DISCUSSION PAPER SERIES

IZA DP No. 1789

Birth Order, Educational Attainment and Earnings: An Investigation Using the PSID

Jasmin Kantarevic
Stéphane Mechoulan

September 2005

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Birth Order, Educational Attainment and Earnings: An Investigation Using the PSID

Jasmin Kantarevic
Ontario Medical Association and IZA Bonn

Stéphane Mechoulan
University of Toronto

Discussion Paper No. 1789
September 2005

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 1789
September 2005

ABSTRACT

Birth Order, Educational Attainment and Earnings: An Investigation Using the PSID

Whether siblings of specific birth order perform differently has been a longstanding open empirical question. We use the family tree structure of the PSID to examine two claims found in the literature: whether being early in the birth order implies a distinct educational advantage, and whether there exists, within large families, a pattern of falling then rising attainment with respect to birth order, to the point where it becomes best to be last-born. Drawing from OLS and family fixed effects estimations, we find that being first-born confers a significant educational advantage that persists when considering earnings; being last-born confers none.

JEL Classification: I2, J1

Keywords: birth order, family size, education

Corresponding author:

Jasmin Kantarevic
Ontario Medical Association
525 University Avenue
Toronto, ON
Canada, M5G 2K7
Email: jasmin_kantarevic@oma.org

1. Introduction

Whether siblings of specific birth order perform differently has been an open empirical question for decades. Surprisingly, economists have paid little attention to this issue, although it is fundamental to our understanding of the intra-household allocation of resources.[1]

In this study, we examine two claims found in the literature: whether being early in the birth order implies a distinct educational advantage; and whether there exists, within large families, a pattern of falling then rising attainment with respect to birth order, to the point that it becomes best to be last-born.

The empirical results presented here, drawn from the Panel Study of Income Dynamics (PSID), show that being first-born or even among the first born does confer an advantage, while being last-born or among the last-born actually confers none. We believe that the reason for the discrepancy between previous findings and ours is that we control for the age of the mother at childbirth.

We find that the age of the mother at childbirth is positively correlated with a child's education. At the same time, the age of the mother at childbirth is mechanically, positively correlated with a child's birth order. Therefore, the two effects of birth order and age of the mother at childbirth may compete against one another. We show that the

---

[1] Birdsall (1979), Behrman (1986), Behrman and Taubman (1986), Kessler (1991). We elaborate on a few other studies in more details in the paper.

omitted variable bias results in a clear offset of the birth order effect. This provides a new key element in our understanding of siblings' competition for parental inputs.

Arguably, a causal interpretation of the previous analysis may suffer from the potential correlation of some of our regressors with the error term in our OLS estimations. Total number of siblings, the age of the mother at childbirth, and other covariates (parental education, whether all siblings report, whether both parents report) could be correlated with unobservable socio-economic characteristics. In particular, the precise causal determination of early motherhood on children's academic outcomes has received considerable attention (*e.g.*, Geronimus, Korenman and Hillemeier, 1994; Hofferth and Reid, 2003; Lopez-Turley, 2003), following an even larger debate on the consequences of early pregnancy on mothers themselves.

Obviously, the age of the mother at childbirth is linked to a number of variables that are expected to affect a child's educational attainment. Younger mothers are more likely to be single and have less human capital, which in turn may disproportionately influence the quality of the maternal inputs to their first-born etc. Also, there are adverse effects of unplanned motherhood that may dissipate over time (Bronars and Grogger, 1993). Yet, even if early motherhood does not *cause* lower educational attainment for a child, it is still possible that first-borns perform relatively better, conditional on early motherhood.

It would be very difficult to find compelling instrumental variables for all our potentially endogenous regressors. Therefore, to provide additional credibility to our results, we use a fixed effects model, which by construction removes variables that are constant *within* a family. As such, we take care of unobserved family-level heterogeneity, which may have previously led to bias. The results on birth order are broadly consistent with our initial ones. Hence, like Black, Devereux, and Salvanes (2005) in their analysis of family composition on children's education in Norway, we believe that our estimated birth order effects do not reflect omitted family characteristics.

To understand better our results, we test whether the effect we find is in the higher educational realm, where financing matters. In particular, we investigate whether birth order influences the probability of completing high school. Both OLS and FE results show that the first born lead is already revealed at the high school stage.[2] Yet, the exact mechanism through which first borns appear to be advantaged is not fully identifiable from our data.[3]

To provide a final check of the robustness of the results, we estimate the impact of birth order on hourly earnings. The same patterns emerge, *e.g.*, when age of the mother at birth is omitted, we find no effect, whereas when we include it, we find a strong positive influence of birth order on hourly earnings.

The rest of the paper is organized as follows. Section two presents our data. In section three, we show how birth order affects various educational outcomes through different estimation strategies, stressing the importance of age of the mother at child birth as a potential source of model misspecification. Section four then extends the analysis to earnings. Finally, section five concludes.

---

[2] We also look at the probability of repeating a grade conditional on high school completion and at the probability of attending post secondary education conditional on high school graduation, none of which seems significantly influenced by birth order.

[3] To the extent that one can separate spacing from siblings' size, we do not find that this premium is influenced by the age difference between first and last born.

2. Description of the data

Our data come from the Childbirth and Adoption History File (CAHF), a special supplemental file of the PSID. The CAHF covers eligible people living in a PSID family at the time of the interview in any wave from 1985 through 2001. Eligible persons are defined as heads or wives of any age and other members of the family unit aged 12-44 at the time of the interview. These individuals are asked retrospective questions about their birth and adoption histories at the time of their first interview. In each succeeding wave these histories are updated.

The population examined here (henceforth the “index persons”) consists of all those for whom the CAHF sample contains records of the childbirth histories of at least one of their parents. The CAHF allows us to compile information on their birth order and the total number of children that their parent(s) report(s).

The index persons with missing information on their birth order or for whom the number of siblings is not ascertained are necessarily excluded from the sample. To ensure that all mothers have completed their fertility so that we correctly identify the total number of siblings, we further limit the sample to those index persons whose mother was older than 44 in the last year she reported. In addition, since our purpose is to look at the effect of birth order, we have restricted attention to families with more than one child. Many studies on birth order have paid attention to attainments of preschool and school children. However, in our data set, we can only look at years of education. Therefore, in this study, our focus is primarily on individuals who have presumably completed their education (*i.e.* index persons at least 25 years of age in 2001).

Siblings are defined based on the childbirth histories of mothers.[4] In addition to the birth order and the number of siblings of the index persons, we have obtained additional demographic information on them and their parents from other PSID files using the unique individual identifiers that are present in the PSID main and supplemental files.

Notably, the PSID suffers from an important attrition bias. More educated people tend to stick with the questionnaire over longer periods of time while less educated ones do not; thus, it *appears* that education is decreasing over cohorts, which is of course untrue according to the U.S. Census. Since a first-born is older than other siblings by definition, this alone could, in theory, produce a spurious positive impact of being first-born on education. We have checked that this problem is of no consequence for our results.[5]

The summary statistics of our sample are presented in Table 1. The detailed description of variables is relegated to the appendix. We found over 8,000 index persons (from more than 3,100 distinct families), older than 24 in 2001, with at least one other sibling, and whose mother has completed her fertility. The main dependent variable – years of completed education – has an average of 12.62 years.[6] About 82% of those

---

[4] The sample shrinks by a third if siblings are defined based on the childbirth histories of fathers, presumably because of the high proportion of single-mother households in the overall sample. In case both parents report, we were able to identify between siblings and half siblings; however, this distinction did not change any of our result. We include a variable expressing whether both parents report in our regressions.

[5] The regressions presented in this article contain age controls that separate cohort from birth order effects. Further, our main results are robust to the exclusion of those age effects.

[6] A more appropriate variable would be education at age 25. However, because this information is unavailable for so many index persons - either because individuals were 25 before 1968, or were present in

selected index persons have at least 12 years of education, *i.e.*, have completed high school.[7] Our measure of earnings – log hourly wage in 2001 – shows an average corresponding to $14.5/hour. However, the information on hourly wages or salary income is not available for most index persons, mainly because the PSID contains this information only for respondents who are head or wife of a household and who are present in the sample in 2001.

The average age in our sample is 39. About half of them are male and 47% are white. The average number of siblings (including the index person) is 4.86. This high number is consistent with the PSID oversampling minorities and low income populations. The first-borns form about 27% of the sample, followed by second-borns (26%), third-borns (17%) etc. 56% of the index persons have all their siblings reporting, and 60% have both of their parents reporting their childbirth history.

We examined two types of family-level variables that can change across parity: family income, and whether the mother is married. Whenever available, we constructed the corresponding information for each of the first 14 years of life of each index person. We then ran our regressions using the average of these variables for three age groups: 1-

---

the sample but did not report education at age of 25 - the number of observations available would drop considerably and we would not be able to run estimations *by* siblings size. All of our other results when running estimations that *control for* siblings size hold when replacing education with education at age 25 for those where such information can be traced.

[7] A negligible fraction of those who declare a certain education level at some point in their life declare less education later. Removing such observations did not alter our results. We therefore choose the latest education level reported as our variable of interest.

6, 7-14, and 1-14. For each age group, the averages are respectively calculated only if the information is present in at least half of the years of the age group (*i.e.*, at least 3 years for age group 1-6; at least 4 years for age group 7-14; and at least 7 years for the age group 1-14 years.) While the inclusion of these variables improves our analysis by controlling for important observed family level-effects that vary across parity, the main limitation is that information on family income cannot be recovered for many index individuals, thus shrinking our sample critically for most sibling sizes.

Lastly, we include two variables describing characteristics of index individuals' parents, namely education (11 years for both mother and father on average) and the age of the parents at birth of index persons (26 for mothers and 29 for fathers on average).

[Insert Table 1]

3. Methods and Results

a. The first-born effect

We first use an OLS estimation with robust standard errors clustered by family unit (identified by the mother), which relaxes the independence assumption between the error terms and requires only that the observations be independent across clusters. Put differently, the procedure allows for correlation of errors within a family.[8] In Table 2, we

[8] We also used random effects procedures, but since they yield almost identical results as those with the family clustered standard errors, those are not reported.

regress completed education on our regressors. We first test the hypothesis that being early in the birth order implies a distinct advantage that is entirely due to the higher probability of coming from a small family.

[Insert Table 2]

Columns (1) and (2) of Table 2 reject that claim but help us understand why it may have been made. In column (1), we omit the number of siblings; therefore, the significant coefficient on firstborn reflects not only the birth order effect but also the probability of coming from a small family. In column (2), the inclusion of the number of siblings leaves the coefficient on firstborn divided by two and less significant.[9]

In column (3), we include age of the mother at childbirth and find a positive and highly significant effect of being first-born on years of education. This effect is confirmed when including the father's characteristics in column (4), where both a father and a mother report. Often, not all siblings report; this is especially the case for large families. To check if we are biasing our results by including such families, in column (5) we restrict our attention to families with complete information on all siblings. The findings are similar there too, further showing that they could not be driven by selective attrition within families by birth order.

---

[9] In fact this is only possible because of the large sample size: when running the estimations on whites or blacks separately for example, the estimate becomes insignificant. The significance found in this specification is therefore somewhat fragile; this explains why previous studies (*e.g.*, Hanushek, 1992) found no effect.

The results presented here are for the impact of being first-born in families of more than one child. This particular procedure takes advantage of the full size of our sample and can be useful when there are not enough observations to run separate estimations for different siblings sizes.[10] We see that it allows us to reveal a significant and robust birth order effect.

However, we still obtain analogous results when looking at individual birth order effects by running estimations by fixed siblings size in Table 3, where all specifications include age of the mother at childbirth. Although large families show higher birth order point estimates, the effect is present in two-sibling families as well. Note that the coefficient on first-born is only weakly significant in column (4) – families of five siblings – likely because of the small sample size.

The reason why the inclusion of the age of the mother at childbirth makes the coefficient on first-born larger in magnitude and more significant is clear: the age of the mother at childbirth is mechanically, positively correlated with the birth order of a child and even more strongly across large families. Conversely, we see that it is positively correlated with a child's education. Then, if having a high birth order carries a negative impact on education, the two effects of birth order and age of the mother at childbirth compete against one another. Therefore the coefficient on firstborn in Table 2 column (2)

---

[10] For example, we found a significant birth order effect when considering separately mothers who first gave birth early and mothers who first gave birth "late" (using various definitions of "early" and "late"), yet we do not have enough observations to split the sample by maternal age and individual siblings size.

reflects an omitted variable bias.[11] The results hold for both males and females, for mothers with or without more than a high school education, for blacks and for whites. We also found that spacing between births, be it that of the first born child with respect to the second born or to the last born child, does not alter the conclusions either.

As noted earlier, a causal interpretation of the age of the mother at childbirth would hinge on the assumption of its exogeneity. Without instrumental variables or a treatment vs. control quasi experiment, it is difficult to draw conclusions.[12] The age of the mother at childbirth, itself positively correlated with birth order, could easily proxy for other unobserved variables such as level of human capital and parental resources.[13]

To address this problem, our fixed-effect estimation (Table 4) removes family characteristics and unobserved family-level heterogeneity. However, it is worth noting that those fixed effects do not solve all endogeneity issues, *e.g.*, it may be the case that

---

[11] The correlation coefficient between firstborn and age of mother at childbirth is −0.43 in the sample used in Table 2 columns (1)-(3).

[12] Also, while it is theoretically possible in the context of very large samples to instrument for siblings size, - using twin births (Black, Devereux, and Salvanes, 2005) or using the fact that parents of two same sex children are slightly more likely to have a third child (Conley, 2004) - one cannot instrument for birth order *per se*.

[13] This is further evidenced by the fact that including a dummy variable indicating whether the mother was married at the time of childbirth instead of (and, obviously, also along with) age of mother at childbirth also results in highly significant birth order coefficients. Results are available upon request.

first born "quality" is affecting subsequent fertility.[14] Also, family fixed effects address family unobservables to the extent that they are constant over time. While we try to incorporate observables such as family income, whether the mother is married, and whether the mother works that vary across birth order to affirm the robustness of our results, we are constrained by the availability of such variables in the data set.

[Insert Table 4]

Unfortunately, the coefficients on age or on age of the mother at childbirth are uninformative in those fixed effects regressions. Deviations from family means for age convey the same information as deviations from family means for age of mother at childbirth. We thus do not provide separate estimations with age and age of mother at childbirth. The age of the mother at first childbirth is certainly relevant, but here, it is differenced out. Still, to assess this issue better, we ran separate regressions, splitting the sample by maternal age.[15] To summarize, the following results are robust to excluding mothers who first gave birth as teens, but we do not have enough observations to meaningfully run the fixed effects regressions by siblings size on that latter group.

---

[14] We thank an anonymous referee for pointing out this caveat.

[15] We thank an anonymous referee for this suggestion.

In Table 4, the fixed effects estimations that only control for age and gender confirm the previous results for the most part.[16] We also provide estimations including the marital status of the mother[17] during the child's first 14 years as an additional covariate: the results do not change by much.[18] Because the information on marital status is retrospective, we do not lose any observation by including this covariate.[19] Clearly, the

---

[16] In column (1) - families of two siblings -, the coefficient on first born becomes significant at the 10% level when including index persons of 23 and 24 years of age, suggesting the non significance when restricting at age 25 stems from a small sample size.

[17] We tried two definitions: number of years married/number of years considered in the age group, and =1 if continuously married over the years considered in the age group, 0 otherwise. To the extent that the results on birth order do not change qualitatively with either of those, we only report the results with the latter definition.

[18] We also ran similar estimations with different age ranges (marital status of mother when the child is 1-6 and 7-14) and obtained similar results.

[19] Unfortunately this is not the case when adding average family income or average employment status of the mother during the child's first years. For example, employment status is based on the PSID individual files which only cover the period 1979-1993; for family income we tap into the PSID family files but there is no record prior to 1968. In addition, recall we ran estimations provided that the information on each of those variables is at least present during half of the age range considered, which caused to lose even more observations. The results yield significant birth order effects for only some siblings sizes when using family income, but employment status of the mother was found to be of no practical use; this is due to the resulting very small sample sizes.

suggestion that first borns are most likely to live their critical development years in a stable household, as opposed to later borns who may experience the divorce of their parents, cannot entirely explain the first born advantage. At the same time, the persistence of birth order effects naturally poses the problem of their origin.

The literature is not able to distinguish between different theories on the topic of birth order. For example, schooling factors and circumstances play a large role in educational outcomes and we cannot ascertain whether they are related to birth order.[20] It could be that first born are sent to better schools, which would in turn provide higher incentives to stay in school longer. Overall, our analysis is necessarily limited in the sense that it cannot discriminate between many competing hypotheses on why birth order appears to be important.[21]

Nevertheless, we checked whether what appears as a first born advantage predominantly comes from financial constraints, *i.e.*, parents sending their first born to college and running out of money for the following siblings. Conley offers the following argument: "in terms of parental investment, the cup starts to run dry as we go down the line (…) Parental resources, it appears, are allotted on a first come, first-served basis."[22] Yet if it turns out that first born perform better beforehand, then a theory based on budget constraints cannot fully account for our results. In Tables 5-7, we estimate the probability

---

[20] We thank an anonymous referee for this point.

[21] Black, Devereux, and Salvanes (2005) review several of those theories at length and we refer the reader to their survey.

[22] Conley (2004), p. 69.

of completing high school, following the same methodology as in Table 2-4. We find that first born have a higher probability of completing high school than later born siblings.

Specifically, Table 5 shows again that the first-born effect is not an artifact of family-size; that it increases in magnitude and significance when including age of mother at birth; that it is robust to including characteristics of father; and to restricting the sample to families in which all siblings report. Table 6 shows that these results also hold when regressions are estimated separately for each family size (except for families of 5, presumably because of the small sample size).

Table 7 is the counterpart of Table 4. Fixed effects regressions controlling for age and gender also support the results found in Tables 5 and 6. When including marital status of the mother as an additional control, the results were confirmed.[23]

Also, we estimated education at age 18 conditional on high school completion (at 18 or later) to see if later borns are more likely to repeat grades, which would support a theory of birth order effects based on cognitive development differences. We did not find any evidence for this. However, the small sample size resulting from selecting index persons with available information at age 18 warrant some caution.

Finally, we did not find strong evidence that conditional on completing high school, first born are any more likely to receive post secondary education, nor that conditional on post-secondary education, there is any advantage to being first born.[24] In summary,

---

[23] We ran into the same small size problems when adding family income and employment status of the mother for different age ranges of each child.

[24] Significant OLS results but insignificant FE results. Results are available from the authors.

financial constraints do not seem to play a major role, and some factors early in life must contribute to the first born premium puzzle.

b.The "last borns effect" in large families

We now test the hypothesis that within large (more than 5 siblings) families, the middle born do worst and that the last-born do better than everyone else. There is some support in the literature that there is a "crunch in the middle": "For almost as long as sociologists have been studying who gets ahead, they have found that kids from large families do more poorly than those from small ones. There is, however, one exception to this: last born children from very large families seem to fare quite well (…) the middle kids do worst."[25] This nonlinear pattern was also advanced by Hanushek (1992)[26] and in the context of time allocation, by Lindert (1977).

There are many ways to replicate these findings. Short of enough observations for each family size when family size is very large, the variables of interest chosen in Table 8 are a dummy indicating whether the child is first born and a dummy indicating whether a child is last born. Note that dummies indicating whether a child is among the first two born and among the last two born would yield similar results.

[Insert Table 8]

[25] Conley (2004): p. 66 and p. 69.

[26] It is actually nicely illustrated on pp. 101-102 (figures 1 and 2) in Hanushek (1992).

When omitting the age of mother at childbirth in column (1), the first born coefficient is insignificant as earlier, but the coefficient on last born is positive and significant. Notice that this does not happen when we run the same regression on smaller families. Under the specification of column (1), the first born does not perform better against an average of the middle born (who supposedly perform worst) and the last born (who supposedly perform well). However, in column (2), once the age of the mother at childbirth is factored in, we find that being last-born confers no advantage but that being first born does.

The fixed effects estimation (column 3) confirms the absence of any upward trend from middle born to last-born.[27] The interpretation of those results is similar to the ones presented earlier and the same qualifications apply.[28]

To summarize, the apparent advantage of being last-born in large families is entirely attributable to the rising age of the mother at childbirth, again with the caveat that the age

[27] Adding whether the mother was continuously married during the children's first years does not change the results qualitatively and similar small size problems arose when adding family income and employment status of the mother for different age ranges of each child. Results are available from the authors.

[28] Regressions restricting the sample to both parents present at childbirth (with or, in the case of fixed effects, without father's characteristics) yield similar results. However, we cannot do the same with large families with complete information on all siblings: the sample becomes too small. Only 25% of families with more than 5 siblings have all their siblings report, which leaves only 633 observations from 100 families. Even the number of siblings is insignificant in such a small sample.

of the mother at childbirth actually picks other elements that are also correlated with it and may affect the outcome of interest.

4. Birth Order and Earnings

Since education is a key in determining earnings, if we find a significant effect of birth order on education, we should likewise find a similar effect on earnings. Our sample is more limited because we only have information on earnings for heads or wives who declare working (about 36% of our initial sample)[29]. Nonetheless, Table 9 shows that the results on hourly earnings display the same patterns as for education, namely a non-significant effect of birth order when age of the mother at birth is omitted, and a significant effect when it is included. Note that with only a few hundred observations on hourly earnings for each siblings size, we cannot run meaningful OLS or fixed effects estimations by siblings size.[30]

[29] Out of 8,318 respondents who are older than 24 in 2001, whose mother completed fertility, and who have at least one sibling: 4,397 are not present in the 2001 PSID Hours of Work and Wages File, 471 are not a head or a wife of the household, 158 did not work any positive hours, and for 264 of them the hourly wage is missing. This leaves us with 3,028 observations with log hourly wage, or about 36 % of the initial 8,318 respondents.

[30] Fixed effects estimations for the entire sample confirm the OLS results though.

5. Conclusion

We have shown how the omission of the age of the mother at childbirth effect results in an underestimation of the impact of being first-born and an overestimation of the impact of being last-born. At this point however, the age of the mother at childbirth should be interpreted broadly as a proxy for a set of maternal inputs. Most importantly, fixed-effects estimations confirmed the presence of a positive first-born effect and the absence of either specific middle born or last-born effects among large families.

From a policy perspective, any effort to treat individuals differently on the basis of birth parity seems unfeasible. Because of the endogeneity of the mother's age at first birth, and without any source of exogenous variations, we cannot ascertain with enough confidence at this point the benefits of a policy that would offset the determinants of early motherhood in order to promote children's education.

Using twins as an instrument, Black, Devereux and Salvanes (2005) advance that "the family size itself has little effect on the quality of each child but more likely impacts the marginal children through the effect of birth order"[31] in their Norwegian sample. Therefore, while we tentatively agree with the idea that smaller family sizes may be responsible for a rise in scholastic performances over cohorts in the U.S., we would like to emphasize that this effect is compounded by a corresponding increase in the proportion of first-born children.

[31] Black, Devereux and Salvanes (2005), p. 2.

References

Amato, P. R. and A. Booth. 1997. *A Generation at Risk: Growing Up in an era of Family Upheaval.* Cambridge, Mass: Harvard University Press.

Birdsall, N. 1979. Siblings and Schooling in Urban Columbia. PhD Dissertation, Yale University, Department of Economics.

Behrman, J. R. 1986. Birth Order, Nutrition and Health; IntraHousehold Allocation in Rural India. Mimeo. University of Pennsylvania.

Behrman, J. R. and P. Taubman. 1986. Birth Order, Schooling, and Earnings. *Journal of Labor Economics* 4 : S121-45.

Black, S. E., P. G. Devereux, K. G. Salvanes. 2005. The More the Merrier? The Effect of Family Composition on Children's Education. *Quarterly Journal of Economics* 120: 669-700.

Bronars, S. G., and J. Grogger. 1994. The Economic Consequences of Unwed Motherhood: Using Twin Births as a Natural Experiment. *American Economic Review* 84: 1141-1156.

Conley, D. 2004. *The Pecking Order*. Pantheon Books, New York.

Geronimus A. T., S. Korenman, M. M. Hillemeier. 1994. Does young maternal age affect child development? Evidence from cousin comparisons in the United States. *Population and Development Review* 20 : 585-609.

Hanushek, E. A. 1992. The Trade-off between Child Quantity and Quality. *Journal of Political Economy* 100 : 84-117.

Hofferth S. L., and L. Reid. 2002. Early Childbearing and Children's Achievement and Behavior over Time. *Perspectives on Sexual and Reproductive Health* 34 : 41-49.

Kehrer K. C. 1979. The Gary Income Maintenance Experiment. *Journal of Human Resources* 14 : 431-33.

Kelly, T. F. and L. Singer. 1971. The Gary Income Maintenance Experiment: Plans and Progress. *American Economic Review* 61 : 30-38.

Kessler, D. 1991. Birth Order, Family Size, and Achievement: Family Structure and Wage Determination. *Journal of Labor Economics* 9 : 413-26.

Lindert, P. H. 1977. Sibling Position and Achievement. *Journal of Human Resources* 12 : 198-219.

Lopez-Turley, R. N. 2003. Are Children of Young Mothers Disadvantaged Because of Their Mother's Age or Family Background? *Child Development* 74 : 465-474.

Appendix: Description of the variables

Variable	Description
Years of completed education[1]	Years of education reported in the most recent year
Completed high school[1]	= 1 if years of completed education greater than or equal to 12
Hourly earnings[5]	= hourly earnings in 2001 if the index person is a head or a wife of a household, and missing otherwise
Age[1]	= the age of index person, based on the year of birth
Male[1]	=1 if the gender of index person is male
White[2]	=1 if the race of mother of index person is white, or if the race of mother of index person is missing but the race of father is white
Number of siblings[3]	The total number of childbirths reported by the mother of index person if mother older than 44 in the last year in which she reported; otherwise, set to missing
Birth order[3] (First-born, etc.)	The birth order of index person
Family income[2]	=Total income of the household. The average family income is calculated only if it is available for at least 50% of years in the relevant time period (i.e. 3+ years for ages 1-6, 4+ years for ages 7-14, and 7+ years for ages 1-14)
Mother married[4]	=1 the mother is continuously married during the relevant period, and 0 if not. The average marital status of mother is calculated only if it is available for at least 50% of years in the relevant time period (i.e. 3+ years for ages 1-6, 4+ years for ages 7-14, and 7+ years for ages 1-14)
Mother working[1]	=1 in each year in which mother is working. The average employment status of mother is calculated only if it is available for at least 50% of years in the relevant time period (i.e. 3+ years for ages 1-6, 4+ years for ages 7-14, and 7+ years for ages 1-14)
Information on all siblings	=1 if all siblings present in the sample and report
Both parents report the childbirth	=1 if both the mother and father of index person report the birth of index person

<u>Data sources</u>:
[1] Individual PSID file
[2] Family PSID file
[3] Childbirth and Adoption History File
[4] Marriage History File
[5] Hours of Work and Wages File

Table 1

DESCRIPTIVE STATISTICS

Variable	Obs.	Mean	Std.Dev.	Min	Max
Index individual					
Years of completed education	8,147	12.62	2.13	1	17
% completed high school	8,147	0.82	-	0	1
Log hourly earnings in 2001	3,028	2.67	0.71	-1.2	5.99
Age (in 2001)	8,318	38.57	8.76	25	89
% male	8,318	0.50	-	0	1
% white	8,318	0.47	-	0	1
Number of siblings	8,318	4.86	2.72	2	16
% first-born	8,318	0.27	-	0	1
% second-born	8,318	0.26	-	0	1
% third-born	8,318	0.17	-	0	1
% fourth-born	8,318	0.11	-	0	1
% fifth-born	8,318	0.07	-	0	1
Information on all siblings	8,318	0.56	-	0	1
Both parents report the childbirth	8,318	0.60	-	0	1
Family income:					
Age 1-6	2,695	13,639	9,283	508	97,660
Age 7-14	4,576	19,993	17,360	1,173	255,393
Age 1-14	3,474	18,850	14,026	1,092	178,480
Mother continuously married:					
Age 1-6	8,318	0.25	-	0	1
Age 7-14	8,318	0.36	-	0	1
Age 1-14	8,318	0.21	-	0	1
Mother working:					
Age 1-6	378	0.45	-	0	1
Age 7-14	1,823	0.56	-	0	1
Age 1-14	1,176	0.55	-	0	1
<u>Mother</u>					
Age at birth	8,292	26.18	5.88	15	48
Years of completed education	8,102	11.03	3.02	1	20
<u>Father</u>					
Age at birth	5,000	29.32	6.54	17	60
Years of completed education	4,892	11.13	3.67	1	17

Includes index persons who have at least one sibling, who are 25 years of older in 2001, and whose mother is at least 44 years old in the last year she reported. The number of distinct families is 3,112.

Table 2

OLS REGRESSION WITH DEPENDENT VARIABLE: EDUCATION

	(1)	(2)	(3)	(4)	(5)
d[first-born]	0.226	0.111	0.383	0.41	0.361
	(0.048)***	(0.049)**	(0.051)***	(0.064)***	(0.06)***
Total number of siblings		-0.102	-0.12	-0.107	-0.124
		(0.015)***	(0.015)***	(0.02)***	(0.029)***
Age of the mother at childbirth			0.058	0.06	0.065
			(0.005)***	(0.011)***	(0.008)***
d[Male]	-0.205	-0.21	-0.226	-0.186	-0.158
	(0.047)***	(0.046)***	(0.046)***	(0.058)***	(0.057)***
Age	0.12	0.158	0.162	0.142	0.084
	(0.027)***	(0.027)***	(0.027)***	(0.039)***	(0.04)**
Age^2	-0.001	-0.001	-0.001	-0.001	-3×10^{-4}
	(3×10^{-4})**	(3×10^{-4})***	(3×10^{-4})***	(5×10^{-4})**	(5×10^{-4})
d[White]	0.153	0.05	-0.05	-0.212	-0.042
	(0.064)**	(0.064)	(0.063)	(0.077)***	(0.083)
Mother's education	0.213	0.197	0.2	0.114	0.236
	(0.012)***	(0.012)***	(0.012)***	(0.017)***	(0.018)***
d[all siblings report]	0.26	0.095	0.18	0.142	
	(0.068)**	(0.07)	(0.069)***	(0.089)	
d[both parents report]	0.526	0.554	0.52		0.532
	(0.063)***	(0.062)***	(0.061)***		(0.081)***
Father's education				0.124	
				(0.013)***	
Age of father at childbirth				0.006	
				(0.009)	
Constant	6.69	6.681	5.038	5.123	5.873
	(0.53)***	(0.532)***	(0.542)***	(0.778)***	(0.775)***
R^2	0.158	0.169	0.189	0.2	0.2
# observations	7,928	7,928	7,928	4,766	4,541
# family clusters	3,112	3,112	3,112	1,869	1,732

(1)-(5): all mothers have completed their fertility (age>44), all respondents assumed to have completed their education (age>24) and have at least one other sibling.
(4): d[both parents report]=1 and (5): d[complete info on all siblings]=1
*: 10% significance; **: 5% significance; ***: 1% significance. Robust standard errors clustered by family.

Table 3

OLS REGRESSION WITH DEPENDENT VARIABLE: EDUCATION

	(1)	(2)	(3)	(4)	(5)
d[first-born]	0.275	0.698	0.715	0.534	0.915
	(0.104)***	(0.136)***	(0.185)***	(0.316)*	(0.236)***
d[secondborn]		0.434	0.463	0.334	0.578
		(0.117)***	(0.16)***	(0.282)	(0.195)***
d[thirdborn]			0.171	0.212	0.368
			(0.14)	(0.247)	(0.175)**
d[fourthborn]				-0.012	0.466
				(0.214)	(0.152)***
d[fifthborn]					0.221
					(0.133)*
Age of the mother at childbirth	0.048	0.079	0.067	0.068	0.07
	(0.012)***	(0.013)***	(0.013)***	(0.023)***	(0.013)***
d[Male]	-0.287	-0.115	-0.035	-0.34	-0.33
	(0.102)***	(0.091)	(0.107)	(0.131)***	(0.087)***
Age	0.217	0.19	0.156	0.075	0.167
	(0.056)***	(0.057)***	(0.049)***	(0.044)*	(0.067)**
Age^2	-0.002	-0.002	-0.002	-3×10^{-4}	-0.002
	(7×10^{-4})***	(7×10^{-4})**	(6×10^{-4})**	(5×10^{-4})	(8×10^{-4})**
d[White]	0.119	-0.124	4×10^{-5}	-0.03	-0.114
	(0.12)	(0.13)	(0.13)	(0.181)	(0.137)
Mother's education	0.257	0.25	0.179	0.124	0.182
	(0.025)***	(0.026)***	(0.023)***	(0.035)***	(0.022)***
d[all siblings report]	0.147	0.212	-0.066	0.351	0.21
	(0.154)	(0.136)	(0.128)	(0.178)**	(0.149)
d[both parents report]	0.241	0.612	0.755	0.542	0.371
	(0.118)**	(0.118)**	(0.126)***	(0.189)***	(0.127)***
Constant	3.339	2.694	4.258	6.607	4.12
	(1.148)***	(1.184)**	(1.011)***	(1.293)***	(1.415)***
R^2	0.185	0.21	0.19	0.114	0.127
# observations	1,398	1,705	1,444	959	2,422
# family clusters	913	811	542	308	538

(1)-(5): all mothers have completed their fertility (age>44), all respondents assumed to have completed their education (age>24).
(1)-(5): Families of 2, 3, 4, 5, and 6 and above siblings respectively.
*: 10% significance; **: 5% significance; ***: 1% significance. Robust standard errors clustered by family.

Table 4

FIXED EFFECTS LINEAR REGRESSION WITH DEPENDENT VARIABLE: EDUCATION (ALL COEFFICIENTS: DEVIATION FROM FAMILY MEANS)

WITHOUT INCLUDING MARITAL STATUS OF MOTHER

	(1)	(2)	(3)	(4)	(5)
d[first-born]	0.216	0.437	0.683	0.594	0.477
	(0.181)	(0.206)**	(0.265)***	(0.379)	(0.224)**
d[secondborn]		0.282	0.412	0.421	0.199
		(0.141)**	(0.214)*	(0.323)	(0.195)
d[thirdborn]			0.138	0.361	0.055
			(0.159)	(0.261)	(0.172)
d[fourthborn]				-0.031	0.181
				(0.207)	(0.149)
d[fifthborn]					-0.012
					(0.125)
Observations	1,422	1,743	1,477	922	2,487
Groups	934	835	562	388	573

INCLUDING MARITAL STATUS OF MOTHER

	(1)	(2)	(3)	(4)	(5)
d[first-born]	0.224	0.436	0.696	0.585	0.492
	(0.181)	(0.206)**	(0.267)***	(0.379)	(0.226)**
d[secondborn]		0.28	0.423	0.413	0.212
		(0.141)**	(0.216)**	(0.323)	(0.197)
d[thirdborn]			0.146	0.363	0.063
			(0.159)	(0.261)	(0.173)
d[fourthborn]				-0.025	0.188
				(0.207)	(0.149)
d[fifthborn]					-0.016
					(0.125)
{Mother continuously married,	-0.196	-0.167	0.112	0.381	-0.105
Age 1-14}	(0.315)	(0.221)	(0.221)	(0.274)	(0.178)
Observations	1,422	1,743	1,477	999	2,487
Groups	934	835	562	328	573

(1)-(5): all mothers have completed their fertility (age>44), all respondents assumed to have completed their education (age>24). (1)-(5): Families of 2, 3, 4, 5, and 6 and above siblings respectively. *: 10% significance; **: 5% significance; ***: 1% significance. Robust standard errors clustered by family. No estimation run with age because including both age and age of mother at birth is redundant in a family fixed effects regression. The regressions also include age of mother at birth and an indicator for males.

Table 5
OLS REGRESSION WITH DEPENDENT VARIABLE: HIGH SCHOOL COMPLETION

	(1)	(2)	(3)	(4)	(5)
d[first-born]	0.037	0.021	0.051	0.045	0.047
	(0.009)***	(0.009)**	(0.01)***	(0.011)***	(0.011)***
Total number of siblings		-0.014	-0.016	-0.011	-0.011
		(0.003)***	(0.003)***	(0.004)**	(0.005)**
Age of the mother at childbirth			0.006	0.007	0.005
			(0.001)***	(0.002)***	(0.001)***
d[Male]	-0.047	-0.048	-0.05	-0.047	-0.037
	(0.009)***	(0.008)***	(0.009)***	(0.01)***	(0.01)***
Age	0.024	0.029	0.029	0.032	0.023
	(0.004)***	(0.004)***	(0.004)***	(0.066)***	(0.007)***
Age^2	-2×10^{-4}	-3×10^{-4}	-3×10^{-4}	-3×10^{-4}	-2×10^{-4}
	(5×10^{-5})***	(5×10^{-5})***	(5×10^{-5})***	(8×10^{-5})***	(9×10^{-5})***
d[White]	7×10^{-4}	-0.014	-0.025	-0.039	-0.017
	(0.012)	(0.012)	(0.012)**	(0.014)***	(0.016)
Mother's education	0.028	0.026	0.026	0.015	0.026
	(0.002)***	(0.002)***	(0.002)***	(0.003)***	(0.003)***
d[complete info on all siblings]	0.035	0.012	0.022	0.029	
	(0.013)***	(0.013)	(0.013)	(0.015)*	
d[both parents report]	0.09	0.094	0.09		0.097
	(0.012)***	(0.012)***	(0.012)***		(0.016)***
Father's education				0.011	
				(0.002)***	
Age of father at childbirth				-0.001	
				(0.001)	
Constant	-0.102	-0.103	-0.287	-0.242	-0.128
	(0.089)***	(0.09)***	(0.093)***	(0.135)*	(0.141)
R^2	0.088	0.094	0.102	0.089	0.085
# observations	7,928	7,928	7,928	4,766	4,541
# family clusters	3,112	3,112	3,112	1,869	1,732

(1)-(5): all mothers have completed their fertility (age>44), all respondents assumed to have completed their education (age>24) and have at least one other sibling.
(4): d[both parents report]=1 and (5): d[complete info on all siblings]=1
*: 10% significance; **: 5% significance; ***: 1% significance. Robust standard errors clustered by family.

Table 6

OLS REGRESSION WITH DEPENDENT VARIABLE: HIGH SCHOOL COMPLETION

	(1)	(2)	(3)	(4)	(5)
d[first-born]	0.038	0.084	0.09	0.031	0.122
	(0.018)**	(0.025)***	(0.036)**	(0.053)	(0.049)**
d[secondborn]		0.071	0.08	2×10^{-4}	0.022
		(0.022)***	(0.031)***	(0.051)	(0.044)
d[thirdborn]			0.031	-0.012	0.035
			(0.028)	(0.041)	(0.037)
d[fourthborn]				-0.061	0.041
				(0.037)	(0.033)
d[fifthborn]					0.007
					(0.028)
age of the mother at childbirth	0.003	0.004	0.009	0.008	0.011
	(0.002)*	(0.002)*	(0.003)***	(0.004)**	(0.003)***
d[Male]	-0.03	-0.041	0.007	-0.094	-0.082
	(0.016)*	(0.016)**	(0.02)	(0.026)***	(0.018)***
Age	0.028	0.028	0.051	0.012	0.037
	(0.007)***	(0.009)***	(0.01)***	(0.008)	(0.011)***
Age^2	-3×10^{-4}	-3×10^{-4}	-5×10^{-4}	-8×10^{-5}	-4×10^{-4}
	(9×10^{-5})***	(10^{-5})***	(10^{-5})***	(9×10^{-5})	(10^{-4})***
d[White]	0.012	-0.124	-0.049	-0.033	-0.022
	(0.021)	(0.023)	(0.026)*	(0.034)	(0.028)
Mother's education	0.024	0.025	0.027	0.018	0.027
	(0.004)***	(0.004)***	(0.005)***	(0.005)***	(0.004)***
d[all siblings report]	-0.011	0.03	-0.034	0.059	0.037
	(0.025)	(0.024)	(0.024)	(0.034)*	(0.03)
d[both parents report]	0.045	0.076	0.125	0.088	0.098
	(0.019)**	(0.022)***	(0.027)***	(0.036)**	(0.026)***
Constant	-0.115	-0.254	-0.927	0.026	-0.644
	(0.16)	(0.2)	(0.2)***	(0.234)	(0.261)**
R^2	0.076	0.084	0.138	0.078	0.088
# observations	1,398	1,705	1,444	959	2,422
# family clusters	913	811	542	308	538

(1)-(5): all mothers have completed their fertility (age>44), all respondents assumed to have completed their education (age>24).
(1)-(5): Families of 2, 3, 4, 5, and 6 and above siblings respectively.
*: 10% significance; **: 5% significance; ***: 1% significance. Robust standard errors clustered by family.

Table 7

FIXED EFFECTS LINEAR REGRESSION WITH DEPENDENT VARIABLE: HIGH SCHOOL COMPLETION (ALL COEFFICIENTS: DEVIATION FROM FAMILY MEANS)

WITHOUT INCLUDING MARITAL STATUS OF MOTHER

	(1)	(2)	(3)	(4)	(5)
d[first-born]	0.033	0.089	0.132	0.076	0.124
	(0.029)	(0.037)**	(0.051)***	(0.073)	(0.051)**
d[secondborn]		0.064	0.107	0.039	0.029
		(0.026)***	(0.041)***	(0.062)	(0.045)
d[thirdborn]			0.048	0.061	0.047
			(0.03)	(0.05)	(0.039)
d[fourthborn]				-0.054	0.049
				(0.04)	(0.034)
d[fifthborn]					0.004
					(0.029)
Observations	1,422	1,743	1,477	999	2,487
Groups	934	835	562	328	573

INCLUDING MARITAL STATUS OF MOTHER

	(1)	(2)	(3)	(4)	(5)
d[first-born]	0.033	0.089	0.134	0.075	0.129
	(0.029)	(0.037)**	(0.051)***	(0.073)	(0.052)***
d[secondborn]		0.063	0.109	0.039	0.033
		(0.026)***	(0.041)***	(0.062)	(0.045)
d[thirdborn]			0.049	0.061	0.049
			(0.03)	(0.051)	(0.039)
d[fourthborn]				-0.053	0.051
				(0.04)	(0.034)
d[fifthborn]					0.004
					(0.028)
{Mother continuously married,	-0.001	-0.058	0.019	0.025	-0.035
Age 1-14}	(0.051)	(0.04)	(0.042)	(0.053)	(0.041)
Observations	1,422	1,743	1,477	999	2,487
Groups	934	835	562	328	573

(1)-(5): all mothers have completed their fertility (age>44), all respondents assumed to have completed their education (age>24). (1)-(5): Families of 2, 3, 4, 5, and 6 and above siblings respectively. *: 10% significance; **: 5% significance; ***: 1% significance. Robust standard errors clustered by family. No estimation run with age because including both age and age of mother at birth is redundant in a family fixed effects regression. The regressions also include age of mother at birth and an indicator for males.

Table 8
REGRESSION WITH DEPENDENT VARIABLE: COMPLETED EDUCATION
(LARGE FAMILIES)

	OLS		FAMILY FIXED EFFECTS
	(1)	(2)	(3)
d[firstborn]	0.114	0.463	0.305
	(0.161)	(0.166)***	(0.141)**
d[lastborn]	0.327	-0.06	0.098
	(0.106)***	(0.123)	(0.137)
Total number of siblings	-0.059	-0.089	
	(0.031)*	(0.033)***	
Age of the mother at childbirth		0.056	
		(0.011)***	
d[Male]	-0.322	-0.331	-0.278
	(0.089)***	(0.087)***	(0.093)***
Age	0.167	0.172	-0.001
	(0.066)***	(0.068)**	(0.01)
Age^2	-0.002	-0.001	
	(2×10^{-4})**	(8×10^{-4})**	
Mother's education	0.175	0.181	
	(0.022)***	(0.021)	
d[all siblings report]	0.115	0.175	
	(0.151)	(0.151)	
d[both parents report]	0.451	0.431	
	(0.129)***	(0.126)***	
Constant	6.757	5.209	12.08
	(1.314)***	(1.327)***	(0.286)***
R^2	0.11	0.13	0.01
# observations	2,422	2,422	2,487
# family clusters	538	538	573

(1)-(3): all mothers have completed their fertility (age>44), all respondents assumed to have completed their education (age>24), all respondent from families > 5 siblings. Robust standard errors clustered by family.
*: 10% significance; **: 5% significance; ***: 1% significance.

Table 9
OLS REGRESSION WITH DEPENDENT VARIABLE: LOG HOURLY WAGE IN 2001

	(1)	(2)	(3)	(4)	(5)
d[first-born]	0.047	0.027	0.063	0.058	0.058
	(0.026)*	(0.027)	(0.029)**	(0.032)*	(0.033)*
Total number of siblings		-0.022	-0.025	-0.026	-0.041
		(0.007)***	(0.007)***	(0.008)***	(0.014)***
Age of the mother at childbirth			0.008	0.008	0.010
			(0.002)***	(0.005)*	(0.030)***
d[Male]	0.301	0.300	0.299	0.347	0.291
	(0.025)***	(0.024)***	(0.024)***	(0.029)***	(0.030)***
Age	0.058	0.067	0.065	0.067	0.079
	(0.013)***	(0.013)***	(0.013)***	(0.017)***	(0.021)***
Age^2	-0.001	-0.001	-0.001	-0.001	-0.001
	(2×10^{-4})***	(2×10^{-4})***	(2×10^{-4})***	(2×10^{-4})***	(3×10^{-4})***
Mother's education	0.048	0.044	0.043	0.020	0.045
	(0.006)***	(0.006)***	(0.006)***	(0.008)**	(0.009)***
d[complete info on all siblings]	0.103	0.063	0.071	0.094	
	(0.032)***	(0.035)*	(0.036)**	(0.044)**	
d[both parents report]	0.092	0.098	0.097	0.171	
	(0.030)***	(0.030)***	(0.030)***	(0.037)***	
Father's education				0.045	
				(0.006)***	
Age of father at childbirth				0.002	
				(0.004)	
Constant	0.266	0.277	0.130	-0.228	-0.142
	(0.262)	(0.261)	0.264	(0.343)	(0.401)
R^2	0.159	0.163	0.165	0.220	0.174
# observations	3,000	3,000	3,000	2,059	1,962
# family clusters	1,575	1,575	1,575	1,075	1,015

(1)-(5): all mothers have completed their fertility (age>44), all respondents assumed to have completed their education (age>24) and have at least one other sibling.
(4): d[both parents report]=1 and (5): d[complete info on all siblings]=1
*: 10% significance; **: 5% significance; ***: 1% significance. Robust standard errors clustered by family.