Sibling Composition and School Enrollment in Ghana*

Kehinde Ajayi

University of California, Berkeley

March 2009

Abstract

This paper reexamines household decisions about investment in child education. In particular, I consider the role of siblings in determining school enrollment. Previous studies have found correlations between sibling characteristics and education outcomes of school children but this evidence has largely resulted from cross-sectional studies, making it difficult to identify a causal mechanism for sibling rivalry. This paper contributes to existing literature by presenting a longitudinal perspective. Using household data from all five rounds of the Ghana Living Standards Survey administered in 1987/88, 1988/89, 1991/92, 1998/99 and 2005/06, I estimate the importance of sibling sex in explaining the educational attainment of school-age children in each of the survey samples, holding the number of siblings and birth order fixed. I find that the sibling sex coefficients have changed over the period of study and I attempt to interpret the underlying motivations for household behavior in light of these changes.

*The author thanks Ghana Statistical Service for providing access to the Ghana Living Standards Survey data.

1 Introduction

There are many reasons to believe that household conditions may impact the educational outcomes of children. One of the most influential of these conditions may be the presence of siblings. Existing literature on the effect of sibling composition primarily examines completed education and employment outcomes of adult or older-aged individuals based on reported sibling characteristics. This paper instead examines the household decision about child enrollment for primary school-aged children.[1] In particular, I focus on three aspects of sibling characteristics: 1) number of siblings, 2) birth order and 3) sex of siblings.

Becker's classic (1960) quantity-quality trade-off model provides the foundation for research on the relationship between family size and children's education. Becker assumes that parents have a fixed budget constraint subject to which they must decide on the number and quality of offspring, where quality is the result of investments in human capital, such as health and education expenses. Parents therefore simultaneously decide on the desired quantity and quality of children choosing between a small number of high quality children or a high number of low quality children at the extremes. A central prediction of this model is that the quality of children in a household should decrease as the number of children increases, thus larger families should have lower investments in child education. Although theoretical implications of the quantity-quality model are clear, empirical studies are difficult to implement because of the problem of endogeneity. A correlation between large family size and low education may reflect idiosyncratic factors that cause a household to desire lots of children and to undervalue education, as opposed to capturing the direct *effect* of large sibship size per se.

Indeed, more recent examinations of Becker's theory suggest that the number of siblings may not matter so much as birth order particularly in the absence of budget constraints. Black, Devereux and Salvanes (2005) use twins as an instrument for the number of siblings in their study of the effect of family size on education and labor market outcomes in Norway. The authors find that the correlation between sibship size and education outcomes essentially disappears once fertility has been exogenously determined. Yet a birth order effect persists, thus first borns in a family of five fair as well as first borns in a family of two. This finding is not entirely inconsistent with Becker's quantity-quality trade-off but reinforces the importance of having limited resources as necessary condition for the theory's predictions to hold.

Sibling sex is another key factor addressed in the literature. A natural extension of Becker's model suggests that if households have a preference for educating sons and face a budget constraint, then having sisters should improve a child's education outcomes. Morduch (2000) analyzes data from two African contexts and examines the notion that there exists a "rivalry for scarce resources in which parents favor sons, and children do better with sisters" (p. 405). This

[1] The phenomenon of delayed enrollment is in itself informative because human capital theory predicts that schooling should begin at the earliest possible age when the value of child's time is lowest (Glewwe and Jacoby, 1995).

son preference may result from several factors including differences in returns to schooling, parental preferences, and cultural norms, but the resulting implication is that having sisters lessens the demand on household resources and thus increases the probability of parents' investment in one's own human capital.

The effect of sibling sex does not appear to be consistent across contexts and may function through alternative mechanisms particularly in the absence of budget constraints. Butcher and Case (1994) study completed education of adults in the United States and find that it is better for girls to have all brothers than all sisters. They posit various possible mechanisms for their findings, including differing "prices" of raising sons and daughters and they also appeal to theories of developmental psychology and the acquisition of masculine and feminine traits in childhood. To the extent that gender-based traits affect attitudes to and capacity for classroom learning, their acquisition may be critical to educational attainment and may be dependent on siblings. Another alternative explanation is the role of reference groups - if parents respond to children based on the availability of a peer group, then a single daughter in a household of sons may be treated differently than if she had a sister. Butcher and Case emphasize that these explanatory theories have different implications for whether the share or absolute number of sisters matters.

This paper proposes to make four main contributions to the literature. First, I seek to provide more descriptive empirical evidence on the relationship between sibling characteristics and education attainment in Africa. Second, I examine trends over time and consider the extent to which the effects of sibling composition evolve as the socio-economic climate changes. Third, I attempt to address potential causal explanations for sibling characteristic effects by estimating differential effects for various subgroups and analyzing the impact of sibling composition using several different specifications. Finally, I focus my analysis on children (as opposed to adults) which has both costs and benefits. On the positive side, studying children in their household environment allows the researcher to construct an objective measure of sibling features (instead of relying on subjective recollections of adult respondents). However, this verifiability comes at the cost of limiting the study to observe short-term outcomes only as several interesting longer term outcomes such as completed education and labor force participation are not yet realized.

The remainder of this paper is structured as follows: the next section develops a theoretical framework for analyzing the relationship between sibling sex and school enrollment, section three describes the data, section four outlines my methodology, section five presents the results and section six concludes.

2 Theoretical Framework

At the foundation of most sibling composition models is Becker's (1960) formulation of a quantity-quality theory in which families' fertility decisions reflect a trade-off between the quantity and quality of children. The key prediction of Becker's model is that families with more children should have lower in-

vestments in education. Beyond this first-order prediction on the relationship between sibship size and educational investment are a series of more nuanced predictions about the effect of siblings' sex and birth order. This section outlines some possible mechanisms through which sibling composition might affect school enrollment and focuses on the effect of variation in sibling sex for two main reasons. First, unlike number of children, the sex of children is relatively exogenous. While households may quite readily decide on the number of children to have, sex selection is less easy to achieve and there is little evidence of strong gender-biased mortality or fertility patterns in Ghana. Nonetheless, gender may be a common basis for discrimination in human capital investment in contexts with gender-based cultural norms or preferences and differential costs of (or returns to) schooling.[2] The persistence of a gender gap in education enrollment reinforces this point and thus presents a second motivation for examining the effect of sibling sex.

2.1 Budget Constraints

If a household has limited resources and faces borrowing constraints and the economic implications of schooling differs by the sex of children then we might be interested in the number (and percentage) of children within the household who are female. Labor supply considerations could also explain differences in investment in education. If siblings provide care for younger siblings or contribute to household economic activity and girls are predominantly involved in supplying household labor, then the number of sisters and particularly the number of older sisters could be important.

2.2 Reference Groups

There may be non-economic explanations for the effect of sibling sex. In particular, if having any other siblings of the same sex changes expectations for the child in question (by altering the child's reference group) then an informative measure of sibling composition would be captured with an indicator variable for presence of any siblings of the same sex. For example, if having other sisters changes a daughter's reference group then we would be interested in the presence of any sisters, and not in the exact number. This model suggests a non-linear effect of having sisters, in which the presence of at least one sister is the critical factor and the presence of additional sisters has a much lesser effect.

3 Data

This paper uses data from all five rounds of the Ghana Living Standards Survey (GLSS), collected in 1987/88, 1988/89, 1991/92, 1998/99 and 2005/06 to further examine these theories. The surveys provide comparable cross-sectional data on

[2] Garg and Morduch (1998) analyze the correlation between sibling sex and health outcomes in Ghana and find a significant positive effect of having sisters.

households in the five periods. For comparison purposes, I model my sample after Morduch (2000) who focuses on adolescent children aged 13 to 16 years old in Tanzania and South Africa. I also replicate the analysis using a sample of younger 6 to 12 year olds to examine changes in effects of sibling composition over the life course. Tables (1a) and (1b) below provide a list of summary statistics.

Table 1a: Summary Statistics (13 to 16 year olds)

	GLSS 1	GLSS 2	GLSS 3	GLSS 4	GLSS 5
Years of schooling	4.83	4.68	5.00	-	4.46
	(3.27)	(3.08)	(3.13)		(2.69)
Currently enrolled	0.64	0.66	0.76	0.83	0.81
	(0.48)	(0.47)	(0.43)	(0.38)	(0.39)
Ever attended school	0.79	0.81	0.83	0.90	0.87
	(0.41)	(0.39)	(0.35)	(0.30)	(0.34)
Number of sisters	1.35	1.30	1.24	1.06	1.20
	(1.45)	(1.39)	(1.31)	(1.16)	(1.35)
Number of older sisters	0.19	0.18	0.19	0.16	0.20
	(0.44)	(0.45)	(0.43)	(0.38)	(0.45)
Number of siblings	2.87	2.85	2.70	2.27	2.52
	(2.44)	(2.50)	(2.21)	(1.87)	(2.32)
Any sisters	0.67	0.64	0.65	0.60	0.62
	(0.47)	(0.48)	(0.48)	(0.50)	(0.48)
Any brothers	0.67	0.68	0.68	0.64	0.63
	(0.69)	(0.47)	(0.47)	(0.48)	(0.48)
Percent sisters	0.40	0.38	0.39	0.38	0.39
	(0.35)	(0.45)	(0.36)	(0.38)	(0.37)
Birth order	1.46	1.41	1.44	1.37	1.45
	(0.67)	(0.65)	(0.65)	(0.58)	(0.66)
Female	0.46	0.46	0.44	0.50	0.49
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Age	14.00	14.00	14.00	14.00	14.01
	(0.83)	(0.82)	(0.82)	(0.83)	(0.82)
Household head male	0.70	0.69	0.68	0.68	0.74
	(0.46)	(0.46)	(0.47)	(0.47)	(0.44)
Household head age	50.64	50.40	50.62	49.68	50.08
	(13.03)	(13.59)	(12.39)	(12.37)	(12.27)
Household head schooled	0.45	0.46	0.48	0.64	0.58
	(0.50)	(0.50)	(0.50)	(0.48)	(0.49)
Household size	9.99	9.94	7.08	6.72	7.11
	(4.64)	(4.88)	(3.04)	(2.74)	(3.50)
Urban			0.34	0.35	0.36
			(0.47)	(0.48)	(0.48)
Observations	1098	1085	1536	2042	2892

Table 1b: Summary Statistics (6 to 12 year olds)

	GLSS 1	GLSS 2	GLSS 3	GLSS 4	GLSS 5
Years of schooling	1.28	1.19	1.48	-	1.05
	(1.54)	(1.47)	(1.59)		(1.41)
Currently enrolled	0.63	0.70	0.72	0.83	0.82
	(0.48)	(0.46)	(0.45)	(0.38)	(0.39)
Ever attended school	0.72	0.75	0.74	0.87	0.83
	(0.45)	(0.43)	(0.44)	(0.34)	(0.38)
Number of sisters	1.44	1.41	1.42	1.25	1.34
	(1.42)	(1.34)	(1.45)	(1.20)	(1.33)
Number of older sisters	0.60	0.58	0.58	0.60	0.62
	(0.85)	(0.81)	(0.84)	(0.82)	(0.87)
Number of siblings	3.04	3.00	3.05	2.54	2.79
	(2.30)	(2.31)	(2.50)	(1.78)	(2.21)
Any sisters	0.72	0.72	0.72	0.70	0.70
	(0.45)	(0.45)	(0.45)	(0.46)	(0.46)
Any brothers	0.75	0.73	0.75	0.70	0.71
	(0.43)	(0.44)	(0.43)	(0.46)	(0.45)
Percent sisters	0.43	0.43	0.43	0.44	0.44
	(0.34)	(0.35)	(0.35)	(0.37)	(0.36)
Birth order	2.34	2.31	2.33	2.23	2.35
	(1.41)	(1.36)	(1.48)	(1.20)	(1.40)
Female	0.49	0.50	0.49	0.49	0.49
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Age	8.27	8.34	8.46	8.45	8.34
	(1.67)	(1.67)	(1.65)	(1.68)	(1.72)
Household head male	0.72	0.70	0.71	0.69	0.78
	(0.45)	(0.46)	(0.45)	(0.46)	(0.41)
Household head age	47.22	47.10	46.06	46.51	46.87
	(13.68)	(13.75)	(12.74)	(12.71)	(12.77)
Household head schooled	0.48	0.52	0.54	0.65	0.56
	(0.50)	(0.50)	(0.50)	(0.48)	(0.50)
Household size	9.38	9.39	7.00	6.54	7.04
	(4.51)	(4.67)	(3.30)	(2.60)	(3.37)
Urban			0.33	0.31	0.30
			(0.47)	(0.46)	(0.46)
Observations	2734	2626	3568	4806	6096

Notably, there has been a slight decrease in the mean education attainment over time but also a decrease in variance suggesting a convergence in education outcomes. Additionally, the share of currently enrolled and ever enrolled children has increased. With regard to family characteristics, household size and number of siblings has decreased.

4 Methodology

The baseline specification also draws on Morduch (2000). For each child i in household j, I estimate an ordinary least squares regression with the number of completed years of schooling (E_{ij}) as the outcome of interest.[3]

$$E_{ij} = \alpha + \beta S_{ij} + \gamma X_{ij} + \delta Z_j + \epsilon_{ij}$$

Right hand variables include a vector of sibling composition measures (S_{ij}), including the number of siblings (and the square), number of sisters (and the square), number of older sisters (and the square), and birth order, as well as sex interacted with sibling composition measures to estimate differential impacts by gender. Child characteristics (X_{ij}) are captured by an indicator for age by year and sex. Also included is a vector of household-level covariates (Z_j) with characteristics of the household head (education, sex, and age) and dummy variables for household income decile, region, and residence in an urban location (these last three measures are excluded from the GLSS 1 and 2 regressions as the data was not available).

Finally, I estimate three additional models to examine Butcher and Case's (1994) predictions about the mechanism for sibling sex effects, using share of sisters and an indicator for the presence of any brother or any sister as the main sibling composition variables. Once again, I interact the sibling composition terms with a gender dummy in order to estimate differential impacts for boys and girls.

5 Results

Table 2 reports the results of the baseline regression.[4] An examination of the five periods reveals that the female coefficient has decreased over time reflecting a move toward gender parity in educational attainment. Accordingly, the coefficient on number of sisters has decreased over time as well. Whenever significant, the sisters coefficient shows up as being positive which supports Morduch's conjecture that households invest less in daughters' education leaving a higher amount of resources available to invest in other children, given a fixed budget constraint. The female and sibling interaction terms do not show up with significance which suggests that the effect of having sisters is relatively equal for male and female children - a girl with sisters is not significantly more advantaged than a boy with sisters after the level effect of being female has been taken into account.

[3] Additional specifications examine current school enrollment, ever attendance, and highest level of education attained as robustness tests, using OLS and probit regressions.

[4] My preferred specification excludes the gender interacted terms although they were included in other specifications not reported here and were never statistically significant.

Table 2: Years of Completed Schooling (13 to 16 year olds)

	GLSS 1	GLSS 2	GLSS 3	GLSS 5
Female	-0.798***	-0.659***	-0.371**	-0.064
Number of sisters	0.350	0.318	0.125	0.254*
Number of sisters^2	-0.096***	-0.057	-0.002	-0.064**
Number of siblings	-0.227	-0.217	0.145	-0.106
Number of siblings^2	0.025*	0.012	-0.013	0.012
Number of older sisters	-0.361	-0.556	0.525	-0.468*
Number of older sisters^2	0.145	0.281	-0.415	0.212
Birth order	0.317	0.301	-0.027	0.143
Observations	1,098	1,085	1,536	2,892

Notes: The dependent variable is completed year of schooling for children aged 13 to 16. OLS regressions are estimated with robust standard errors. Specifications include an indicator for child's age by year, as well as education, sex, and age of the household head and dummy variables for household income decile, region, and residence in an urban location (these last three measures are excluded from the GLSS 1 and 2 regressions as the data was not available). Asterisks indicate significance levels: * $p<0.05$; ** $p<0.01$, and *** $p<0.001$. Data on completed years of schooling was missing from GLSS 4 so the sample is excluded from this analysis.

The results largely support the predictions of Becker's quantity-quality model. With the exception of data in GLSS 3, the coefficient on siblings is negative as expected (albeit insignificant). A negative coefficient is consistent with the notion that households with a larger number of children are less likely to invest in their childrens' education. Results for the younger cohorts are generally of a smaller magnitude than those for the adolescent group (see Table 3 below).

Table 3: Years of Completed Schooling (6 to 12 year olds)

	GLSS 1	GLSS 2	GLSS 3	GLSS 5
Female	-0.113*	-0.083	-0.081*	-0.003
Number of sisters	0.077	-0.031	0.002	0.085*
Number of sisters^2	-0.018	0.013	-0.006	-0.015**
Number of siblings	-0.042	0.016	0.007	-0.037
Number of siblings^2	0.005	-0.003	0.004	-0.002
Number of older sisters	0.115	0.038	0.105	-0.087*
Number of older sisters^2	-0.039	-0.006	-0.008	0.019*
Birth order	0.035	0.019	-0.028	0.069**
Observations	2,734	2,626	3,568	6,096

Notes: same as above.

5.1 Mechanisms

Tables (4) and (5) report results from the additional specifications which examine possible mechanisms through which sibling sex exerts its influence. There is some evidence to suggest that the share of sisters matters, but the results are mostly insignificant.

Table 4: Having a Sister and Years of Completed Schooling (13 to 16 year olds)

	GLSS 1	GLSS 2	GLSS 3	GLSS 5
Female	-1.012**	-0.618*	-0.143	-0.124
Any sisters	-0.072	0.650*	0.359	0.090
Female × any sisters	0.276	-0.038	-0.359	0.102
Number of siblings	-0.031	-0.265*	0.163*	-0.056
Number of siblings^2	-0.008	0.013	-0.011*	0.002
Birth order	0.289*	0.252	-0.042	0.067
Observations	1,098	1,085	1,536	2,892

Notes: same as above.

Table 5: Share of Sisters and Years of Completed Schooling (13 to 16 year olds)

	GLSS 1	GLSS 2	GLSS 3	GLSS 5
Female	-0.735**	-0.625*	-0.163	-0.155
Percent sisters	0.137	0.338	0.510*	-0.109
Female × percent sisters	-0.232	-0.076	-0.543	0.246
Number of siblings	-0.018	-0.129	0.182**	-0.018
Number of siblings^2	-0.008	0.005	-0.012*	-0.000
Birth order	0.296*	0.245	-0.043	0.066
Observations	1,098	1,085	1,536	2,892

Notes: same as above.

5.2 Robustness Checks

5.2.1 Other Outcomes

To check the robustness of these results, I replicated the analysis using three alternative measures of educational attainment: 1) completed education level, 2) current enrollment status and 3) whether a child had ever attended school. The results were qualitatively similar for these additional specifications.

5.2.2 Other Definitions of Siblings and Reference Groups

The main specification defines siblings as biological brothers and sisters in the household and as other children with the same relationship to the household

head for those who are not biologically related. It is plausible to imagine that there may be differential effects of sibling sex for children based on their relationship to the household head and that the presence of other non-related children might also affect education outcomes. I therefore estimate additional specifications to examine the effects of having any other children of school age in the household. I also estimate differential effects for school-aged children who are not related to the household head or spouse. Results are estimated for each sample separately.

When I focus on children of the household head only, I find that the sibling effect is stronger but the sister effect smaller. Moreover, estimates from a model that includes all children in the household as reference group are qualitatively similar to estimates from the baseline specification (results not reported here).

5.2.3 Attrition

On final potential source of bias in results is missing data. In most cases there was only a moderate incidence of non-response with attrition rates of around 3 percent. A notable exception is in the GLSS 3 sample of young (6 to 12 year old) children in which 9 percent of the sample was dropped due to missing data on educational attainment. The incidence of missing data was lower for school attendance and current enrollment outcomes. A potentially useful exercise for future would be to examine whether attrition is systematically correlated with any of the key variables.

Table 6: Attrition Due to Missing Data

	Cohort	Full (N)	Regression (N)	Missing	Attrition Rate
GLSS 1	6 to 12	2812	2734	78	2.77 %
	13 to 16	1136	1098	38	3.35 %
GLSS 2	6 to 12	2723	2626	97	3.56 %
	13 to 16	1130	1085	45	3.98 %
GLSS 3	6 to 12	3929	3568	361	9.19 %
	13 to 16	1547	1536	11	0.71 %
GLSS 4	6 to 12	4954	4806	148	2.99 %
	13 to 16	2116	2042	74	3.50 %
GLSS 5	6 to 12	6118	6099	19	0.31 %
	13 to 16	2957	2892	65	2.20 %

6 Conclusion

Overall, this analysis provides more suggestive evidence that the presence of sisters has a positive effect on the educational attainment of children and that this effect has decreased over time as Ghana moves towards achieving gender

parity in school enrollment. However, there are some limitations to acknowledge in this study and opportunities for future research. One challenge to overcome is the potential endogeneity of household composition, particularly at older ages as there may be selective migration of household members. In particular,a comparison of Tables (1a) and (1b) reveals that the percent of sisters is lower at older ages (39 percent for adolescents as compared with 44 percent for the younger age-group). This difference may reflect endogenous decisions by households concerning their desired child quality outcomes, which would bias regression estimates. I have attempted to address this concern by focusing on relatively young children (under the age of 16), and estimating additional specifications which include the total number of children in the house instead of just the number of siblings. Another concern is the small sample size in certain years which may contribute to the low significance of results under certain specifications. A possible approach to address this issue would be to use census data or a larger sample to achieve more statistical power. Lastly, as stated earlier, this study only considers current enrollment, not completed schooling or labor force outcomes. Thus, there is the potential that some catch up occurs later in life so that the effects observed in childhood dissipate over time even if certain children cut short or postpone their schooling. It would be interesting to observe longer term outcomes using panel data or an alternative analysis of adults using reported information on siblings.

References

[1] Becker, Gary (1960), "An Economic Analysis of Fertility", in Gary Becker, ed. Demographic and Economic Change in Developed Countries, Princeton University Press: Princeton, NJ.

[2] Black, Sandra, Paul Devereux and Salvanes (2005). "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education", Quarterly Journal of Education, 669-699.

[3] Butcher, Kristin and Anne Case (1994). "The Effect of Sibling Sex Composition on Women's Education and Earnings", Quarterly Journal of Economics, 109(3): 531-563.

[4] Garg, Ashish and Morduch, Jonathan (1998), "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana", Journal of Population Economics, 11(4): 71-93.

[5] Glewwe, Paul and Hanan G. Jacoby (1995), "An Economic Analysis of Delayed Primary School Enrollment in a Low Income Country: The Role of Early Childhood Nutrition", The Review of Economics and Statistics, 77(1): 156-169.

[6] Morduch, Jonathan (2000), "Sibling rivalry in Africa", American Economic Review 90(2): 405-409.