WORKING PAPERS

The effect of the number of siblings on education in sub-Saharan Africa: evidence from a natural experiment

Mathias KUEPIE[1, 2]
Michel TENIKUE[1, 3]

CEPS/INSTEAD, Luxembourg[1]
DIAL, France[2]
University of Namur, Belgium[3]

Working Paper No 2012-28
June 2012

The effect of the number of siblings on education in sub-Saharan Africa: evidence from a natural experiment[1]

Mathias Kuepie

CEPS/INSTEAD, Luxembourg
DIAL, France

Michel Tenikue

CEPS/INSTEAD, Luxembourg
University of Namur, Belgium

Abstract

The objective of this paper is to investigate the effect of the number of siblings on education in urban sub-Saharan Africa. The birth of twins is considered as a natural experiment that affects the number of siblings but has no direct effect on education. This event is used as instrumental variable in a two-stage least-squared estimation approach to investigate the causal effect of the number of siblings on school achievement. Equations are estimated on subsamples of singleton children born before the twins. The results show that an exogenous fertility increase significantly inhibits human capital accumulation. However, the magnitude of the marginal effect seems small: one additional sibling decreases the total number of school grade by nearly one-tenth. In a context of high fertility, the total effect might become very detrimental.

Keywords: twins, fertility, education, sub-Saharan Africa.

[1] Corresponding author: michel.tenikue@ceps.lu

Introduction

The lack of human capital has been identified as one of the major sources of economic stagnation. It limits productivity gains and constrains social progress (Schultz 1961, Becker 1993). Many developing countries, in particular sub-Saharan Africa countries, are characterized by lack of human capital. Efforts from national and international organization undertaken through the Millennium Development Goals and devoted to improving education have provided limited results. Human capital indicators are still low in sub-Saharan: about one-third of the adult population remains uneducated, and only a small proportion of children have completed their primary education (World Bank, 2011). This underinvestment in education is the outcome of numerous factors, from both the supply and demand sides of education. In this study, we focus on the demand side factors and, more specifically, we are interested in the relation between family fertility behaviour and household child education.

From a theoretical perspective, the link between fertility and socioeconomic outcomes (including, but not only, education) can be studied with the capillarity theory framework. Arsène Dumont (1890) defines capillarity theory as follows: "Just as a colon has to be thin to allow liquid to rise by capillarity, the family size has to be small to allow the family to rise on the social scale". The main mechanism underlining the capillarity theory is the "dilution" of parental inputs (Blake 1981). Parental resources are finite, and as the number of children in the family increases, the resources accrued to any one child necessarily decline. Siblings are competitors for parents' time, energy and financial resources, and so the fewer the better. The quantity-quality model (Becker & Lewis 1973) also provides a framework to investigate the relationship between family size and children's outcomes. The model presumes that households allocate resources to each child to improve their quality. A direct implication of this model is a trade-off between per child investment (quality) and the number of children in the family (quantity).

From an empirical perspective, the literature on the relation between quality and quantity of children is huge and diverse. The papers covers different areas in the world, including the following countries or regions: the US (Blake 1981, Downey 1995, Cáceres-Delpiano 2006, Conley & Glauber 2006), Europe (Black et al 2005, Goux & Maurin 2005), Israel (Angrist et al 2010), Asia (Knodel & Wongsith 1991), Kenya (Gomes 1984), Botswana (Chernichovsky 1985), Ghana (Montgomery, Kouame & Oliver 1995), Ivory Coast (Montgomery et al. 1995), Malaysia (Sudha 1997), China (Lu & Treiman 2005 , Rosenzweig & Zhang 2009), Hungary (Van Eijck & De Graaf 1995) and Cameroon (Eloundou-Enyegue & Williams 2006). The theory predicts a negative relation between quantity and quality of children while the empirical literature provides mixed results. In developed countries, previous studies in the 70s and 80s display a consistent negative relationship between the number of siblings and the schooling (Sewell 1968, Becker & Lewis 1973, Blake 1981, Becker & Tomes 1986). Recent studies provide less clear-cut conclusions. Black et al (2005), based on Norwegian data, show that when there is a negative correlation between children's school achievement and fertility, this correlation vanishes when endogenous fertility is controlled for. Cáceres-Delpiano (2006), analysing US census data, comes to the conclusion that an exogenous increase in fertility doesn't affect child quality per se (measured by school grade repetition) but obliges households to reallocate resources (for example by enrolling children in public schools instead of private ones) to avoid reduction in child quality. Goux and Maurin (2005) also show that it is not an increase in family size per se that causes lower school performance in France, but rather the fact of living in an overcrowded family (in terms of the number of people per room). In the same vein, Angrist et al (2010) find no evidence for the quantity-quality trade-off in Israel. A recent paper by Conley and Glauber (2006) demonstrates a clear negative relationship between fertility and children's school performance.[2]

In developing countries, the literature also shows mixed conclusions. In some contexts a negative relationship is found (Ivory Coast, Ghana), while in others, a positive relationship is observed (Kenya, Botswana). These results raise the possibility of systematic variations in the relation across societies as noted by Eloundou-Enyegue and Williams (2006).[3] Most papers do not take into account the fact that family size and children's schooling are jointly determined in a household decision model (Becker & Lewis 1973, Baland & Robinson 2000). The correlations highlighted may then be spurious due to endogeneity. In a seminal empirical paper,

[2] Especially for boys.
[3] In some societies, parents can externalize a part of the education costs and rely on "extended family buffering" (Eloundou-Enyegue & Williams 2006). This can take the form of fosterage (Vandermeersch 2000, Rakoto-Tiana 2011) or transfers (Azam & Gubert 2005, Lu et al.2007, Temesgen 2007, Kuepie 2011).

Rosenzweig and Wolpin (1980) addressed the endogeneity problem. They used twin birth as an instrument for fertility in an instrumental variable estimation framework. They found that an increase in fertility reduced children’s school achievement in rural India. In a more recent paper, using a similar approach, Rosenzweig and Zhang (2009) highlighted a negative effect of fertility on human capital formation in China. However, the magnitude of the effect is moderate.

This paper contributes to this debate in providing new evidence from sub-Saharan Africa. It uses data from 29 recent Demographic and Health Surveys (DHS) in 20 countries. DHS data provide information on women’s birth history and fertility preferences. They also contain socioeconomic information including school attainment and a wealth index of households surveyed. We first discuss the issue of endogeneity between fertility and children’s school achievement in sub-Saharan African urban areas. Our identification strategy is based on an instrumental variable approach. The advent of twin births is used as an instrument for fertility (Rosenzweig & Wolpin 1980, Cáceres-Delpiano 2006, Rosenzweig & Zhang 2009). We then estimate the relation between family size and children’s education with a two-stage least-squared (2SLS) approach. The results show that high fertility is detrimental to children’s education.

The paper is organized as follows. Section 2 describes the data and discusses the estimation strategy. Section 3 discusses the results and Section 4 provides concluding remarks.

1) Data and methods

We use data from Demographic and Health Surveys (DHS). The DHS programme was originally developed by the US Agency for International Development (USAID). Since 1984, DHS have collected, analysed and disseminated accurate and representative data for more than 200 surveys in more than 75 countries. DHS data are collected with the support of ICF Macro, based in the United States. The samples are representative at national and sub-national levels.[4] DHS survey methodologies and questionnaires are standardized so that data are comparable across countries. The surveys offer detailed information on various subjects including demographic events

[4] http://www.measuredhs.com/

and reproductive health. We use 29 DHS data sets for our analysis. Table 1 below provides the list of countries and years of surveys. We use a large number of surveys because our identification strategy is based on twins' births. The birth of twins is a relatively rare event. Indeed, on average, only 1.5% of all births are twins. We pooled all the surveys to have enough twins in our data set. In all regressions, a country fixed effect is included to account for country-specify characteristics.

Table 1: List of countries and years of the DHS surveys

Country	Year (s) of the survey (s)
Benin	2001, 2006
Burkina Faso	2003
Cameroon	2004
Congo	2005
Democratic Republic of Congo	2007
Ethiopia	2000, 2005
Ghana	2003, 2008
Kenya	2003, 2008
Guinea	2005
Lesotho	2004
Liberia	2007
Madagascar	2003-04, 2008-09
Malawi	2000, 2004
Mali	2001, 2006
Mozambique	2003
Namibia	2000, 2006-07
Niger	2006
Nigeria	2003, 2008
Senegal	2005
Tanzania	2004-05

The main variables used are total fertility, twin births and children's schooling. To collect data on fertility, each woman is first asked to report her birth history, then for each child born alive, information is collected on birth order, birth date, twinning and gender. Computing total fertility and identifying twin births is then straightforward. DHS also collect data on education and literacy for all household members. In this study, the indicator of human capital accumulation is the number of completed school grades.

Regression models control for child gender and age, for mother's age and education, and for a household wealth index defined as the number of durable goods owned by the household. The list of durable goods contains 11 items including a radio, TV, stove, bicycle and car.

The sample consists of singleton children aged 7 to 20 living in urban areas. When they are twins, only their elder brothers or sisters are included in the sample used to run the corresponding regression.[5]

Estimation strategy

We denote by Y_i the number of completed school grades for a given child; by X_i a vector of covariates (sex, age of the child, household economic level and the education level of the mother) and n_i the number of siblings or mother's fertility level. For a given child i, the link between education and fertility is defined as follows:

$$Y_i = X_i\beta + \theta\, n_i + u_i \qquad (1)$$

where u_i is the error term.

The fertility level reflects parental choice, and as such, it is endogenous in a decision household model (Baland & Robinson 2000). We therefore need to explicitly treat the number of siblings as endogenous:

$$n_i = Z_i\delta + v_i \qquad (2)$$

where Z is a vector of variables with $X \subset Z$ and (u,v) correlated. Instrumental variable methods are often used to isolate the exogenous effect of the number of siblings. Two sets of instruments have been frequently used in the literature. The first relies on the arrival of multiple births in the family (Rosenzweig & Wolpin 1980, Black et al 2005, Cáceres-Delpiano 2006, Rosenzweig & Zhang 2009). A twin birth generates an exogenous/unexpected shock on the family size and is not the outcome of parental choice. The second set uses the gender composition of children in the family (Angrist & Evans 1998, Goux & Maurinc 2005, Conley & Glauber 2006, Angrist et al, 2010). The argument of authors who have used the gender composition of children is that parents prefer having children of mixed gender. When the first two children are of the same gender, parents tend to have one additional child. The

[5] We provide explanations on the rationale behind this choice below.

gender composition is then presumed to satisfy the exclusion restriction, so that it affects the number of children with no direct link to education. In addition, it should be correlated sufficiently to the family size. The correlation between gender composition and the number of children has been observed in many developed countries and in Asia. It is shown that the gender composition of the first two elder children significantly affects the probability of having an additional child (Angrist & Evans 1998, Goux & Maurinc 2005, Conley & Glauber 2006, Filmer et al 2009). In sub-Saharan Africa, however, Filmer et al (2009) showed that the effect of the gender composition of children on the family size is weak. In our database, we also observe a weak relationship between the gender composition of children and the number of children (Table 1a in appendix).

Using the gender composition of children as an instrument for the number of children is also questioned by Conley (2000). He documented a direct link between the gender composition of siblings and educational achievement. The underlying mechanism is not clear, however the link is there. Thus, in the sub-Saharan African context, using gender composition as an instrument for the number of children would lead to inconsistent estimates (Stock et al 2002).

In this paper, we opt for using the birth of twins as an instrument for the number of children. Having twins is random and twinning has a direct link to the number of children. Having a twin birth is, then, included in vector Z of equation 2. This approach has already been used in the literature (Rosenzweig & wolpin 1980, Angrist & Evans 1998, Cáceres-Delpiano 2006, Rosenzweig & Zhang 2009). In line with these authors, we restrict our sample to singleton children born before the twins. This restriction is motivated by two reasons. The first is that twins present some biological frailties materialized, for instance, by their small weight at birth. The restriction helps to distinguish the pure effect of sibling size increase due to twin birth from the effect of the lower endowment of twins. The second is that children coming after twins may be adversely affected by some post-twin birth effects, such as the initial increase in the family size. This restriction helps to avoid confounding factors related to “the post-treatment effect “(Cáceres-Delpiano 2006).

However, as noted by Rosenzweig & Zhang (2009), when the sample is restricted to elder children, the twin-birth effect is underestimated. This is due to the fact that parents may allocate different resources to children, according to their initial biological endowments.[6] Given that twins are less well off at birth than other children, rational parents may decide to reinforce these initial differences through unequal resource allocation among children. Thus, the 2SLS results presented in this paper can be interpreted as a lower bound of the actual effect of having a twin birth on schooling.

For our analysis, in line with Cáceres-Delpiano (2006), the sample is divided into subsamples based on the number of children. Given that fertility is high in sub-Saharan Africa (actual and desired average number of children is around five), we create four subsamples. The first subsample consists the first two singleton children in families with at least three children. In this case, the instrumental variable is the dummy variable defined by: 1 whether the third birth was a twin and 0 otherwise. The second subsample comprises the first three singleton children in families with at least four children. The instrumental variable is then defined by: 1 whether the third birth was a twin and 0 otherwise. The last two subsamples are defined similarly with families with at least five and six children.

3 Results

Descriptive statistics

Table 2 presents the mean and the standard deviation of the main variable used in the regression analysis. The statistics are computed on the first (n-1) children in families with at least n children. In the table, we can see that about 5% of children with at least three siblings have a junior sibling who is a twin.

[6] We do not have information on the birth weight of children. This information would have allowed the effect of allocating resources based on biological endowments to be mitigated.

Table 2: Descriptive statistics on the first "n-1" oldest singletons in family with at least "n" children

	Family with (« n »)			
	At least three children	At least four children	At least five children	At least six children
Number of children	4.59	5.41	6.28	7.17
	(1.54)	(1.46)	(1.39)	(1.30)
Number of completed grades	4.38	4.40	4.38	4.30
	(3.33)	(3.32)	(3.32)	(3.29)
Gender (dummy male =1)	0.51	0.51	0.52	0.52
	(0.50)	(0.50)	(0.50)	(0.50)
Age	12.63	12.96	13.25	13.48
	(3.85)	(3.81)	(3.77)	(3.74)
Age of the mother	34.16	35.55	36.96	38.39
	(5.36)	(5.36)	(5.24)	(5.06)
Education of the mother (years)	5.69	5.03	4.42	3.86
	(5.52)	(5.33)	(5.38)	(5.42)
Household wealth index	3.86	3.69	3.55	3.43
	(2.41)	(2.36)	(2.34)	(2.30)
Have younger twin sibling (dummy)	0.05	0.05	0.04	0.04
	(0.22)	(0.21)	(0.21)	(0.20)
N observations	33409	36785	34033	28183

(Standard deviation in parentheses)

The dependent variable is the number of completed school grades of singleton children aged 7-20 years. For children who have never attended school, this variable is set to zero. The average number of completed grades is around four for all the subsamples.

Figure 1 plots the average number of completed school grades by age. It also shows the theoretical number of grades a child is supposed to have completed at a given age. These theoretical values are computed under the following assumption: a child starts school at the age of six and advances one grade per year. The figure displays an important discrepancy between the theoretical and observed number of completed grades. It also highlights the deficit in schooling of children of school age in countries covered in the data set. For instance, at the age of 15, children display on average a deficit of three years of successful schooling. This deficit is a threat to fulfilling the Millennium Development Goals (Pôle de Dakar 2007).

Figure 1: Number of completed school grades and theoretical school grades by child age

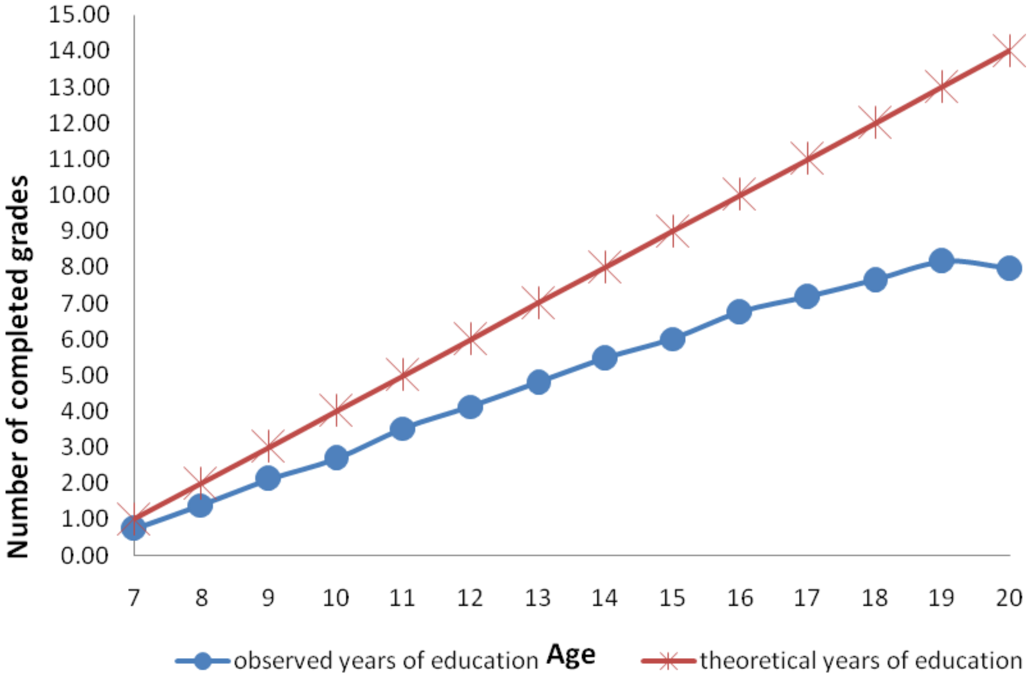
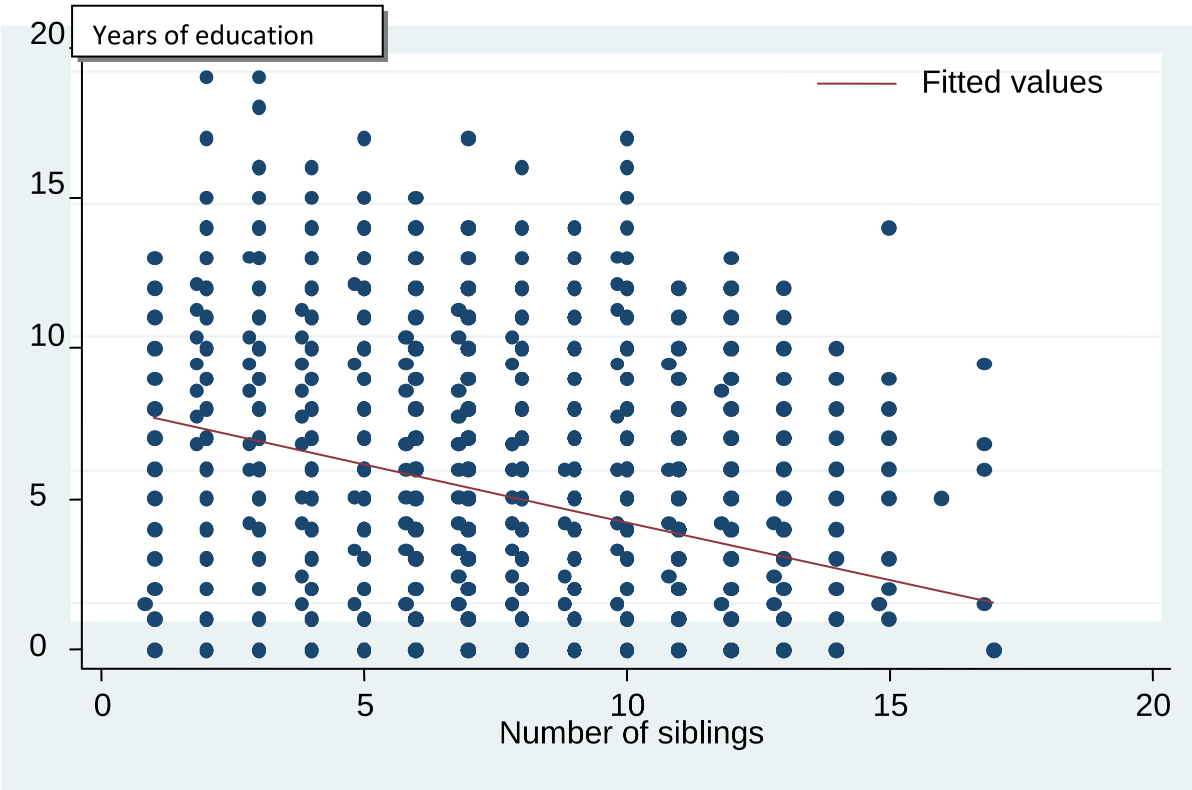


Figure 2 depicts a scatter plot of the number of completed grades for a child and the number of siblings. The regression line on the plot indicates a negative (unconditional) correlation between the number of completed grades and the number of siblings. The higher the number of siblings, the lower the number of completed grades. This correlation does not reflect a causal relationship because the number of siblings is endogenous in a household decision-making model. Our estimation strategy addresses the endogeneity problem.

Figure 2. Scatter plots and linear regression line of the number of successful years of school on the number of siblings



First-stage regression

The results of the first-stage regression are presented in Table 3. They show the link between the number of children and twinning after controlling for the socioeconomic and demographic characteristics of a mother. The results are presented for different subsamples corresponding to the number of siblings. Twinning is measured as a dummy variable that takes the value one for families with twins and zero otherwise. The regression coefficient shows that having twins increases the sibling size by almost 1.5 children for a singleton child who has 3 siblings or more. The magnitude of the coefficient is 1.4 for a singleton child who has 6 siblings or more. It remains between 1 and 2 for all subsamples. This is understandable and is illustrated by the following example. When a child has one sibling and her mother gives birth, her number of siblings increases either by one (in the case of a singleton birth) or by two (in the case of a twin birth). If one moves from the subsample of children who have one sibling or more to a subsample of children who have five siblings or more, the magnitude of the coefficient decreases. This pattern is due to the fact that as the number of children increases, the probability of a mother having a twin birth decreases.

The effect of a twin birth on the number of siblings is statistically significant. The p value of the coefficient is smaller than 0.01. Thus, the advent of a twin birth, which in essence is a random event, is significantly correlated to the number of siblings. It is therefore a reliable instrument.

In the first-stage regression, other control variables are included, as recommended in the two-stage least-square estimation procedure (2SLS). In particular, the mother's number of completed grades and birth order are included. These variables are also determinants of the number of completed grades (Tenikue & Verheyden 2010). Table 3 indicates that, as often noticed in the literature, educated mothers tend to have a smaller number of children.

Table 3. First stage OLS regression of the determinant impact of twin births on the number of siblings of elder singletons

	Family size			
VARIABLES	At least three children	At least four children	At least five children	At least six children
Twin birth	1.494***	1.460***	1.438***	1.388***
	(0.0419)	(0.0460)	(0.0514)	(0.0595)
Child gender – (Male =1)	0.0193	0.0103	0.0173	0.0280**
	(0.0134)	(0.0125)	(0.0126)	(0.0133)
Child age	0.206***	0.197***	0.185***	0.163***
	(0.00312)	(0.00320)	(0.00335)	(0.00379)
Mother's education	-0.0390***	-0.0366***	-0.0294***	-0.0237***
	(0.00395)	(0.00456)	(0.00404)	(0.00363)
Mother's age	-0.0325***	-0.0378***	-0.0400***	-0.0368***
	(0.00241)	(0.00256)	(0.00270)	(0.00307)
Household wealth index	-0.0734***	-0.0644***	-0.0603***	-0.0529***
	(0.00476)	(0.00518)	(0.00522)	(0.00559)
Child is 2nd born	0.590***	0.507***	0.442***	0.344***
	(0.0115)	(0.0124)	(0.0144)	(0.0184)
Child is 3rd born		1.062***	0.905***	0.729***
		(0.0187)	(0.0200)	(0.0238)
Child is 4th born			1.429***	1.136***
			(0.0268)	(0.0306)
Child is 5th born				1.584***
				(0.0385)
Constant	3.208***	3.894***	4.735***	5.591***
	(0.0808)	(0.0837)	(0.0932)	(0.107)
Countries fixed effects included				
Observations	33,409	36,785	34,033	28,183
R-squared	0.381	0.345	0.313	0.271

Robust standard errors are in parentheses. Standard errors are clustered at mother level. *** p<0.01, ** p<0.05, * p<0.1

OLS estimates

OLS estimate of the impact of the number of siblings on school achievement is biased. However, we first present OLS results so that they can be compared with those obtained by 2SLS. OLS estimates exhibit an important negative effect of the number of siblings on school achievement (Table 4). Each additional brother or sister reduces the number of completed grades by about 0.2 years. The magnitude of the effect is stable across subsamples. Also, it is similar to those obtained in the literature (see, for example, Eloundou-Enyegue and Williams [2006] for a review).

The 2SLS estimates

Table 4 presents the estimated coefficient of the 2SLS. The last two rows indicate the statistics of the Hausman test for endogeneity. Our instrument passes the test and indeed our identification strategy holds.

Estimated coefficients by 2SLS are significant for the subsample of families who have at least four children and the subsample of families who have at least five children. The magnitude of the effect is around 0.09. Thus, an unexpected increase in the number of siblings by one reduces the number of completed school grades by 0.09 for a child who lives in a family with at least four or five children. When the sample includes small families (with three children) and large families (at least six children), the estimated 2SLS coefficient remains negative but not significant. This feature suggests that an increase in fertility is more detrimental to the human capital accumulation of children who live in families of average size.[7] The effect is relatively low. The small size of the effect is in part due to within-household reallocation of resources to account for strengthened resource constraints (Cáceres-Delpiano 2006, Angrist et al 2010).

The magnitude of the estimated effect by 2SLS is smaller than the effect obtained by OLS. Estimated 2SLS coefficients are two to three times smaller than OLS coefficients. Such correction of the magnitude of the effect of number of siblings on schooling by 2SLS is persistent in the literature (Black et al 2005, Goux & Maurinc 2005, Cáceres-Delpiano 2006, Rosenzweig & Zhang 2009, Angrist et al

[7] The average and the median number of siblings is between four and five. The average desired number of children is also around five in urban sub-Saharan Africa.

2010). OLS overstates the impact of fertility on children’s outcomes because families who prefer having more children (quantity) are forced to reduce education (quality). Conversely, families who are more interested in quality have to reduce quantity (Cáceres-Delpiano 2006).

It is important to stress that while the marginal effect of the number of siblings on the number of completed grades seems small (about 1/10 of a grade), the total effect might turn out to be important, in particular in larger families. For instance, a child with five siblings is expected to have a deficit of half a grade compared to a singleton child, all else equal. In a context where the average number of completed grades is four, a half grade lost represents 12.5% of the average number of completed grades.

Table 4: Impact of fertility on school achievement of elder singleton children, OLS and 2SLS estimates

	At least three children		At least four children		At least five children		At least six children	
VARIABLES	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Number of siblings	-0.216***	-0.0626	-0.216***	-0.0885*	-0.229***	-0.0989*	-0.220***	-0.0398
	(0.0128)	(0.0438)	(0.0137)	(0.0504)	(0.0152)	(0.0600)	(0.0175)	(0.0759)
Child gender – male	-0.0635***	-0.0668***	-0.0365	-0.0386*	0.00849	0.00543	0.0678**	0.0616**
	(0.0224)	(0.0224)	(0.0222)	(0.0222)	(0.0241)	(0.0242)	(0.0274)	(0.0276)
Child age	0.605***	0.572***	0.576***	0.550***	0.548***	0.523***	0.521***	0.490***
	(0.00547)	(0.0104)	(0.00560)	(0.0112)	(0.00624)	(0.0126)	(0.00694)	(0.0142)
Child is 2nd born	0.0405*	-0.0530	0.0575**	-0.00913	0.0814**	0.0212	0.0836	0.0184
	(0.0228)	(0.0345)	(0.0272)	(0.0373)	(0.0372)	(0.0458)	(0.0566)	(0.0627)
Child is 3rd born			0.00972	-0.130**	0.0296	-0.0929	0.0732	-0.0637
			(0.0360)	(0.0641)	(0.0444)	(0.0703)	(0.0610)	(0.0829)
Child is 4th born					0.0124	-0.181*	0.0293	-0.184
					(0.0541)	(0.102)	(0.0696)	(0.112)
Child is 5th born							0.0549	-0.241*
							(0.0790)	(0.145)
Mother's age	0.0366***	0.0417***	0.0378***	0.0427***	0.0444***	0.0498***	0.0480***	0.0548***
	(0.00388)	(0.00420)	(0.00422)	(0.00465)	(0.00462)	(0.00518)	(0.00519)	(0.00596)
Mother's education	0.0806***	0.0866***	0.0855***	0.0903***	0.0786***	0.0825***	0.0667***	0.0711***
	(0.00817)	(0.00882)	(0.00959)	(0.0103)	(0.00991)	(0.0105)	(0.0105)	(0.0111)
Household wealth index	0.270***	0.282***	0.287***	0.295***	0.309***	0.317***	0.330***	0.340***
	(0.00821)	(0.00917)	(0.00881)	(0.00967)	(0.00922)	(0.0101)	(0.00998)	(0.0111)
Constant	-5.249***	-5.750***	-5.095***	-5.594***	-5.060***	-5.679***	-5.018***	-6.029***
	(0.141)	(0.196)	(0.146)	(0.237)	(0.175)	(0.322)	(0.212)	(0.464)
Countries fixed effects included								
Observations	33,409	33,409	36,785	36,785	34,033	34,033	28,183	28,183
R-squared	0.633	0.630	0.601	0.599	0.564	0.562	0.526	0.522
H0 tot_enf exo		13.63		7.032		5.172		6.080
P-value_H0		0.000223		0.00801		0.0230		0.0137

Robust standard errors are in parentheses. Standard errors are clustered at mother level. *** p<0.01, ** p<0.05, * p<0.1

Robustness checks

The first robustness check is the choice of the outcome variable. Instead of using the number of completed grades, we turn to the variable current school enrolment. This variable is contemporaneous and defined as a dummy. It takes the value one whenever a child is currently enrolled in school and zero otherwise. The model is still estimated by 2SLS. In fact, the second stage is a linear probability model with robust (to heteroscedasticity') standard errors. Table 5 presents the impacts of fertility on school enrolment. The results are very similar to those obtained with the number of completed grades as dependent variable. In all subsamples, the OLS estimated effect is negative and statistically significant, while the 2SLS estimated effect is negative and statistically significant only for subsamples of children living in families with four children or more and families with five children or more. The magnitude of the statistically significant 2SLS effect on school enrolment is about 2%.

Table 5: Impact of fertility on school enrolment of elder singleton children; OLS and 2SLS estimates.

	At least three children		At least four children		At least five children		At least six children	
VARIABLES	OLS	2SLS	OLS	2SLS	OLS	2SLS	OLS	2SLS
Number of siblings	-0.0093***	-0.0081	-0.0134***	-0.0197**	-0.0151***	-0.0194**	-0.0182***	-0.00664
	(0.00193)	(0.00703)	(0.00202)	(0.00796)	(0.00226)	(0.00930)	(0.00269)	(0.0113)
Child gender – male	0.0242***	0.0242***	0.0297***	0.0298***	0.0340***	0.0341***	0.0380***	0.0376***
	(0.00394)	(0.00394)	(0.00392)	(0.00392)	(0.00423)	(0.00424)	(0.00481)	(0.00483)
Child age	-0.0227***	-0.0230***	-0.0240***	-0.0227***	-0.0248***	-0.0240***	-0.0242***	-0.0261***
	(0.000918)	(0.00170)	(0.000945)	(0.00182)	(0.00104)	(0.00202)	(0.00119)	(0.00220)
Child is 2nd born	0.00706*	0.00627	0.00446	0.00780	0.00473	0.00671	0.0102	0.00609
	(0.00405)	(0.00576)	(0.00480)	(0.00622)	(0.00627)	(0.00748)	(0.00899)	(0.00980)
Child is 3rd born			0.00304	0.0100	-0.00230	0.00174	0.0130	0.00428
			(0.00610)	(0.0104)	(0.00736)	(0.0112)	(0.00976)	(0.0128)
Child is 4th born					-0.00864	-0.00226	0.00247	-0.0111
					(0.00900)	(0.0162)	(0.0113)	(0.0172)
Child is 5th born							0.00292	-0.0160
							(0.0129)	(0.0222)
Mother’s age	0.000461	0.000503	0.00125*	0.00100	0.00172**	0.00155*	0.00225**	0.00269***
	(0.000615)	(0.000654)	(0.000663)	(0.000722)	(0.000752)	(0.000842)	(0.000891)	(0.000985)
Mother’s education	0.00722***	0.00727***	0.00735***	0.00711***	0.00738***	0.00725***	0.00675***	0.00704***
	(0.000812)	(0.000850)	(0.000950)	(0.000957)	(0.000982)	(0.000995)	(0.00109)	(0.00114)
Household wealth index	0.0255***	0.0256***	0.0270***	0.0266***	0.0293***	0.0290***	0.0328***	0.0334***
	(0.00115)	(0.00127)	(0.00120)	(0.00132)	(0.00128)	(0.00141)	(0.00148)	(0.00161)
Constant	0.168***	0.164***	0.194***	0.219***	0.223***	0.243***	0.215***	0.150**
	(0.0175)	(0.0281)	(0.0190)	(0.0357)	(0.0229)	(0.0487)	(0.0302)	(0.0687)
Countries fixed effects included								
Observations	33,409	33,409	36,785	36,785	34,033	34,033	28,183	28,183
R-squared	0.283	0.283	0.286	0.285	0.284	0.284	0.278	0.278
H0 tot_enf exo		0.0355		0.686		0.227		1.092
P-value_H0		0.850		0.407		0.634		0.296

Robust standard errors are in parentheses. Standard errors are clustered at mother level. *** p<0.01, ** p<0.05, * p<0.1

The second and last robustness check is the handling of children who have never attended school. In the OLS and 2SLS regressions of Table 4, we set to zero, the number of school grade of children who have never attended school. These children represent 8 to 10 % of different subsamples. We explicitly account for this overstated number of zeros in our dependent variable by running Tobit and IV-Tobit regressions. Results are presented in Table 2a in the appendix. They are qualitatively similar to the estimates in Table 4.

4 Conclusion

The objective of this paper was to investigate the effect of the number of siblings on school achievement in the urban African context. Given endogeneity of the number of siblings in a household decision model, a correlation between the number of siblings and the education of children cannot be seen as a causal relationship. A two-stage least squares (2SLS) is then used to handle the endogeneity problem. Based on empirical evidence, we discarded the gender composition of siblings as an instrument for the number of siblings in the sub-Saharan African context. We instead use the advent of a twin birth as instrument. Having a twin birth is a random event and has a direct link to the number of siblings.

Given that having a twin birth is not very common, we use a very large data set. We pooled 29 recent Demographic and Health Survey data from 20 sub-Saharan countries. DHS data sets provided suitable data for our analysis and are comparable across countries. In our analysis, we account for country-specific effects and we cluster standard errors of estimated coefficient at child mother level. The results show a negative and significant relationship between the number of siblings and the schooling in families with at least four or five children. In smaller families and larger families, the effect is negative but not significant.

Bibliography

Angrist J. , Evans W. , 1998, Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size, *The American Economic Review*, Vol. 88, No. 3., pp. 450-477.

Angrist, J ., Lavy, V. & Schlosser A. (2010) “Multiple Experiments for the Causal Link Between the Quantity and Quality of Children”. Journal of Labor Economics, vol. 28(4), pp.773-823.

Azam, J-P, Gubert, F. (2005) “Those in Kayes. The Impact of Remittances on their Recipients in Africa”. Revue Economique, 56(6), pp.1331-1358.

Baland, J-M., Robinson, J. (2000) "Is Child Labor Inefficient? " Journal of Political Economy, vol.108(4), pp.663-679.

Basu, K., Van, P. (1998) "The Economics of Child Labor". American Economic Review, vol. 88(3), pp.412-27.

Becker, G. (1993). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* Chicago: University of Chicago Press, 3rd edn.

Becker, G., Lewis, H. (1973) "On the Interaction Between the Quantity and Quality of Children". Journal of Political Economy, vol. 81(2), pp.S279-88.

Becker, G., Tomes, N. (1986) "Human Capital and the Rise and Fall of Families". Journal of Labor Economics, vol. 4(3) July, pp.S1-39.

Black, S., Devereux, P. & Salvanes, K. (2005) "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education". The Quarterly Journal of Economics: MIT Press, vol. 120(2) May, pp.669-700.

Blake, J. (1981) “Family Size and the Quality of Children”. Demography 18(4) , pp.421-442.

Bloom, D., Canning, D. & Sevilla, J. (2003) *The Demographic Dividend: A New Perspective on the Economic Consequences of Population Change*. Population Matters Series. Santa Monica, California: Rand.

Cáceres-Delpiano, J. (2006) "The Impacts of Family Size on Investment in Child Quality". Journal of Human Resources, vol. 41(4) p738-754

Chernichovsky, D. (1985) "Socioeconomic and Demographic Aspects of School Enrollment and Attendance in Rural Botswana". Economic Development and Cultural Change, vol. 33(2), pp.319-332.

Conley, D. (2000) “Sibship Sex Composition: Effects on Educational Attainment”. Social Science Research, 29, pp.441–457.

Conley, D., Glauber, R. (2006) "Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility". Journal of Human Resources, vol. 41(4) pp. 722-737

Downey, D. (1995) "When Bigger Is Not Better: Family Size, Parental Resources, and Children's Educational Performance". American Sociology Review, 60, pp.746-61.

Dumont, A. (1890) *Dépopulation et Civilisation. Etude démographique*. xii + p.520. Paris: Lecrosnier et Babé.

Edmonds, E. (2006) "Understanding Sibling Differences in Child Labor". Journal of Population Economics, Springer, vol. 19(4), pp.795-821.

Eloundou-Enyegue, P., Williams, L. (2006) "Family Size and Schooling in Sub-Saharan African Settings: A Reexamination". Demography, vol. 43(1) , pp. 25-52.

Emerson, P., Souza, A. (2002) "Birth Order, Child Labor and School Attendance in Brazil". Working Papers 0212, Nashville, Tennessee: Department of Economics, Vanderbilt University.

Filmer, D., Friedman, J. & Schady, N. (2009) "Development, Modernization, and Childbearing: The Role of Family Sex Composition". World Bank Econ Rev, vol. 23 (3), pp.371-398.

Gomes, M. (1984) "Family Size and Educational Attainment in Kenya". Population and Development Review, vol. 10, pp.647-60.

Goux, D., Maurinc, E. (2005) "The Effect of Overcrowded Housing on Children's Performance at School". Journal of Public Economics, 89 pp.797– 819.

Greene, W. (2003) *Econometric Analysis*. 5th edn. London: Prentice-Hall International.

Jha, P., Kumar, R., Vasa, P., Dhingra, N., Thiruchelvam, D., Moineddin, R. (2006) "Low Male-to-Female Sex Ratio of Children Born in India: National Survey of 1·1 Million Households". The Lancet, vol. 367 (9506), pp.211-218.

Knodel, J., Napaporn, H., Werasit, S. (1990) "Family Size and the Education of Children in the Context of Rapid Fertility Decline". Population and Development Review, vol. 16(1), pp.31-62.

Knodel, J., Wongsith, M. (1991) "Family Size and Children's Education in Thailand: Evidence from a National Sample". Demography, 28, pp.119-31.

Kuepie Mathias (2011) ″L'impact de la Solidarité Familiale sur le Niveau de Vie et la Pauvreté des Personnes Agées au Mali ″. CEPS/INSTEAD Working Papers, n°2011-49.

Locoh, T. (1988) "Structures familiales et changements sociaux" in : Tabutin, Dominique (ed.), Population et société en Afrique au Sud du Sahara , Paris: L'Harmattan, pp. 441-478.

Lu, Yao and Treiman, D. (2005) "The Effect of Family Size on Educational Attainment in China: Cohort Variations". Mimeo. Los Angeles: Department of Sociology, UCLA.

Lu, Yao, Treiman, Donald J. (2007) "The Effect of Labor Migration and Remittances on Children's Education Among Blacks in South Africa". Revised version of a paper presented at the Annual Meeting of the Population Association of America, Los Angeles, 30 March-2

April. [Los Angeles: UCLA, California Center for Population Research, Population Working Paper PWP-CCPR-2007-001.]

Montgomery, M., Aka, K. & Raylynn, O. (1995) ″The Tradeoff Between Number of Children and Child Schooling: Evidence from Côte d'Ivoire and Ghana″. LSMS Working Papers 111. Washington, DC: The World Bank.

Pôle de Dakar (2007) "Dakar+7 – Education for All in Africa : Top Priority for Integrated Sector-Wide Policies". UNESCO/BREDA publication

Rakoto-Tiana N. (2012) ″Child Fostering and School Enrollment in Rural Madagascar. Empirical Evidence Based on Longitudinal Data.″ DIAL working paper DT/2012/01

Rosenzweig, M.R. & Wolpin, K.I. (1980) "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment". Econometrica, vol. 48(1)January, pp. 227-240.

Rosenzweig, M.R. & Zhang,J. (2009) "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's 'One-Child Policy". Review of Economic Studies Limited, vol 76 (3) , pp.1149–1174.

Schultz T.W. (1961) "Investment in Human Capital". The American Economic Review, vol 51(1), pp.1-17.

Sewell, W. (1968) "Parents' Education and Children's Educational Aspirations and Achievements". American Sociological Review, Vol 33 (2), pp.191-209.

Stock J.H, Wright, J.H. & Yogo M. (2002) "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments". American Statistical Association Journal of Business & Economic Statistics, vol. 20(4), October 2002. DOI 10.1198/073500102288618658

Sudha, S. (1997) "Family Size, Sex Composition and Children's Education: Ethnic Differentials over Development in Peninsular Malaysia". Population Studies, 51, pp.139-51.

Temesgen, K. (2007) "Do Remittances Encourage Investment in Education? Evidence from Eritrea". Journal of African studies, vol. 4(1).

Tenikue, M., Verheyden, B. (2010) "Birth Order and Schooling: Theory and Evidence from Twelve Sub-Saharan Countries". Journal of African Economies, vol. 19(4), pp.459-495.

Eijck Van, K., De Graaf, P. (1995) "The Effects of Family Structure on the Educational Attainment of Siblings in Hungary". European Sociological Review, 11, pp.273-292.

VANDERMEERSCH C. (2000) ″Les Enfants Confiés au Senegal″. Economic Demographics
Thesis. Paris : Institut d'Études Politiques de Paris, 502p+ annexes.

World Bank (2011) Africa Development Indicators, Washington D.C: The World Bank, 196p.

Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data.* Cambridge, Massachusetts: The MIT Press.

Appendix

Table 1a. Impact of gender composition of elder children on the probability of having an additional child (linear probability model, mother aged 40 or more)

VARIABLES	Families with five or more children	Families with four or more children	Families with three or more children	Families with two or more children
0-1 girl; 4-5 boys (ref)				
5-4 girls; 0-1 boy	-0.0393			
	(0.0276)			
3 boys; 2 girls	-0.0158			
	(0.0231)			
3 girls; 2 boys	-0.0288			
	(0.0237)			
0 girls; 4 boys (ref)				
4 girls; 0 boys		-0.0135		
		(0.0382)		
3 boys; 1 girl		-0.0183		
		(0.0251)		
3 girls; 1 boy		-0.0314		
		(0.0259)		
2 boys; 2 girls		-0.0623**		
		(0.0246)		
0 girls; 3 boys (ref)				
3 girls; 0 boys			0.00707	
			(0.0230)	
2 boys; 1 girl			0.0154	
			(0.0166)	
2 girls; 1 boy			0.00370	
			(0.0171)	
0 girls; 2 boys (ref)				
2 girls; 0 boys				-0.00785
				(0.0136)
1boy; 1 girl				-0.00950
				(0.0109)
Observations	2,724	3,659	4,604	5,429
R-squared	0.094	0.096	0.108	0.082

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

The variables mother age, education, household wealth and country fixed effects are included in the regression but coefficients are not shown.

Table 2a. Impact of fertility on school achievement of elder singletons, Tobit and IV-Tobit estimates

	At least three children		At least four children		At least five children		At least six children	
VARIABLES	Tobit	IV- Tobit	Tobit	IV- Tobit	Tobit	IV- Tobit	Tobit	IV- Tobit
Number of siblings	-0.225***	-0.0707	-0.229***	-0.107*	-0.248***	-0.117*	-0.246***	-0.0389
	(0.0143)	(0.0497)	(0.0154)	(0.0580)	(0.0175)	(0.0704)	(0.0206)	(0.0894)
Child gender – male	-0.0555**	-0.0588**	-0.0186	-0.0206	0.0393	0.0363	0.116***	0.109***
	(0.0255)	(0.0255)	(0.0256)	(0.0256)	(0.0282)	(0.0283)	(0.0326)	(0.0329)
Child age	0.647***	0.614***	0.621***	0.596***	0.596***	0.571***	0.573***	0.538***
	(0.00604)	(0.0118)	(0.00633)	(0.0129)	(0.00719)	(0.0148)	(0.00820)	(0.0168)
Child is 2nd born	0.0333	-0.0608	0.0449	-0.0190	0.0820*	0.0218	0.114*	0.0388
	(0.0259)	(0.0392)	(0.0310)	(0.0427)	(0.0421)	(0.0526)	(0.0639)	(0.0714)
Child is 3rd born			-0.0164	-0.150**	0.00249	-0.120	0.101	-0.0562
			(0.0413)	(0.0738)	(0.0511)	(0.0818)	(0.0699)	(0.0961)
Child is 4th born					-0.0331	-0.227*	0.0174	-0.227*
					(0.0629)	(0.119)	(0.0808)	(0.131)
Child is 5th born							0.0313	-0.309*
							(0.0926)	(0.171)
Mother's age	0.0405***	0.0456***	0.0418***	0.0465***	0.0509***	0.0563***	0.0575***	0.0654***
	(0.00446)	(0.00481)	(0.00495)	(0.00543)	(0.00556)	(0.00620)	(0.00644)	(0.00732)
Mother's education	0.0912***	0.0973***	0.0974***	0.102***	0.0904***	0.0943***	0.0780***	0.0831***
	(0.00916)	(0.00984)	(0.0109)	(0.0116)	(0.0112)	(0.0119)	(0.0120)	(0.0127)
Household wealth index	0.310***	0.322***	0.329***	0.337***	0.357***	0.365***	0.387***	0.398***
	(0.00934)	(0.0104)	(0.0101)	(0.0111)	(0.0108)	(0.0118)	(0.0119)	(0.0131)
Constant	-6.294***	-6.798***	-6.227***	-6.705***	-6.400***	-7.021***	-6.503***	-7.663***
	(0.164)	(0.224)	(0.173)	(0.277)	(0.214)	(0.383)	(0.261)	(0.554)
Countries fixed effects included								
Observations	33,409	33,409	36,785	36,785	34,033	34,033	28,183	28,183
H0 tot_enf exo		10.63		4.842		3.759		5.693
P-value_H0		0.00111		0.0278		0.0525		0.0170

Robust standard errors are in parentheses. Standard errors are clustered at mother level. *** p<0.01, ** p<0.05, * p<0

CEPS
INSTEAD

3, avenue de la Fonte
L-4364 Esch-sur-Alzette
Tél.: +352 58.58.55-801
www.ceps.lu