

Results

Tables 7 and 8 present the results of these estimations for girls and boys respectively. In most cases, the signs of the signi..cant regressors come to be the expected ones (see Section 3). The number of children at home has operated in the hypothetical direction: this variable seems to worse children's school outcomes[10]. On the other hand, according to the previous tables, family's wealth, parents' education (especially mother's education) and, in the case of daughters, the fact of having married parents have positive and signi..cant e¤ects on o¤spring school engagement. Maternal education seems to have a greater positive e¤ect than father's education on the children educational attainment. This fact is consistent with the suggestions of the literature. A possible explanation (see Saha, 2005) is that mothers tend to spend more time directly assisting children with school work. As it could be seen in the tables below, considering CLAD results, each additional year of mother education reduces educational

Sianesi (2003) "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing".

[10]This not necessary mean that a higher population growth implies lower standard of living. Simon (1998) states that "population growth has positive e¤ects in the long run, although there are costs in the short run. (...) For the ..rst decades of its life, an additional child certainly is a burden not only on its parents but also on others. Brothers and sisters must do with less of everything except companionship. Tax payers must sough up additional funds for schooling and other public services. Neighbors hear more noise. During these early years the child produces nothing material, and the income of the family and the community is spread more thinly than if the baby had not been born. And when the child grows up and ..rst goes to work, jobs are squeezed a bit, and the output and pay per working person go down. All this clearly is an economic loss for other people. Just as surely, however, an additional person is also a boon. The child or immigrant will pay taxes later on, contribute energy and resources to the community, produce goods and services for the consumption of others, and make e¤orts to beautify and purify the environment. Perhaps most signi..cant for the more-developed countries is the contribution that the average person makes to increasing the e¢ ciency of production through new ideas and improve methods. The real population problem, then, is not that there are too many people or that too many babies are being born. The problem is that others must support each additional person before the person contributes in turn to the well-being of others".

gap of sons by 0.021 while each additional year of father education reduces educational gap of sons only by 0.008. One exception in the signs theoretically predicted seems to be the positive sign of quantity_of_people_with_income_at_home, perhaps suggesting that more members of the family on the market could mean smaller child care time, and thus worse children's educational outcomes. Regarding the family structure issue, the results suggest that girls living with two-biological married parents experience better outcomes on educational attainment. Considering CLAD results, the fact of having married parents reduces educational gap of daughters by 0.094. In the case of the sons, though the sign of the coe¢ cient is the same, it is not signi..cant in any estimation method used. Thus, negative cohabiting e¤ects on educational attainment seem to be more pronounced against daughters. A possible explanation is that instability of the cohabitation unions (Brown, 2004) has a deeper in‡uence on daughters with ages among [8,14] because of the di¤erent psychological characteristics of boys and girls at those ages. There's a rising literature in the psychological ..eld which discusses gender-speci..c learning di¤erences. For instance, Sax (2005) asserts that the brain of girls and boys develops di¤erently; the brain is wired di¤erently; girls hear better; and girls and boys respond to stress di¤erently: Sax argues that stress enhances learning in males and the same stress impairs learning in females. This last fact could be related with the girls' worse school outcome than the boys', as a consequence of the instable environment of cohabitation.

Table 7

Estimates of educational gap (all heteroskedasticity-robust); only girls among [8,14].

Number of observations: 2055 - (estimated standard errors in parentheses) - *** significant at 1%; ** at 5%; * at 10%

Dependent Variable: educational gap

Notes: in the case of CLAD and SCLS, some explanatory regressors were not included in order to avoid lack of computational convergence.

	OLS	PROBIT	TOBIT	OLOGIT	CLAD	SCLS
married parents	**-.069 (.031)****	-.289 (.318)	-.100 (.068)	**-1.03 (.551)***	**-.094 (.046)***	**-.123 (.099)***
quantity children age<15	.013 (.002)***	.073 (.026)***	.022 (.005)***	.154 (.043)***	.034 (.017)*	.037 (.015)*
quantity people age>59	-.010 (.008)	-.091 (.111)	-.025 (.025)	-.120 (.172)		.001 (.141)
quantity people with income	.010 (.004)**	.113 (.045)**	.023 (.010)**	.217 (.070)***	.039 (.020)*	.018 (.025)
log income per capita	-.035 (.065)	-.178 (.718)	-.024 (.162)	-.550 (1.27)	-.057 (.024)*	-.030 (.935)
(log income per capita)^2	.001 (.003)	-.006 (.045)		.021 (.079)		-.001 (.067)
father's education	-.006 (.004)	-.103 (.042)**	-.022 (.009)**	-.196 (.072)***	-.003 (.004)	.041 (.075)
(father's education)^2	.000 (.000)	-.001 (.002)		-.001 (.004)		-.003 (.006)
mother's education	-.018 (.005)***	-.120 (.046)***	-.032 (.010)***	-.237 (.082)***	-.021 (.011)*	-.074 (.074)*
(mother's education)^2	.000 (.000)	.001 (.002)		.004 (.004)		.002 (.008)
(father_educ)x(mother_educ)	.000 (.000)	.006 (.003)*	.001 (.001)	.007 (.006)		
(married_p)x(father_educ)	.004 (.003)	.071 (.033)**	.014 (.007)*	.129 (.059)**		
(married_p)x(mother_educ)	.002 (.004)	-.041 (.040)		-.021 (.069)		
parents' public job	.021 (.011)*	.088 (.128)	.047 (.032)	.140 (.122)		.168 (.132)
home_aid	.009 (.032)	.076 (.515)	.038 (.129)	-.059 (.801)		
mother_inactive	-.001 (.010)	.092 (.104)	.007 (.025)	.130 (.132)		-.071 (.104)
father_inactive	.034 (.021)	.282 (.241)	.077 (.055)	.041 (.313)		.118 (.132)
children private education	.010 (.008)	.186 (.110)*	.049 (.027)*	.193 (.174)		
parents' subjob	.008 (.007)	.001 (.076)	.010 (.018)	.006 (.118)		-.043 (.117)
children scholarship	.051 (.051)		.009 (.096)	1.12 (.611)*		-.003 (.163)
remittances	.226 (.163)		.409 (.205)**	1.58 (2.32)		.277 (.607)
absolute wealth	-.214 (.072)***	-1.04 (.686)	-.338 (.156)**	-2.55 (1.15)**		-.257 (.678)
relative wealth	.158 (.061)***	.803 (.643)	.240 (.147)	1.62 (1.06)		-1.01 (1.06)
constant	.427(.263)	2.35 (2.78)	.513 (.620)		.528 (.214)*	.656 (3.30)
R-squared	.219					
Pseudo R-squared		.148		.096		
Sigma			.262			

Table 8

Estimates of educational gap (all heteroskedasticity-robust); only boys among [8,14].

Number of observations: 2012 - (estimated standard errors in parentheses) - *** significant at 1%; ** at 5%; * at 10%

Dependent Variable: educational gap

Notes: in the case of CLAD and SCLS, some explanatory regressors were not included in order to avoid lack of computational convergence.

	OLS	PROBIT	TOBIT	OLOGIT	CLAD	SCLS
married parents	-.045 (.033)	-.187 (.303)	-.072 (.062)	-.547 (.452)	-.002 (.033)	-.007 (.042)
quantity children age<15	.014 (.002)***	.081 (.026)***	.023 (.005)***	.103 (.038)***	.029 (.008)*	.021 (.012)*
quantity people age>59	.001 (.011)	.069 (.104)	.011 (.022)	-.099 (.163)	.015 (.030)	-.006 (.083)
quantity people with income	.013 (.004)***	.164 (.043)***	.029 (.008)***	.270 (.061)***	.028 (.008)*	.021 (.021)
log income per capita	-.036 (.059)	.159 (.633)	.062 (.136)	.015 (.922)	-.054 (.018)*	.170 (.805)
(log income per capita)^2	.001 (.003)	-.031 (.039)		-.026 (.058)		-.013 (.056)
father's education	-.010 (.004)**	-.074 (.041)*	-.019 (.008)**	-.136 (.066)**	-.008 (.001)	-.024 (.018)
(father's education)^2	.000 (.000)	.001 (.002)		.005 (.003)		.001 (.002)
mother's education	-.024 (.005)***	-.220 (.048)***	-.048 (.010)***	-.358 (.080)***	-.021 (.004)*	-.039 (.040)
(mother's education)^2	.001 (.000)***	.007 (.002)***	.001 (.000)***	.012 (.003)***		.001 (.004)
(father_educ)x(mother_educ)	.000 (.000)	-.001 (.003)	.000 (.000)	-.003 (.004)		
(married_p)x(father_educ)	.002 (.003)	.035 (.031)	.007 (.006)	.037 (.052)		
(married_p)x(mother_educ)	.003 (.004)	.001 (.038)	.003 (.008)	.037 (.068)		
parents' public job	.008 (.0129)	-.097 (.142)	-.001 (.034)	.157 (.119)		.037 (.076)
home_aid	-.001 (.039)	.023 (.413)	.017 (.112)	-.249 (.904)		
mother_inactive	.002 (.011)	-.035 (.105)	.002 (.022)	.065 (.116)		.007 (.036)
father_inactive	.031 (.028)	.309 (.241)	.064 (.054)	.277 (.315)		.039 (.121)
children private education	.007 (.009)	.150 (.114)	.039 (.027)	.155 (.182)		
parents' subjob	.005 (.007)	.025 (.078)	.007 (.017)	-.089 (.110)		.018 (.053)
children scholarship	-.124 (.023)***	-1.21 (.701)*	-.311 (.127)**	-1.28 (.895)		-.045 (.088)
remittances	-.028 (.030)		-1.04 (.101)***			
absolute wealth	-.194 (.079)**	-1.39 (.713)**	-.330 (.148)**	-2.23 (1.05)**		-.187 (.377)
relative wealth	.138 (.066)**	1.28 (.657)*	.242 (.137)*	1.44 (.972)	.036 (.131)	.065 (.439)
constant	.547 (.237)**	1.94 (2.45)	.376 (.523)		.554 (.135)*	-.168 (2.92)
R-squared	.221					
Pseudo R-squared		.155		.095		
Sigma			.255			

Robustness Check

Also we introduced and tested two suggestions of Berlinski et al. (2007). Firstly, these authors study the determinants of the levels of completed education among individuals aged 7-15 in Uruguay. Children can enroll in the ..rst grade of primary education if they become 6 before the 10th. of May. Since the ECH Survey gives no information on birth date, they restrict the sample to the months of January to April. Secondly, Berlinski et al. study the e¤ect of pre-primary education on children's subsequent school outcomes and they suggest a positive relationship. Thus, in this paper, we also introduced the binary regressor pre-primary education and restricted the sample to the months of January to April. But the new regressor has no signi..cative impact on the educational gap and the results are similar to the tables 7 and 8 (see tables 12 and 13 in the Annex)

Testing Endogeneity

The term "endogenous" in econometrics is used to describe any situation where an explanatory variable is correlated with the disturbance. One way in which endogeneity could arise is from the "omitted variables problem" and it might have appeared in the applied part of this paper because of the possible linkage between the variable "parents' marital status" and the unobserved "parents' irresponsibility". With the intention of eliminate, or at least mitigate, the possible omitted variable bias, we introduced proxy variables.

Proxy binary variables for unobserved "parents' irresponsibility" (takes value one in case of parents' irresponsibility):

a) The survey asks the parents who have a job and didn't work last week for the reasons of this attitude. If they answer: "because of bad weather or not too much work to do", then "parents' irresponsibility" takes value one.

b) The survey asks the parents who have a job if they would like to work more hours. If they answer: "yes, but I did nothing to work more hours" or "yes, but I am not searching for other job", then "parents' irresponsibility" takes value one.

c) The survey asks the unemployed parents if they did anything to find a job last week. If they answer: "nothing", then "parents' irresponsibility" takes value one.

In this paper, these different proxy variables were aggregated in one dummy variable which takes value one if any of the dummies above is different from zero. We tested its significance using Tobit, CLAD and SCLS models, for boys and girls separately, with "educational gap" as the dependent variable. In no one of these models, the coefficients of this proxy variable of "parents' irresponsibility" were significantly different from zero (see next table 9). Thus, we did this exploratory exercise but we were not able to find a good proxy of parents' irresponsibility. The variables employed as proxy could be disputed but they were selected due to the restriction of variables available in the ECH survey.

Table 9

Searching for a proxy of parents' irresponsibility

Tobit(MLE) (heteroskedasticity-robust), CLAD and SCLS estimates of educational gap.

Dependent Variable: educational gap - (estimated standard errors in parentheses) - *** significant at 1%; ** at 5%; * at 10%

Notes: in the case of CLAD and SCLS, some explanatory regressors were not included in order to avoid lack of computational convergence.

	girls among [8,14] - 2055 observations			boys among [8,14] - 2012 observations		
	TOBIT	CLAD	SCLS	TOBIT	CLAD	SCLS
parents' irresponsibility	.001 (.025)	-.024 (.026)	-.011 (.104)	.022 (.023)	.001 (.026)	.007 (.123)
married parents	-.100 (.068)	-.080 (.050)	-.121 (.108)	-.070 (.062)	-.002 (.034)	-.006 (.041)
quantity children age<15	.022 (.005)***	.036 (.019)*	.037 (.018)*	.023 (.005)***	.029 (.007)*	.021 (.015)
quantity people age>59	-.025 (.025)		.005 (.188)	.011 (.022)	.014 (.030)	-.008 (.064)
quantity people with income	.023 (.010)**	.035 (.018)*	.018 (.028)	.028 (.008)***	.028 (.009)*	.021 (.024)
log income per capita	-.023 (.162)	-.057 (.029)*	-.035 (1.08)	.062 (.136)	-.053 (.020)	.186 (1.23)
(log income per capita)^2			-.001 (.079)			-.014 (.086)
father's education	-.022 (.009)**	-.001 (.004)	.042 (.079)*	-.019 (.008)**	-.008 (.004)*	-.024 (.023)*
(father's education)^2			-.003 (.006)			.001 (.001)
mother's education	-.032 (.010)***	-.022 (.010)	-.072 (.053)	-.047 (.010)***	-.021 (.005)*	-.038 (.033)
(mother's education)^2			.002 (.006)	.001 (.000)***		.001 (.002)
(father_educ)x(mother_educ)	.001 (.001)			.000 (.000)		
(married_p)x(father_educ)	.014 (.007)*			.007 (.006)		
(married_p)x(mother_educ)				.002 (.008)		
parents' public job	.047 (.032)		.166 (.143)	-.002 (.034)		.038 (.103)
home_aid	.038 (.129)			.018 (.112)		
mother_inactive	.006 (.033)		-.063 (.121)	-.015 (.029)		
father_inactive	.076 (.055)		.119 (.146)	.058 (.054)		.038 (.116)
children private education	.049 (.027)*			.039 (.027)		
parents' subjob	.010 (.018)		-.040 (.140)	.005 (.018)		.019 (.066)
children scholarship	.010 (.097)		-.012 (.247)	-.307 (.127)**		-.044 (.134)
remittances	.410 (.205)**		.269 (.662)	-1.04 (.101)***		
absolute wealth	-.338 (.156)**		-.259 (.742)	-.339 (.149)**		-.178 (.482)
relative wealth	.240 (.147)		-1.00 (1.17)	.250 (.138)*	.037 (.121)	.046 (.522)
constant	.513 (.620)	.529 (.260)	.663 (3.76)	.376 (.523)	.551 (.151)*	-.231 (4.44)
Sigma	.262			.255		

Treatment Evaluation

Table 10 - Average E¤ect (on Educational Gap) of Treatment (Married Parents) on the Treated (ATT)

	Girls	Boys
ATT	-.0129	-.0022
n. treat	1588	1733
n. contr.	258	235
Treated	.0757	.0951
Controls	.0886	.0973
S.E.	.0163	.0165
T-stat	-0.79	-0.13

Note: estimation with the Kernel Matching method

The point estimates indicate that having married parents reduce the educational gap and the e¤ect of the "treatment" (having married parents) is greater in the case of daughters. However the ATT is not signi…cantly di¤erent from zero neither in the daughters' case nor in the boys' one. Thus, using the propensity score and the Kernel matching method, there's no strong evidence to support the positive in‡uence of having married parents on their children attainment. In order to evaluate this result properly, we should bear in mind the results of the next table 11: though the Kernel matching method made comparisons between treated and control individuals who are as similar as possible, this similarity is far from perfect. As it is shown in this table, the mean of some characteristics of the individuals continue to be di¤erent after the matching. This fact denotes that there are no observable characteristics which are not included in the matching, producing that the matching is not so satisfactory.

Table 11
Descriptive Statistics for the Treated, not Treated and Matched groups

Household environment of girls with age [8,14]

Variable	Sample	Mean Treated	Mean Control	t-test t	t-test p>t
log income per capita	Unmatched	8.0068	7.2913	13.27	0.000
	Matched	7.8457	7.7696	3.03	0.002
any parent unemployed	Unmatched	.14524	.23643	-3.78	0.000
	Matched	.16058	.19161	-2.30	0.022
mother's education	Unmatched	10.378	6.9806	12.69	0.000
	Matched	9.5318	9.0267	4.13	0.000
father's education	Unmatched	10.023	6.8605	11.51	0.000
	Matched	9.1814	8.7265	3.49	0.000
relative wealth	Unmatched	.37738	.16771	13.60	0.000
	Matched	.32916	.2957	4.86	0.000
household ownership	Unmatched	.68893	.45736	7.44	0.000
	Matched	.65806	.65434	0.22	0.825
more than one family at home	Unmatched	.0089	.04651	-4.90	0.000
	Matched	.00945	.00841	0.31	0.757
number of individuals per room	Unmatched	2.1911	3.0303	-11.30	0.000
	Matched	2.2707	2.2271	1.17	0.243

Household environment of boys with age [8,14]

Variable	Sample	Mean Treated	Mean Control	t-test t	t-test p>t
log income per capita	Unmatched	7.9726	7.3224	11.29	0.000
	Matched	7.9292	7.7866	5.48	0.000
any parent unemployed	Unmatched	.13506	.22553	-3.71	0.000
	Matched	.13849	.14793	-0.79	0.427
mother's education	Unmatched	10.224	7.4447	10.15	0.000
	Matched	10.022	9.4638	4.34	0.000
father's education	Unmatched	9.9226	7.0128	10.29	0.000
	Matched	9.706	9.4067	2.11	0.035
household ownership	Unmatched	.67136	.47234	6.06	0.000
	Matched	.66532	.63145	2.09	0.037
more than one family at home	Unmatched	.00788	.0383	-4.14	0.000
	Matched	.00808	.00665	0.49	0.622
sum of parents age	Unmatched	83.18	78.783	4.74	0.000
	Matched	82.885	81.825	2.23	0.026

6. Conclusions

There's a growing body of research on the determinants of children's school performance and not few methodological problems appear in previous investigations about the determinants of educational gap. This paper has extended prior research considering -besides the possible existence of endogeneity- censored regression models -such as Tobit Model- and semiparametric alternative approaches -such as the Censored Least Absolute Deviations Estimator and the Symmetrically Censored Least Squares Estimator. Drawbacks and advantages of the different estimation methods have been discussed. In the empirical application, this study introduces indicators of family environment and focuses on the impact of the parents' marital status on their children educational attainment. In particular, the results suggest positive influence of having married parents on daughter's educational outcomes, after controlling for household background variables such as parents' education, income per capita, wealth and number of children. This finding is consistent with previous investigations (see Brown 2004 for an extensive summary) and with the theoretical hypotheses that both the impermanence of cohabiting unions and their incomplete institutionalization (unclear family roles, rights, and obligations) set the stage for a family environment that may undermine child development. Finally, this paper includes an application of the propensity score approach for treatment evaluation of parents' marital status: all the point estimates indicate that having married parents has a positive effect on children schooling outcomes, but the results are no robust since confidence intervals does not span zero. This present study contributes to the economic literature in this field by applying more suitable estimation methods and by checking through the possible faults or omissions of methods used in previous investigations.

For further research, four considerations about the empirical application: First, a significant shortcoming of the survey used in this paper, is that it does not have longitudinal data or cohort information[11]: there's no information available about the marriage history information of the biological parents[12]. Thus, one drawback of Continuous Households Survey (ECH) is that it does not provide measurement of the duration of the di¤erent family structures or the number of di¤erent family transitions that children have experienced (so long term or cumulative e¤ects of family structure can't be observed). Second, besides taking into account data from all the available years of the Continuous Households Survey, in order to contribute to unravel the complexities of family issues, it could be useful to wide the range of family structures and also test the di¤erent incidence of, for instance, the two-biological-parent families, stepfamilies and female-headed households over the children education attainments. Moreover, it could be interesting to evaluate also for the other years of the ECH survey if cohabiting e¤ects on educational attainment could be biased against daughters -a kind of unwelcome discrimination- as it is suggested in this paper. Third, this investigation could be completed testing also not only children school engagement but also other behavioral and emotional e¤ects. Fourth, one major problem with the data used for the empirical application is that there is no measure of the children ability which should be positive correlated with school performance.

On the theoretical ..eld of estimation methods for censored regression models, other semiparametric alternatives for censored models could be evaluated.

[11]Brown (2004) quotes previous research which using longitudinal data also suggests a positive relationship between two-biological married parents and child well-being.

[12]Longitudinal data with individual life trajectories would allow us to observe, for instance, how parents' attitudes about cohabitation in‡uence the child's subsequent marital and cohabitation experience (Axinn and Thornton, 1993)

References

Andersen, Lykke. 2001. "Social Mobility in Latin America: Links with Adolescent Schooling." Research Network Working Paper R-433, Inter-American Development Bank, July 2001.

Axinn, William G. and Thornton, Arland. 1993. "Mothers, Children and Cohabitation: the Intergenerational Effects of Attitudes and Behaviour." American Sociological Review. April 58:2.

Becker, Gary S. 1988. "Family Economics and Macro Behavior." American Economic Review. March 1988, 78:1, pp. 1-13.

Becker, Sascha and Ichino, Andrea. 2002. "Estimation of average treatment effects based on propensity scores", The Stata Journal (2002) Vol.2, No.4, pp. 358-377.

Behrman, Jere R.; Birdsall, Nancy and Székely, Miguel. 2000. "Intergenerational Mobility in Latin America: Deeper Markets and Better Schools Make a Difference." In New Markets, New Opportunities? Economic and Social Mobility in a Changing World. Carnegie Endowment for International Peace; Washington, D.C.: Brookings Institution Press, 2000.

Berlinski, Samuel; Galiani, Sebastian; Manacorda, Marco. 2007. "Giving Children a Better Start: Preschool Attendance and School-Age Profiles." World Bank Policy Research Working Paper 4240, June 2007.

Bjorklund, Anders; Ginther, Donna K. and Sundstrom, Marianne. 2005. "Does Marriage Matter for Kids? The impact of Legal Marriage on Child Outcomes"

Brown, Susan L. 2004. "Family Structure and Child Well-Being: The Significance of Parental Cohabitation." Journal of Family and Marriage. May, 66:2.

Cameron, A. Colin and Pravin K. Trivedi. 2005. "Microeconometrics. Methods and Applications". Cambridge University Press.

Case, Anne and Angus Deaton. 1999. "School Inputs and Educational Outcomes in South Africa". The Quarterly Journal of Economics, Vol. 114, No. 3. (Aug., 1999), pp. 1047-1084.

Chay, Kenneth Y. and James L. Powell. 2001. "Semiparametric Censored Regression Models." Journal of Economic Perspectives. Fall 2001, 15:4, pp. 29-42.

Cid, Alejandro; Ignacio Presno and Luis Viana. 2004. "Institutions, Family and Economic Performance." Revista de Ciencias Empresariales y Economía. Año III. 2004.

Datcher-Loury, Linda. 1988. "E¤ects of Mother's Home Time on Children's Schooling." The Review of Economics and Statistics. August 1988, 70:3, pp. 367-73.

Elwood, David T. 2000. "Anti-Poverty Policy for Families in the Next Century: From Welfare to Work - - and Worries". Journal of Economic Perspectives. Winter 2000, 14:1, pp. 187-198.

Filgueira, Carlos; Fernando Filgueira and Alvaro Fuentes. 2003. "School Attainment and Transitions to Adulthood in Latin America." In Cox Edwards, Alejandra; Duryea, Suzanne; Ureta, Manuelita. Editors. 2003. "Critical Decisions at a Critical Age: Adolescents and Young Adults in Latin America." Inter-American Development Bank.

Francesconi, Marco; Jenkis, Stephen P.; Siedler, Thomas. 2006. "Childhood Family Structure and Schooling Outcomes: Evidence for Germany". Discussion paper 610. German Institute for Economic Research.

Fishback, P.; Baskin, J. 1991. "Narrowing the Black-White Gap in Child Literacy in 1910: The Roles of School Inputs and Family Inputs." The Review of Economics and Statistics, Vol. 73, No. 4. (Nov., 1991), pp.725-728.

Heckman, James J. and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." American Economic Review. May 2001, 91:2, pp. 145-49.

Honoré, Bo E. and James L. Powell. 1994. "Pairwise Di¤erence Estimators for Censored and Truncated Regression Models." Journal of Econometrics. September/October, 64:1-2, pp. 241-78.

Margo, R. 1987. "Accounting for Racial Di¤erences in School Attendance in the American South, 1900: The Role of Separate-but-Equal". The Review of Economics and Statistics, Vol. 69, No. 4. (Nov., 1987), pp.661-666.

McKenzie, David and Hillel Rapoport. 2005. "Migration and Education Inequality in Rural Mexico." Stanford Center for International Development. Working Paper Nº 258. September 2005.

Manning, Wendy D. and Daniel T. Lichter. 1996. "Parental Cohabitation and Children's Economic Well-Being." Journal of Family and Marriage. November, 58:4.

McLanahan, Sara. 1985. "Family Structure and the Reproduction of Poverty". American Journal of Sociology, January, 90:4, pp. 873-901.

Powell, James L. 1984. "Least Absolute Deviations Estimation for the Censored Regression Model." Journal of Econometrics. July, 25:3, pp. 303-25.

Powell, James L. 1986. "Symmetrically Trimmed Least Squares Estimation for Tobit Models." Econometrica. November, 54:6, pp. 1435-60.

Psacharopoulos, George and Ana Maria Arriagada. 1989. "The Determinants of Early Age Human Capital Formation: Evidence from Brazil." Economic Development and Cultural Change, Vol. 37 (4), July 1989.

Raley, R. Kelly and Elizabeth Wildsmith. 2004. "Cohabitation and Children's Family Instability." Journal of Family and Marriage. February, 66:1.

Saha, Rumki. 2005. "The Determinants of the Changing Educational Gender Gap in Brazil." December 2005. Cornell University. Department of Economics.

Sax, Leonard. 2005. "Why Gender Matters. What Parents and Teachers Need to Know about the Emerging Science of Sex Di¤erences". Doubleday.

Simon, Julian L. 1998. "The Ultimate Resource 2". Princeton University Press.

Wooldrige, Je¤rey M. 2002. "Econometric Analysis of Cross Section and Panel Data". Massachusetts Institute of Technology.

Table 12

Estimates of educational gap (all heteroskedasticity-robust).

Sample restricted to the months of January to April; only girls among [8,14]

Number of observations: 687 - (estimated standard errors in parentheses) - *** significant at 1%; ** at 5%; * at 10%

Dependent Variable: educational gap

	OLS	PROBIT	TOBIT	OLOGIT
married parents	**-.153 (.068)****	-.788 (.733)	**-.445 (.206)****	-1.131 (1.139)
preschool attendance	-.021(.020)	.005 (.197)	-.035 (.060)	.025 (.284)
quantity children age<15	.025 (.006)***	.161 (.062)***	.062 (.018)***	.208 (.080)**
quantity people age>59		-.056 (.251)	.016 (.076)	-.109 (.374)
quantity people with income	.015 (.012)	.154 (.099)	.050 (.033)	.242 (.146)*
log income per capita	-.020 (.090)	.018 (.888)	-.008 (.286)	-.590 (1.177)
(log income per capita)^2	.001 (.005)	-.004 (.060)		.056 (.080)
father's education	.013 (.009)	.030 (.095)	.017 (.029)	.134 (.145)
(father's education)^2		-.003 (.006)		-.010 (.008)
mother's education	-.038 (.011)	-.281 (.091)***	-.088 (.028)***	-.480 (.169)***
(mother's education)^2		.002 (.005)	.001 (.001)	.008 (.006)
(father_educ)x(mother_educ)			-.002 (.002)	-.001 (.011)
(married_p)x(father_educ)	-.005 (.007)	.020 (.077)	.007 (.023)	-.008 (.106)
(married_p)x(mother_educ)	.020 (.010)**	.105 (.079)	.052 (.024)**	.152 (.138)
parents' public job	.034 (.024)	.332 (.348)	.135 (.114)	.288 (.286)
home_aid	.048 (.061)	1.712 (.782)**	.536 (.267)**	1.211 (1.207)
mother_inactive	-.019 (.025)	.520 (.235)**	.077 (.080)	.065 (.264)
father_inactive	-.027 (.038)	-.032 (.464)	-.074 (.145)	-.151 (.514)
children private education		.236 (.282)	.078 (.099)	.218 (.438)
parents' subjob	.006 (.015)	-.027 (.179)		.286 (.252)
remittances	.256 (.201)		.610 (.241)**	1.843 (1.879)
absolute wealth	-.358 (.156)**	-4.198 (1.470)***	-1.206 (.471)**	-5.182 (2.060)**
relative wealth	.265 (.129)**	2.839 (1.439)**	.744 (.458)	3.010 (2.130)
constant	.411 (.371)	1.300 (3.410)	.451 (1.096)	
R-squared	.340			
Pseudo R-squared		.308		.156
Sigma			.355	

Table 13

Estimates of educational gap (all heteroskedasticity-robust)

Sample restricted to the months of January to April; only boys among [8,14]

Number of observations: 642 - (estimated standard errors in parentheses) - *** significant at 1%; ** at 5%; * at 10%

Dependent Variable: educational gap

	OLS	PROBIT	TOBIT	OLOGIT
married parents	-.086 (.055)	-.815 (.606)	**-.308 (.136)****	**-2.223 (.975)****
preschool attendance	-.007 (.019)	.176 (.179)	.011 (.046)	.060 (.227)
quantity children age<15	.007 (.005)	.080 (.060)	.023 (.014)*	.002 (.072)
quantity people age>59	-.029 (.015)*	-.020 (.193)	-.017 (.043)	-.030 (.279)
quantity people with income	.027 (.008)***	.327 (.084)***	.064 (.016)***	.384 (.104)***
log income per capita	-.145 (.117)	.401 (1.554)	.027 (.049)	-.069 (1.950)
(log income per capita)^2	.006 (.006)	-.054 (.098)		-.032 (.123)
father's education	.004 (.009)	.023 (.091)	.015 (.019)	-.157 (.128)
(father's education)^2		.001 (.005)		.011 (.006)*
mother's education	-.044 (.011)***	-.462 (.103)***	-.113 (.021)***	-.643 (.182)***
(mother's education)^2	.001 (.000)***	.015 (.004)***		.018 (.004)***
(father_educ)x(mother_educ)		-.010 (.006)	-.002 (.001)*	-.017 (.007)**
(married_p)x(father_educ)	-.008 (.006)		-.005 (.016)	.034 (.119)
(married_p)x(mother_educ)	.021 (.008)**	.143 (.086)*	.057 (.018)***	.343 (.170)**
parents' public job	.009 (.020)	-.137 (.375)	.026 (.093)	-.152 (.246)
home_aid	-.033 (.032)		-1.265 (.194)***	-32.61 (.574)***
mother_inactive	-.003 (.025)	-.115 (.219)	-.025 (.049)	.088 (.228)
father_inactive	-.005 (.061)	.025 (.480)	.031 (.135)	-.250 (.663)
children private education	.012 (.019)	.353 (.267)	.124 (.067)*	.211 (.354)
parents' subjob	.013 (.016)	-.075 (.177)	-.009 (.041)	-.094 (.216)
absolute wealth	-.130 (.161)	-1.206 (1.762)	-.209 (.393)	-1.529 (2.068)
relative wealth	.129 (.137)	1.233 (1.568)	.191 (.364)	1.412 (1.888)
constant	1.044 (.472)**	1.923 (5.992)		
R-squared	.362			
Pseudo R-squared		.309		.129
Sigma			.279	