Maryland Population Research Center

WORKING PAPER

Education, Birth Order, and Family Size

PWP-MPRC-2013-016
June 2013

Authors:

Jesper Bagger
University of London

Javier A. Birchenall
University of California

Hani Mansour
University of Colorado
Denver

Sergio S. Urzúa
University of Maryland

NBER WORKING PAPER SERIES

EDUCATION, BIRTH ORDER, AND FAMILY SIZE

Jesper Bagger
Javier A. Birchenall
Hani Mansour
Sergio Urzúa

Working Paper 19111
http://www.nber.org/papers/w19111

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2013

Earlier versions of this paper were presented at Royal Holloway, Catholic University of Chile and University of Maryland. We benefited from helpful comments by participants at the seminars and also from valuable comments by Sandra Black, Paul Devereux, Jessica Goldberg, and Kjell Salvanes. Jesper Bagger gratefully acknowledges financial support from The Cycles, Adjustment, and Policy research unit, CAP, Department of Economics and Business, Aarhus University. Sergio Urzua appreciates the support of Centro de Microdatos at the University of Chile through the Millennium Science Initiative sponsored by the Chilean Ministry of Economics, Development and Tourism, Project NS100041. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Education, Birth Order, and Family Size
Jesper Bagger, Javier A. Birchenall, Hani Mansour, and Sergio Urzúa
NBER Working Paper No. 19111
June 2013
JEL No. C23,C26,D50,E20,E24,J12,J13

ABSTRACT

We introduce a general framework to analyze the trade-off between education and family size. Our framework incorporates parental preferences for birth order and delivers theoretically consistent birth order and family size effects on children's educational attainment. We develop an empirical strategy to identify these effects. We show that the coefficient on family size in a regression of educational attainment on birth order and family size does not identify the family size effect as defined within our framework, even when the endogeneity of both birth order and family size are properly accounted for. Using Danish administrative data we test the theoretical implications of the model. The data does not reject our theory. We find significant birth order and family size effects in individuals' years of education thereby confirming the presence of a quantity-quality trade off.

Jesper Bagger
Royal Holloway
University of London
Egham Hill
Surrey TW20 0EX
United Kingdom
jesper.bagger@rhul.ac.uk

Javier A. Birchenall
Department of Economics
University of California
2127 North Hall
Santa Barbara, CA 93106
jabirche@econ.ucsb.edu

Hani Mansour
Department of Economics
University of Colorado Denver
Campus Box 181, P.O. Box 173364
Denver, CO 80217
hani.mansour@ucdenver.edu

Sergio Urzúa
Department of Economics
University of Maryland
3105 Tydings Hall #3115C
College Park, MD 20742
and NBER
urzua@econ.umd.edu

1 Introduction

Understanding the effect of changes in family size on children's education has been an ongoing concern in economics. Existing theoretical discussions have followed Becker and Lewis's [6] seminal paper. Under the assumption that parents treat all children equally, these models predict that an increase in family size reduces children's education.[1] A separate strand of the literature, proceeding mostly on empirical grounds, has shown that the equal treatment assumption overly simplifies how education is distributed within the household. For example, Black et al. [12] investigate the relationship between family size and education, while taking into account differences in the order of birth of a child, their *birth order*. There are two important results coming from this empirical literature. The first is that birth order negatively influences a child's education. The second is that after controlling for birth order, there is no evidence supporting a negative effect of family size on a child's education. In other words, the existing empirical evidence do not support the hypothesis of a quantity-quality trade-off.[2]

This paper formulates a model of fertility choice that incorporates parental preferences for specific birth orders. The theory delivers theoretically consistent birth order and family size effects on children's educational attainment. Specifically, the theory predicts that an exogenous increase in family size reduces average human capital within the household, even when birth order effects are present. The theory also highlights the limitations of existing empirical strategies in identifying the effect of family size in the presence of birth order predispositions; primarily due to not fully accounting for the fact that family size and birth order are jointly determined. Guided by the theory, we introduce and implement a simple empirical strategy that identifies the effect of family size on children's human capital separately from birth order effects. We test the model's predictions using a population-wide comprehensive administrative panel dataset from Denmark.

We find that birth order has a strong negative effect on a child's education, consistent with existing empirical studies. Controlling for family fixed effects and with a linear birth order effect,

[1] The literature on family size and education is too large to provide a detailed account here. From a theoretical perspective, Barro and Becker [4], Becker, Murphy and Tamura [7], Becker and Tomes [8], Doepke [14], and Galor and Weil [15] study family size and education in models of economic growth. Some empirical studies include Ahn et al. [1], Angrist et al. [3], Behrman et al. [11], Glick et. al [16], Hanushek [18], Li et al. [23], Millet and Wang [25], Parish and Willis [28], Qian [29], and Schultz [32]. These studies typically abstract from birth order effects.

[2] Recently Mogstad and Wiswall [26] have argued that this result may not be robust to nonlinear family size effects, an issue we shall return to in this paper as well.

an additional birth order reduces years of schooling by little less than one fifth of a year. Results are similar in magnitude when we allow for nonlinear birth order effects. However, contrary to the previous literature, we find that family size, once carefully analyzed in the light of the theory, has a strong negative effect on the distribution of education in the household; particularly on the household's average. Quantitatively, in the case of linear family size effects, an additional child reduces the average number of years of schooling in the household by about one tenth of a year. In a more flexible specification we find evidence of nonlinear family size effects. Overall, we provide evidence supporting the existence of a trade-off between quality and quantity of children.

Our findings relating to the effect of family size appears at odds with previous findings in e.g. Black et al. [12]. We use our model to shed light on the reasons behind the different results. Consider two households with different family sizes but equal expenditures on children. Since expenditures are equal, per child education should on average be lower in the larger household. This effect of family size is a consequence of limited parental resources, or the quantity-quality trade-off. But differences in family size necessarily translates into differences in birth orders within the two families, changing the distribution of education within the household. Importantly, the average birth order is higher in the larger family. A comparison of the educational outcomes of children with the same birth order in families of different sizes confounds the quantity-quality trade-off with differences in the average birth order effect in the two families. Specifically, if parents have a predisposition to favor low birth orders, we show that the coefficient on family size in a regression of educational attainment on birth order and family size provides a biased-towards-zero estimate of the family size effect and thus leads to a false rejection of the quantity-quality trade-off. This is true even when the endogeneity of both birth order and family size are properly accounted for as in e.g. Black et al. [12]. The bias is not due to endogeneity; it arises because one *cannot* manipulate family size, holding the within-family distribution of birth orders constant.

Our empirical strategy overcome this complication in two simple steps. In the first step, we estimate birth order effects controlling for family-fixed effects. In the second step, we "net out" the estimated birth order effects from the child's education and estimate the effect of an exogenous change in family size using an instrumental variable approach. One can also think of the two steps as identifying birth order effects from within-family variation in educational attainment in the first step, while the second step utilizes between-family variation in educational attainment to estimate the family size effect, or the quantity-quality trade-off.

The rest of the paper is organized as follows. Section 2 introduces our model of fertility in which parents have general preferences for birth order. Section 3 discusses the empirical implications associated with the theory. Section 4 describes our data. Section 5 presents our main results, and section 6 concludes.

2 A model of birth order and family size

This section presents a simple model of fertility in which parents treat children differentially. Our focus is on parents' choices regarding family size and their children's human capital. We keep in the background other endogenous choices that parents may make (e.g., choices regarding parental consumption and time allocation). A general model is presented in the Appendix. We assume that family size is a continuous variable and that parents make all decisions in a single stage of choice.[3] Children are born sequentially but within this single stage.

Let N represents family size. Let i represent the order in which children are born, their *birth order*. Birth order is jointly realized with family size, $i \in [0, N]$. A child's human capital is a function of her birth order, $h(i)$. Parents are altruistic; their utility depends on family size, N, and the human capital at each birth order, $h(i)$, as in

$$U(h, N) \equiv \int_0^N u(h(i), i)di, \tag{1}$$

where the function $u(h(i), i)$ denotes the parents' utility function. This represents the indirect utility function that nets out the other choices parents make (i.e., choices other than those of N and $h(i)$) after taking into account constraints other than the budget constraint

$$\int_0^N p(i)h(i)di \leq Y, \tag{2}$$

in which Y is the parents' total spending in their children and $p(i) > 0$ is the (given) cost per unit of human capital for a child born at order i. The cost $p(i)$ captures pure pecuniary advantages for some birth orders.

[3] In the empirical analysis, we take family size to be a discrete variable. The central predictions of the model are not sensitive to whether family size is modeled as a continuous or a discrete variable.

The utility function $u(h(i), i)$ captures birth order differences in child endowments, ability, and pure parental preferences.[4] Human capital will differ across birth orders as long as $\partial^2 u(h,i)/\partial h \partial i \neq 0$ or $\partial p(i)/\partial i \neq 0$; see Proposition 1(iii) below. We assume that $u \in \mathcal{C}^2$ and satisfies Inada conditions, particularly $\lim_{h \to 0} \partial u(h,i)/\partial h = \infty$ for all (h, i). We assume stronger concavity conditions than usual to ensure that the problem is globally concave.[5] Parental preferences are gender-neutral and there is no mortality.

Parents maximize their utility (1) subject to the budget set (2). Therefore, optimal decisions are represented by a family size $N^* \in [0, N^+]$, where N^+ is a biological upper bound of fertility, and by a bounded *human capital profile* $h^* \in \mathcal{C}^1$. Optimal family size satisfies

$$u(h^*(N^*), N^*) = \lambda p(N^*) h^*(N^*), \tag{3}$$

and human capital satisfies

$$\frac{\partial u(h^*(i), i)}{\partial h(i)} = \lambda p(i), \tag{4}$$

for all $i \in [0, N^*]$, where λ is the Lagrange multiplier associated with (2).

The first-order conditions are fairly intuitive. Expression (3) says that an additional child adds the term $u(h^*(N^*), N^*)$ to the sum of utilities of the other children. Increasing N, however, requires additional spending; parents must invest $h^*(N^*)$ at a cost $p(N^*)$. Expression (4) says that the utility gains from human capital investments at birth order i must, at the margin, equal their cost. Combining (3) and (4) yields

$$\frac{1}{p(N^*)} \frac{u(h^*(N^*), N^*)}{h^*(N^*)} = \frac{1}{p(i)} \frac{\partial u(h^*(i), i)}{\partial h(i)}, \tag{5}$$

for all $i \in [0, N^*]$. The left-hand side of (5) is the per dollar average utility of the "last" child and the right-hand-side is the (constant) per dollar marginal utility of the other children. Thus expression (5) simply says that the average gains from the "last" child must compensate the marginal losses associated with a reduction in the human capital of inframarginal children.

[4]The literature on the intrahousehold allocation of resources usually relies on an *earnings production function* $f(h(i), e(i), i)$, where $e(i)$ is a child endowment; see, e.g., Behrman et al. [10]. If parental utility is written as $u(h(i), i) \equiv u(f(h(i), e(i), i), i)$, the index i can be seen as the result of the differences just listed. Our results do not rely on a separate identification of the contribution of endowments, intellectual ability, and pure preferences.

[5]The maximization of (1) subject to (2) is an *isoperimetric* problem; see Hestenes [20]. We assume $\partial^2 u(h,i)/\partial h^2 < 0$ for all (h, i). To ensure that the optimal choices of N are interior, we assume $\partial u(h,i)/\partial i - h(\partial u(h,i)/\partial h)(\partial p(i)/\partial i)/p(i) < 0$ for all (h, i). This second order condition is standard in variational problems with optimal endpoints; see Vincent and Brusch ([35], Theorem 3.1).

To characterize the human capital profiles, we can write (4) as a differential equation

$$\frac{\partial h^*(i)}{\partial i} = \Delta(h^*(i), i), \tag{6}$$

for all $i \in [0, N^*]$, where

$$\Delta(h,i) \equiv \frac{\partial u(h,i)/\partial h}{\partial^2 u(h,i)/\partial h^2}\left(\frac{\partial p(i)/\partial i}{p(i)}\right) - \frac{\partial^2 u(h,i)/\partial h \partial i}{\partial^2 u(h,i)/\partial h^2}. \tag{7}$$

Parental choices involve two decisions. First, the optimal family size N^*, which is determined by (3), and, second, the optimal human capital associated with birth order i, which follows from expression (6) and terminal condition $\{h^*(N^*), N^*\}$. In this context, $\Delta(h,i)$ defines the shape of the human capital profiles for any family size, including the optimal choice N^*.

Proposition 1 *(i) If $\Delta(h,i) < (>)0$, for all (h,i), then human capital is a decreasing (increasing) function of birth order. (ii) If $\Delta(h,i) = \delta$, then the human capital profile is linear with slope δ. (iii) If $\Delta(h,i) = 0$, then human capital is allocated equally among all children, e.g., the human capital profile is "flat."*

Proof. The proof follows from (6). Equivalently, consider the integral equation: $h^*(i) = h^*(N^*) - \int_i^{N^*} \Delta(h^*(x), x)dx$. (i) follows from the differentiation under the integral sign. In (ii), $h^*(i) = [h^*(N^*) - \delta N^*] + \delta i$. Finally, (iii) implies that $h^*(i) = h^*(N^*)$, which is constant across birth orders. ■

Case (i) in proposition 1 shows that the model yields human capital profiles of any general shape depending on the preference and cost differentials. In case (ii), changes in family size only alter the intercept of the human capital profile. However, as the proof makes clear, the change in the intercept is *not* independent of the slope parameter δ. As discussed below, this point plays a central role in the empirical implementation of the model. Case (iii) on the other hand, assumes that parents treat all of their children equally, e.g., $\Delta(h,i) = \partial^2 u(h,i)/\partial h \partial i = \partial p(i)/\partial i = 0$. This special case is the Becker and Lewis [6] formulation.

Proposition 1 however, does not describe the effects of exogenous changes in family size on children's human capital, an idea motivating the entire empirical literature analyzing the "quality-quantity trade-off". But identifying exogenous changes in family size is not an easy task, since the optimal family size N^* is endogenous. This explains why in this literature *twin* *births* are commonly used as exogenous sources of variation (instruments) for N^* (Rosenzweig and Wolpin [30]). The conceptual experiment behind this logic is as follows. Parents plans for their children's human capital satisfy (6), and parents plans for family size satisfy (3) but only for families without twins. Let $\{h^*(i|N^*) : i \in [0, N^*]\}$ be the optimal human capital of child with birth order i in a family of size N^*. But choosing family size involves the possibility of multiple births. Suppose that twinning is unanticipated and undiversifiable, and denote by $Z \in \{0, z\}$ the occurrence of a twin birth. Therefore, if there are twins, family size becomes $N^* + z$ with $z > 0$ representing the "additional exogenous children." Proposition 2 characterizes how human capital changes for existing children changes with an exogenous increase in family size.

Proposition 2 *The human capital profile of families of size N^* dominates (i.e., it lies above) the human capital profile of families of size $N^* + z$ with $z > 0$. That is, $h^*(i|N^*) > h^*(i|N^* + z)$, for all $i \in [0, N^*]$.*

Proof. Human capital profiles are solutions to (6). Because of the existence and uniqueness theorem for differential equations, one and only one integral curve passes through each terminal point (Hestenes ([20], Appendix, Theorem 3.1).[6] Accordingly, any two human capital profiles with different end values should not cross (Hestenes ([20], Appendix, Theorem 4.1). Next, assume to the contrary that $h^*(i|N^*) \leq h^*(i|N^* + z)$ for all $i \in [0, N^*]$. Both plans exhaust spending. Then,

$$\int_0^{N^*} p(i)[h^*(i|N^*) - h^*(i|N^* + z)]di = \int_{N^*}^{N^*+z} p(i)h^*(i|N^* + z)di. \tag{8}$$

Since spending in the additional children is positive (e.g., the right-hand side of (8)), $h^*(i|N^*) \leq h^*(i|N^* + z)$ for $i \in [0, N^*]$ is a contradiction. ■

This result is a direct consequence of the decline in available resources per child due to the increase in family size. In Proposition 2, total spending Y is constant, but the results would be maintained if the presence of additional children lowers parental spending.[7]

[6] These arguments are similar in spirit to Brock's [13] analysis of changes in the planning horizon in the neoclassical growth model. An alternative proof based on standard comparative statics is given in the Appendix, as an intermediate step of the proof of Proposition 3. Our results might be proved by *monotone comparative statics* methods which have generalized Brock's [13] results; see, e.g., Amir [2] and Milgrom and Shannon [24]. These methods may be useful to obtain more general results than those presented here.

[7] The problematic case is when Y increases "too much" so that the budget constraint (2) is relaxed. Let $Y'(z)$

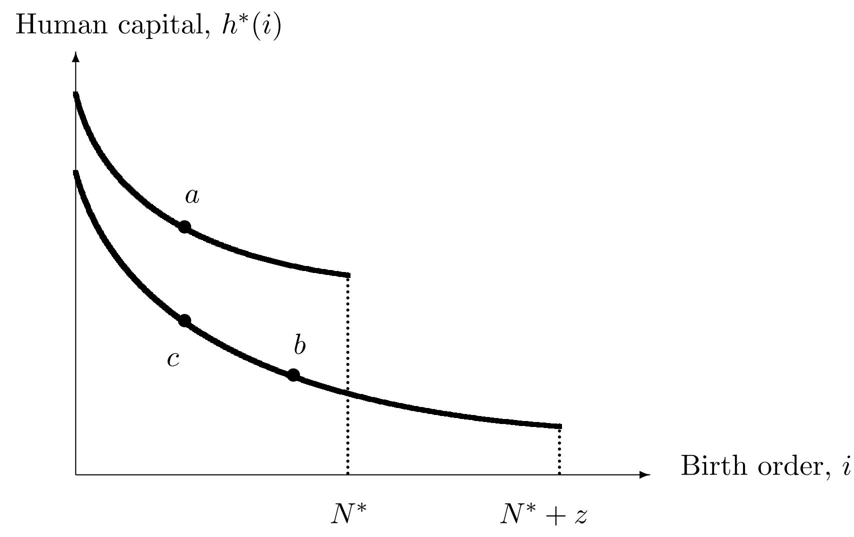

Figure 1: Human capital profiles under negative birth order effects.

Figure 1 represents graphically Proposition 2 in the case of negative birth order effects; the case of positive birth order effects is represented in Figure 2. Both figures indicate that children's human capital profiles in larger families should be below the human capital profiles of smaller families, regardless of the direction of the parental birth order predisposition.

In Proposition 2, twins "shift" family size but leave the shape of the human capital profile unaffected. This means that parental predispositions toward birth order do not change with twinning. That is, the preference and cost differentials subsumed in $\Delta(h, i)$ should be independent of the presence of twins: the additional children z should influence $\Delta(h^*(i|N^*), i)$ only through $h^*(i|N^*+z)$. This exclusion restriction allows us to find a solution for (6) while treating N^* and $N^* + z$ as different terminal points. In our empirical analysis below, we will focus on twins at the last birth because the occurrence of twins at that stage best matches the theoretical description just presented.

Proposition 2 compares individual-level human capital for the same birth orders in families of size N^* and families of size $N^* + z$, but it is silent about the human capital investments of the additional children, i.e., children of birth orders $i \in (N^*, N^* + z]$, and consequently, about the effect of the exogenous change on household's average human capital. Let $H(N^*)$ denote

denote parental spending for parents with additional children. If the extra resources, $Y'(z) - Y$, are larger than the extra cost in (8), parents with additional children will be able to invest more than (or at least as much as) parents without additional children.

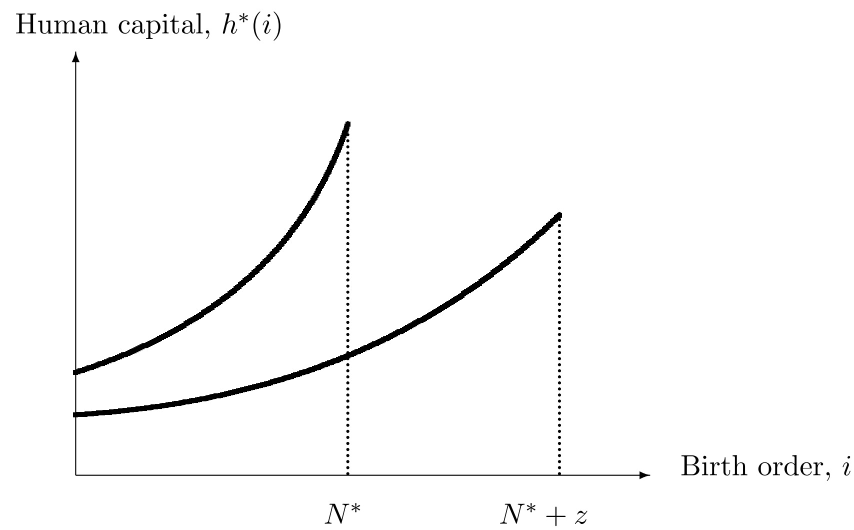

Figure 2: Human capital profiles under positive birth order effects.

the average human capital of families of size N^*,

$$H(N^*) \equiv \frac{1}{N^*} \int_0^{N^*} h^*(i|N^*)di. \tag{9}$$

Proposition 3 is our general version of the "quantity-quality trade-off" in the presence of birth order effects.

Proposition 3 *Consider the following conditions:*

(a) $\Delta(h, i) \leq 0$*, for all* (h, i)*,*

(b) $\Delta(h, i) > 0$ *and* $\partial^2 u(h, i)/\partial h \partial i \geq -\partial^2 u(h, i)/\partial h^2$*, for all* (h, i)*.*

If either (a) *or* (b) *hold, then average human capital of families of size* N^* *is larger than average human capital of families of size* $N^* + z$*. That is,* $H(N^*) > H(N^* + z)$*.*

Proof. See Appendix ■

To get a better intuition for this result, and conditions (a) and (b), it is useful to notice that an exogenous increase in family size affects average human capital through two different channels: an *income channel* and a *substitution channel*. As in Proposition 2, the income channel arises because the available resources per child decrease as family size increases. The substitution channel arises because the unit cost of human capital, $p(i)$, may be lower for the additional children (i.e., $i \in (N^*, N^* + z]$) than for the existing children (i.e., $i \in [0, N^*]$). Indeed, the unit cost of human capital for the additional children may be so low that average human capital may increase as family size increases.

Condition (a) guarantees that the substitution channel works in the same direction as the income channel. Thus, as family size increases, the average cost per unit of human capital also increases. In this case, substituting toward the additional children tightens the budget constraint. With that in mind, the intuition behind this condition is simple: the additional children lower average human capital because they contribute negatively to the family's average; see Figure 1. Condition (b) on the other hand, guarantees that the substitution channel is weaker than the income channel. The inequality $\partial^2 u(h,i)/\partial h \partial i \geq -\partial^2 u(h,i)/\partial h^2$ ensures that reducing the human capital of the existing children generates marginal utility losses that are sufficiently strong to limit parents' substitution toward the additional children.

A particular case of our "quantity-quality trade-off" formulation does not require either condition (a) or condition (b) to hold. It relies, however, on a special assumption.

Proposition 4 *Suppose that $p(i) = p$ for all i. Then average human capital decreases nonlinearly with family size.*

Proof. The result is a trivial consequence of expression (2). If $p(i) = p$, then the budget constraint can be written simply as

$$pNH \leq Y, \tag{10}$$

which implies that $H(N^*) = Y/pN^* > Y/p(N^* + z) = H(N^* + z)$. ■

Proposition 4 assumes that all children face the same human capital cost. This assumption eliminates the *substitution* channel and forces N and H to enter *multiplicatively* in the budget set (10). This result is central in Becker and Lewis [6]'s formulation of the "quantity-quality trade-off and illustrates why the conventional framework cannot be used to study birth order effects, because, by assumption, all children are treated equally. Under equal treatment, individual and average human capital exactly coincide: $h^*(i) = H(N^*)$ for all i. Propositions 3 and 4, do not assume that children are treated equally. Importantly, we shall see in the next section that Proposition 4 describes the empirically relevant model formulation.

On the contrary, the "quantity-quality trade-off" in Propositions 3 and 4 is more general than Becker and Lewis [6] formulation. It allows for a general analysis of the study of birth order effects.

In closing, we note that, throughout this section, we have used a utility function $u(h(i), i)$ and a linear aggregator across children's utilities. The Appendix presents the more general aggregators used by Behrman et al. [10] and Barro and Becker [4]. We have also abstracted from many considerations relevant to the allocation of resources within a household. For example, by treating parental spending as given, we have abstracted from the resource allocation between parents and children. We also consider parental consumption in the Appendix. The Appendix demonstrates the robustness of the basic theoretical predictions to these generalizations.

3 Empirical implications of the theory

To facilitate the exposition of the econometric model, the remainder of the paper treats birth order i as a discrete variable. We index birth order specific objects by subscript i, and family specific objects by the subscript j. For example, h^*_{ij} is the (optimal) human capital of a child with birth order i in family j.

To understand the empirical implications of the theory, assume that the utility function (1) takes the single index form

$$u_j(h_{ij}, i) = u_j(h_{ij} - \delta_i - v_j - \varepsilon_{ij}), \tag{11}$$

where δ_i and v_j represent birth order and family specific preferences, respectively, and ε_{ij} represents a family and birth order specific idiosyncratic preference shock; that is, $\mathbb{E}[\varepsilon_{ij}|i, j] = 0$ for all i and j. The terms δ_i and v_j in (11) are related to the "birth order effect" and the "family size effect" whose full content is specified below.

Recall that parental choices of N^*_j and h^*_{ij} satisfy (3) and (4), respectively. Let μ_{ij} be the *inverse function* of (4), i.e., $\mu_{ij} \equiv \{\partial u_j / \partial h_{ij}\}^{-1}(\lambda_j p_{ij})$. Thus, μ_{ij} is an implicit function of parental spending Y_j, family size N^*_j (via the Lagrange multiplier λ_j), family-specific preferences (via the utility function u_j), and birth order i (via the human capital cost p_{ij}). The following regression equation may now be obtained from expressions (4) and (11):

$$h^*_{ij} = \delta_i + \mu_{ij} + v_j + \varepsilon_{ij}, \tag{12}$$

for $i = 1, ..., N_j^*$.

In order to endow (12) with empirical content, we need to impose further restriction on μ_{ij}. Particularly, if $p_{ij} = p_j$, then μ_{ij} does not vary by birth order i and, with a slight abuse of notation, can be subsumed into the family fixed effect v_j. This is why we noted above that Proposition 4 represents the empirically relevant formulation of our theory. The family fixed effect now represents family specific preferences, expenditures, and prices, and the human capital equation now reads

$$h_{ij}^* = \delta_i + v_j + \varepsilon_{ij}. \tag{13}$$

Family size N_j^* is not explicit in (13), although it is part of the family fixed effect v_j. Averaging expression (13) at the family level yields

$$v_j = H(N_j^*) - \bar{\delta}(N_j^*), \tag{14}$$

where $H(N_j^*)$ is family j's average human capital and $\bar{\delta}(N_j^*)$ is the *average birth order effect*,

$$\bar{\delta}(N_j^*) = \frac{1}{N_j^*} \sum\nolimits_{i=1}^{N_j^*} \delta_i. \tag{15}$$

The theory suggests a nonlinear relationship between H_j and N_j^* (see Proposition 4). Therefore, we adopt the following semi-nonparametric relationship in the empirical analysis:

$$H(N_j^*) = \alpha + \eta(N_j^*) + \xi_j, \tag{16}$$

where $\eta(N_j^*)$ is an unrestricted function of N_j^*, ξ_j is a function of parental spending and preferences. N_j^* is endogenous in relation to ξ_j. Substituting (14)-(16) into (13) yields

$$h_{ij}^* = \alpha + \delta_i + \eta(N_j^*) - \bar{\delta}(N_j^*) + \xi_j + \varepsilon_{ij}, \tag{17}$$

for $i = 1, ..., N_j^*$. We refer to $\partial\eta(N_j^*)/\partial N_j^*$ as the *family size effect* on the grounds that average human capital is also the human capital of the "typical" child in a family of size N_j^*. That is, $H(N_j^*)$ is the human capital of a child for which $\delta_i = \bar{\delta}(N_j^*)$; see (16) and (17).

Expression (17) shows that an exogenous change in family size, from N_j^* to $N_j^* + z$, has two distinct effects on an individual's human capital. The first is the family size effect, $H(N_j^* +$

$z) - H(N_j^*) = \eta(N_j^* + z) - \eta(N_j^*)$. As we remarked after Proposition 3, this effect is associated with our general "quantity-quality trade-off" and holds independently of birth order effects. The second is the average birth order effect, $\bar{\delta}(N_j^* + z) - \bar{\delta}(N_j^*)$. This effect arises due to the parent's predisposition toward certain birth orders. In effect, as we stress throughout the analysis, one cannot manipulate family size without manipulating the birth order configuration within the family.

Expression (17) imposes no special restrictions on the shape of the human capital profiles. A special case is that of linear functions: $\delta_i = \delta i$ and $\eta(N_j^*) = \beta N_j^*$. In this case, the birth order and family size effects are each represented by a single parameter: δ and β, respectively. The average birth order effect is $\bar{\delta}(N_j^*) = \delta(N_j^* + 1)/2$, which is linear in N_j^* with slope $\delta/2$.[8] Another special and testable case is Becker and Lewis [6] formulation: $\delta_i = 0$ for all i. In this case, $\bar{\delta}(N_j^*) = 0$ and only the family size effect is relevant.

In proceeding it is useful to restate (17) in vector notation (no additional restriction imposed). Define $\boldsymbol{\iota}_A$ to be a vector whose A-th entry equals 1 and all other entries equal 0. The dimension of $\boldsymbol{\iota}_A$ is N^+. Similarly, let $\boldsymbol{\delta} \equiv (\delta_1, ..., \delta_{N^+})$, $\boldsymbol{\beta} \equiv (\beta_1, ..., \beta_{N^+})$, and $\bar{\boldsymbol{\delta}} \equiv (\bar{\delta}_1, ..., \bar{\delta}_{N^+})$. Our main outcome equation (17) can now be written as:

$$h_{ij}^* = \alpha + \boldsymbol{\iota}_i' \boldsymbol{\delta} + \boldsymbol{\iota}_{N_j^*}' (\boldsymbol{\beta} - \bar{\boldsymbol{\delta}}) + \xi_j + \varepsilon_{ij}, \tag{18}$$

for all $i = 1, ..., N_j^*$. The parameters of interest are contained in the vectors $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$, the effects of birth order and family size on human capital accumulation, respectively. Consequently, our empirical work is based on the following regression model:

$$h_{ij}^* = a + \boldsymbol{\iota}_i' \mathbf{d} + \boldsymbol{\iota}_{N_j^*}' \mathbf{b} + \epsilon_{ij}, \tag{19}$$

where $\mathbf{d} = (d_1, ..., d_{N^+})$ and $\mathbf{b} = (b_1, ..., b_{N^+})$ are vectors of parameters, and ϵ_{ij} is an error term. Notice that ϵ_{ij} contains the family fixed effect ξ_j.[9] Since $\mathbf{d} = \boldsymbol{\delta}$ and $\mathbf{b} = \boldsymbol{\beta} - \bar{\boldsymbol{\delta}}$, (19) is simply a

[8] This case further illustrates that summations can be replaced by integrals without loss of generality. For example, the continuous analog of (15) is

$$\bar{\delta}(N_j^*) = \frac{1}{N_j^*} \int_0^{N_j^*} \delta(i) di.$$

When $\delta(i) = \delta i$, $\bar{\delta}(N_j^*) = \delta N_j^*/2$, which is also linear with slope $\delta/2$.

[9] Throughout, we suppose that information on individual human capital h_{ij}^*, birth order i, family size N_j^*, and family identifiers is available. Fertility is completed for all families, and families have at least two children. That

reparameterization of (18).

3.1 Existing empirical strategies

The overriding issue in an empirical analysis of education, birth order and family size is a dual endogeneity problem. Family size is chosen in response to parental preferences and spending, and the cost of human capital, all of which are part of the composite family fixed effect. A family's birth order configuration depends deterministically on family size, and thus also correlates with the family fixed effect. It is thus clear that the Ordinary Least Squares (OLS) estimator does not admit consistent estimation of the parameters of interest $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$.

Black et al. [12] proposed an empirical strategy for dealing with the dual endogeneity problem. Stated in the notation of our model, their outcome equation allows for a linear effect of family size and nonlinear effects of birth order. That is,

$$h^*_{ij} = a + \boldsymbol{\iota}'_i\mathbf{d} + bN^*_j + \epsilon_{ij}, \tag{20}$$

and their parameter of interest is b, i.e. the regression coefficient on family size N^*_j on individual human capital formation h^*_{ij} controlling for birth order i. Birth order effects are nuisance parameters in this setup.[10] Black et al. [12] do not discuss the theoretical foundations motivating this outcome equation

From (18) it is clear that a regression equation involving nonlinear birth order effects $\boldsymbol{\iota}'_i\mathbf{d}$ and a linear family size effect bN^*_j, as does (20), might not be appropriate. The reason is that the regression coefficient on family size picks up variation in human capital induced by difference in the average birth order effect across families. If birth order effects are nonlinear, average birth order effect varies nonlinearly across families as well. Mogstad and Wiswall [26] also argue, although from an empirical point of view, that a regression equation like (20), where family size enters linearly, is unnecessarily restrictive. They devise a non-parametric Instrumental Variables (IV) strategy for the case where family size enters nonlinearly, i.e. takes the form $\boldsymbol{\iota}'_{N^*_j}\mathbf{b}$. This extension allows an outcome equation of a form that is consistent with our theoretical model (see (19)). In the following, we consider the Mogstad and Wiswall [26] extension of the Black

is, we cannot estimate b_1 and thus we normalize the first entry in $\mathbf{b}$ to zero. To avoid collinearity problems, we further normalize the first entry in $\mathbf{d}$ and the second entry in $\mathbf{b}$ to zero. It is straightforward to include exogenous regressors; we do so in the empirical analysis.

[10] Black et al. [12] subsequently estimate birth order effects using the within-family estimator.

et al. [12] estimation procedure.

Using (19), consider the first $k-1$ children in families with k or more children ($k \geq 2$). The birth order configuration within each of these families is identical and, therefore, independent of the family fixed effect. In terms of (19), $\epsilon_{ij} \perp \boldsymbol{\iota}_i' | (i \leq k-1, N_j^* \geq k)$. Family size indicators of course remain endogenous regressors. Mogstad and Wiswall [26] show how one can exploit exogenous variation in family size induced by twin births to develop a nonparametric IV estimator for $\mathbf{b}$ in (19). The procedure can be implemented for different value of $k \geq 2$. For our purposes it suffices to consider the case $k = 2$. Denote the resulting estimator by $\widehat{\mathbf{b}}_{\text{IV}}$.

If the instrumental variable is valid (in the usual sense), it is evident that $\widehat{\mathbf{b}}_{\text{IV}}$, interpreted within the context of our model, has the following probability limit:

$$\text{plim}\, \widehat{\mathbf{b}}_{\text{IV}} = \mathbf{b} = \boldsymbol{\beta} - \bar{\boldsymbol{\delta}}. \tag{21}$$

Hence, $\widehat{\mathbf{b}}_{\text{IV}}$ consistently estimates the causal effect of family size on mean individual human capital conditional on birth order i. Indeed, it is straightforward to show that $b_N = \mathbb{E}[h_{ij}^*|i, N] - \mathbb{E}[h_{ij}^*|i, N = 2]$ for $N \geq 3$ (recall the data is such that $N > 2$ and we normalize $b_2 = \beta_2 - \bar{\delta}_2 = 0$, cf. footnote 9). Notice that our theory of fertility implies that $\mathbf{b} < \mathbf{0}$ (element-by-element, see Proposition 2).

However, the causal effect of family size on mean individual human capital conditional on birth order i confounds two separate effects. Indeed, $b_N = \{\mathbb{E}[H_j|N] - \mathbb{E}[H_j|N = 2]\} + \{\mathbb{E}[h_{ij}^* - H_j|i, N] - \mathbb{E}[h_{ij}^* - H_j|i, N]\} = \beta_N - \bar{\delta}_N$. The first effect, β_N, reflects the difference in human capital between a child of average birth order in a family of size N and a child of average birth order in a family of size $N = 2$. This is our generalized family size effect (see Proposition 3). The second effect, $-\bar{\delta}_N$, reflects the change in the average birth order effect between families of size N and families of size $N = 2$. The two effects are confounded here because, as discussed in the context of our model, one cannot manipulate family size without manipulating the birth order configuration. Disentangling the two effects is important. Indeed, testing Proposition 3 requires a estimator of $\boldsymbol{\beta}$, and, as we show further below, understanding whether dispersion in human capital acquisition arises within or between families, also requires one to separate $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$. This illustrates the contribution and advantages of our approach over the previous literature.

3.2 A theory-driven empirical strategy

We now present a simple theory-driven empirical strategy that solves the dual endogeneity problem associated with birth order and family size. This strategy delivers consistent estimates of the full vectors of birth order effects $\boldsymbol{\delta}$ and family size effects $\boldsymbol{\beta}$, even when parents have predispositions towards certain birth orders. It draws its motivation from the decomposition of the variance of individual human capital. In particular, using expression (18), the variance of individual human capital can be written as

$$Var[h^*_{ij}] = \mathbb{E}\{Var[\boldsymbol{\iota}'_i\boldsymbol{\delta} + \varepsilon_{ij}|j]\} + Var[\boldsymbol{\iota}'_{N^*_j}\boldsymbol{\beta} + \xi_j]. \tag{22}$$

from where we conclude that the within-family variation in human capital, $\mathbb{E}\{Var[\boldsymbol{\iota}'_i\boldsymbol{\delta} + \varepsilon_{ij}|j]\}$, is only a function of $\boldsymbol{\delta}$ (and not of $\boldsymbol{\beta}$), whereas the between-family variation, $Var[\boldsymbol{\iota}'_{N^*_j}\boldsymbol{\beta} + \xi_j]$, is only a function of $\boldsymbol{\beta}$ (and not of $\boldsymbol{\delta}$). Hence, we can identify birth order effects from within-family variation and family size effects from between-family variation in human capital.[11] We next apply this simple logic to overcome the dual endogeneity problem of birth order and family size.

In the empirical analysis we augment (19) with individual level controls $\mathbf{x}_{ij}$ (age and sex) and family level controls $\mathbf{w}_j$ (mother's age and education and father's age and education). We treat the controls as strictly exogenous.

Estimating birth order effects. Let $\widehat{\mathbf{d}}_{\text{Within}}$ denote the within-family OLS estimate of the birth order effect in (19). The within-family estimator is equivalent to the family fixed effect estimator of the birth order effects. Applying the within-family transformation to (19) produces the following estimation equation:

$$h^*_{ij} - H(N^*_j) = \left[\boldsymbol{\iota}_i - \frac{1}{N^*_j}\sum\nolimits_{i=1}^{N^*_j}\boldsymbol{\iota}_i\right]'\mathbf{d} + \left[\mathbf{x}_{ij} - \frac{1}{N^*_j}\sum\nolimits_{i=1}^{N^*_j}\mathbf{x}_{ij}\right]'\mathbf{g} + [\epsilon_{ij} - \xi_j]\,, \tag{23}$$

where we have used that $(1/N^*_j)\sum_{i=1}^{N^*_j}\epsilon_{ij} = \xi_j$ and $\mathbf{g}$ contains the parameters associated with the vector of individual level controls. The within-family transformation sweeps out all family

[11] The use of within and between sources of variation is loosely reminiscent of Hausman and Taylor [19]. In a linear panel data model, Hausman and Taylor [19] propose an IV estimator that separately exploit these sources of variation. They make use of "internal" instruments obtained from exogenous regressors. We do not have regressors well suited for this purpose. Instead, we exploit conventional "external" instrumental variables.

level variables, including family size N_j^*. Moreover, the error-term $[\epsilon_{ij} - \xi_j]$ has been purged from the family fixed effect, thus solving the endogeneity problem in relation to the vector of birth order effects. It is standard to show that

$$\text{plim}\, \hat{\mathbf{d}}_{\text{Within}} = \boldsymbol{\delta}.$$

In other words, *the within-family OLS estimator* $\hat{\mathbf{d}}_{\text{Within}}$ *correctly identifies* $\boldsymbol{\delta}$.[12]

Estimating family size effects. Family size effects are estimated using between-family variation in average human capital, which is not contaminated by birth order effects. Averaging (19) at the family level yields

$$H(N_j^*) = a + \left[\frac{1}{N_j^*}\sum_{i=1}^{N_j^*} \boldsymbol{\iota}_i\right]' \mathbf{d} + \boldsymbol{\iota}'_{N_j^*}\boldsymbol{b} + \left[\frac{1}{N_j^*}\sum_{i=1}^{N_j^*} \mathbf{x}_{ij}\right]' \mathbf{g} + \mathbf{w}'_j\mathbf{f} + \xi_j, \tag{24}$$

where $\mathbf{f}$ contains the parameters associated with the vector of family level controls.

Our goal is to estimate $\boldsymbol{\beta}$, the causal effect of family size N_j^* on the human capital of the "typical" child. To this effect, notice that $\left[(1/N_j^*)\sum_{i=1}^{N_j^*} \boldsymbol{\iota}_i\right]' \mathbf{d} = \boldsymbol{\iota}'_{N_j^*}\bar{\boldsymbol{d}}$ where $\bar{\boldsymbol{d}} \equiv (1/N_j^*)\sum_{i=1}^{N_j^*} d_i$, and that an estimate of $\mathbf{g}$ is available from (23). Now, rewrite (24) as

$$\left(H(N_j^*) - \left[\frac{1}{N_j^*}\sum_{i=1}^{N_j^*} \mathbf{x}_{ij}\right]' \hat{\mathbf{g}}\right) = a + \boldsymbol{\iota}'_{N_j^*}\boldsymbol{c} + \mathbf{w}'_j\mathbf{f} + \xi_j, \tag{25}$$

where $\boldsymbol{c} \equiv \bar{\boldsymbol{d}} + \boldsymbol{b}$. The left hand side of (25) is family level average human capital net of the effect of (strictly exogenous) individual level controls,[13] and family size N_j^* is endogenous in relation to the family fixed effect ξ_j. We overcome this problem by employing an IV estimator. Specifically, following the literature, we use twinning as a source of exogenous variation in family size.[14] Denote the resulting IV estimator of $\boldsymbol{c}$ by $\hat{\mathbf{c}}_{\text{Between}}$. Using (18) and (25), it is standard

[12] Black et al. [12] uses the same strategy for estimating birth order effects.

[13] Alternatively one could leave the individual level controls averaged at the family level on the right hand side and re-estimate $\mathbf{g}$ in (25).

[14] As we noted in our discussion following Proposition 2, we can think of a twin births as a random and an unplanned event, uncorrelated with any time-variant or time-invariant family characteristics, but correlated with family size. Using Chinese data, Rosenzweig and Zhang [31] show that twinning influences the education of non-twin siblings because of a change in family size, but also because of a change in the allocation of resources within the household. Specifically, they show that in rural areas in China parents shift resources from less endowed children to their healthier siblings. This concern is mitigated in the context of a developed country such as Norway (Black et al. [12] and Mogstad and Wiswall [26]) or Denmark (this paper).

to show that

$$\text{plim}\, \hat{\mathbf{c}}_{\text{Between}} = \boldsymbol{\beta}. \tag{26}$$

In other words, *the between-family IV estimator* $\hat{\mathbf{c}}_{\text{Between}}$ *correctly identifies* $\boldsymbol{\beta}$.

With respect to the specification of the model, our empirical analysis includes various specifications of the family size profile $\boldsymbol{\iota}'_{N^*}\boldsymbol{\beta}$. In the simple case where the family size profile is linear, i.e. $\boldsymbol{\iota}'_{N^*}\boldsymbol{\beta} = \beta N^*$, and thus characterized by a single parameter β, only a single instrumental variable is required for identification. In this case we use the occurrence of a twin birth in the last birth in a family as an instrumental variable for N_j^*. The IV regression is just-identified and is computed via a 2SLS procedure.

In the general case of a nonlinear family size profile, (25) contains a vector of endogenous variables $\boldsymbol{\iota}_{N^*}$. We have excluded families with $N = 1$ and treat $N = 2$ as the reference case. Hence, $\boldsymbol{\iota}_{N^*}$ contains $N^+ - 2$ endogenous variables, and identification thus requires (at least) $N^+ - 2$ instrumental variables. Following Mogstad and Wiswall [26] let $\tilde{z}_j^k$ take the value one if family j experienced a twin-birth in the k-th birth and the value zero otherwise. $\tilde{z}_j^k$ is defined in such a way that $\tilde{z}_j^k = 0$ for $N_j^* \leq k - 1$ (i.e. $\tilde{z}_j^k$ is zero when information on twin birth at parity k is missing due to truncation at N_j^*). Hence, $\tilde{z}_j^k$ is systematically related to the family fixed effect ξ_j in the subpopulation with $N_j^* > 2$. Formally, $\mathbb{E}[\tilde{z}_j^k \xi_j | \mathbf{w}_j] \neq 0$ for $k > 2$. Hence, the vector $\tilde{\mathbf{z}}_j = (\tilde{z}_j^2, \tilde{z}_j^3, ..., \tilde{z}_j^{N^+-1})'$ does not qualify as a vector of instrumental variables for $\boldsymbol{\iota}_{N^*}$.

Angrist et al. [3] and Mogstad and Wiswall [27] show how this problem of partially missing information can be circumvented. Indeed, let

$$z_j^k = \left(\tilde{z}_j^k - \mathbb{E}[\tilde{z}_j^k | \mathbf{w}_j, N_j^* \geq k] \right) \mathbf{1}(N_j^* \geq k) \tag{27}$$

for $k = 3, 4, ..., N^+ - 1$, where $\mathbf{1}(\cdot)$ is the indicator function. Then $\mathbf{z}_j = (z_j^2, z_j^3, ..., z_j^{N^+-1})'$ is a vector of valid instrumental variables, i.e. $\mathbb{E}[\mathbf{z}_j \xi_j | \mathbf{w}_j] = \mathbf{0}$. We estimate $\mathbb{E}[\tilde{z}_j^k | \mathbf{w}_j, N_j \geq k]$ in (27) by regressing $\tilde{z}_j^k$ onto all variables in $\mathbf{w}_j$ as well as their interactions within the relevant subsamples defined by $N_j^* \geq k$.

The variance-covariance matrix of the two-step estimator is estimated by applying the block-bootstrap, treating families as the blocks (Horowitz [21]).

4 Data

Our analysis data is extracted from IDA (*Integreret Database for Arbejdsmarkedsforskning*), a comprehensive Danish administrative panel dataset for the period 1980-2006 with annual observations on all individuals aged 15-70 and residing in Denmark with a social security number (CPR number). IDA contains detailed individual-level information on socioeconomic characteristics, including gender, education, and income. The data is constructed and collected for administrative purposes and contains very few measurement errors. Moreover, the data is population wide with a long period of observation. We can link children and parents, and thus identify siblings; we define siblings as children born to the same mother.

The unit of observation is an individual, and all outcome measurements are conducted in 2006. We impose a series of selection criteria, similar to those of Black et al. [12] who use Norwegian administrative data; these criteria are detailed in the Appendix.[15] We supplement IDA with information on the precise date of birth, which allows us to identify twinning (or rather multiple births). Information on date of birth is obtained from Statistics Denmark; the merging procedure is documented in the Appendix. We are left with all individuals in 2006 aged 25-41 whose mother was both alive and present in IDA at some point during 1980-2006, and who satisfy standard selection criteria.

Descriptive statistics. Table 1 presents some descriptive statistics for the full sample of 1,256,031 children in 581,159 families, excluding single child families. The average individual is 38.2 years old and has 12.9 years of education. 52 percent are males and their parents did not complete high school. On average, an individual has 1.7 siblings.

Table 2 presents the distribution of family size, including single child families. 50.2 percent of families have two children and less than one-third of the families have more than two children.[16]

[15]The most important aspects in our selection criteria are the following: First, we only retain children aged 25 or above in 2006 to ensure that our outcome measurements represent complete education. Second, we exclude families with children born after 1991 (aged 0-14 in 2006). This ensures that our family size measure represents completed fertility. Third, we exclude families in which at least one member (a child or one of the parents) has missing education data, and families where the mother was below 17 or above 49 when giving birth. The set of children in IDA consist of all individuals with a valid pointer to the mother's unique identifier, whose mother was alive, aged 15-70 and present in IDA at some point during 1980-2006. These restrictions reduce our sample size from 2,453,843 individuals (distributed over 1,272,874 families) to 1,256,031 individuals (distributed over 581,159) families. See Appendix for further details.

[16]The distribution of children in our sample (Denmark) differs from that reported by Black et al. [12] (Norway). The differences can be attributed to the lower fertility rate in Denmark and to the fact that Norwegian women marry more and earlier than Danish women; see Tsuya [34].

The average number of children in the family is 2.4. Table 3 presents the average education by family size and birth order. This table shows a clear negative association between an individual's education and family size, as well as between an individual's education and her birth order. Similar patterns are documented for the mother's and father's education.

Our empirical strategy exploits the within- and between-family variation in education. As a final set of descriptive statistics we note that total variation in educational attainment in our sample is 7.975, with 35 percent being attributable to within-family variation and 65 percent to between-family variation. These values imply that understanding differences in education within the family is as important as understanding differences in education between families.

5 Results

OLS results. We start by estimating the impact of family size on a child's education, as specified in expression (19), by OLS. The first column of Table 4 reports results from a linear specification of family size on child's education, controlling for the age and sex of the child. As expected, the family size coefficient is negative and implies that an additional child decreases schooling by a little more than a quarter of a year. Because family-specific characteristics might impact the choice of completed family size and educational choices, column 4 adds unrestricted indicators for the mother's and father's education, and a 5-year interval set of indicator variables for mother's and father's age. Adding demographic controls reduces the magnitude of the relationship by about 30 percent but the coefficient remains statistically significant.

In order to account for birth order effects, column 5 adds a linear control for birth order. Similar to previous findings (i.e., Black et al. [12]), the coefficient on family size is considerably reduced to -0.04, but remains significant. Allowing for a more flexible estimation by including indicator variables for birth order in column 6 does not change the results markedly.[17] The impact of family size is -0.066, smaller but comparable to the coefficient (-0.013) reported in Black et al. [12] using Norwegian data.

Birth order effects, whether included linearly or nonlinearly, are negative, large and highly significant. They suggest, for example, that a third child in the family has, on average, 0.61 fewer years of education (column 6 in Table 4). As we discussed earlier, the negative birth

[17] Notice however, that, according to our model, a regression equation where educational attainment is a linear function of family size but a nonlinear function of birth orders is likely to be misspecified.

order effects could simply reflect family-specific unobservable factors. In columns 7 and 8 of Table 4, we report results from estimating birth order effects while controlling for family fixed effects. The family indicators capture any time-invariant characteristics, including completed family size. Once we control for family fixed effects, the linear birth order effect is reduced by about a half to -0.178. Similar changes in magnitude occur when we estimate the regression including nonlinear birth order effects. The birth order results are similar in magnitude to the results reported by Black et al. [12].

IV results. The coefficients on family size reported in Table 4 are still likely to be biased, even when controlling for a rich set of demographic controls. We follow initially the instrumental variable approach that Black et al. [12] implement. As shown in section 3.1, this approach does not identify the family size effect, or the quantity-quality trade-off, as defined in Proposition 3, but it serves as an illustrative benchmark for the main empirical analysis to come.

We estimate the family size effect using three different samples: the first sample includes the educational outcomes of first born children in families with two or more children, the second sample includes the outcomes of the first two children in families with three or more children, and the third sample includes the outcomes of the first three children in families with four or more children. We instrument family size with twins at second birth in the first sample, twins at third birth in the second sample, and twins at fourth birth in the third sample, all while controlling for demographic characteristics of the parents and including linear birth order effects. Examining the outcomes of children born before the $k-$th birth avoids the problem that a twin birth changes the birth order of children born after the twin birth in addition to changing the family size.

Table 5 reports the results of this exercise. In all three samples, the first stage results show that there is a strong and highly significant relationship between a twin birth and family size. The OLS results, reported in column 1, show that there is a negative and statistically significant relationship between family size and education. The coefficient on family size in the first sample is -0.09 and increases to about -0.23 in families with four or more births. The instrumented family size effects are reported in column 3, and they vary by sample. The coefficient on family size, when estimated using a sample of first born children in families with two or more children, is about -0.024 but is not statistically distinguishable from zero. Similarly, the coefficient on family size estimated using a sample of the first three children in families with four or more children is positive, although insignificant. The coefficient on family size, when using a sample of the first two children in families with three or more children, is negative, sizable (-0.095), and significant at the 5 percent level. With the exception of this latter finding, the results are consistent with Black et al.'s [12] finding of no effect of family size on children's education.

Two-step estimation results. We now present the estimates from our two-step strategy. We showed in Section 3.1 that this strategy identifies both birth order effects and the family size effect as it is defined in Proposition 3. Our two main specifications are reported in Tables 6 and 7. Table 6 contains results for the case of linear birth order effects and a linear family size effect. Table 7 contains results for the case of nonlinear birth order effects and a nonlinear effect of family size. We further present a specification with nonlinear birth order effects and a linear family size effect in Table 8. As we have already mentioned, the latter setup is likely to misspecified, but we include it because it is a common specification in the literature and thus facilitate comparisons between our results and results reported in previous studies.

Recall that the first step estimates birth order effects, while controlling for family fixed effects. This is similar to the way birth order effects are estimated in Black et al. [12]. Column 7 of Table 4 reports this result when including a linear birth order term. For convenience, we report again these results in column 1 of Table 6. Likewise, the nonlinear birth order effects reported in column 8 of Table 4 are reproduced in column 1 of Table 7.

Column 2 of Table 6 reports OLS results of estimating equation (24), controlling for the average age and sex composition in the family. The results indicate that an increase of one child in the family size reduces the average educational level in the family by about 0.24 years. This coefficient is comparable to the -0.186 OLS coefficient we report in Table 4, column 2. Column 3 of Table 6 adds demographic controls; the coefficient is reduced to -0.163 but it remains highly significant. This estimate is larger in magnitude compared to the OLS coefficient (-0.040) we report in Table 4, column 5.

The OLS estimates do not account for the endogeneity of family size. Our IV strategy uses the incidence of twins at last birth to instrument for family size. Using twins at last birth ensures that desired family size is, on average, the same for families with singletons and for families with a twin birth. Moreover, using twin birth at last birth ensures that the family size changes without also changing the birth order of subsequent children.

Column 4 of Table 6 reports the IV coefficient of family size without demographic controls. We note that the Minimum eigenvalue statistic indicates that our instruments are strong. The estimated coefficient, -0.102, is significant at the one percent level. In addition to linear controls for birth order, column 5 of Table 6 controls for demographic characteristics. The estimated coefficient, -0.115, is also significant at the 1 percent level. Overall, the household-level IV estimates counter the individual-level IV estimates in Tables 4 and 5, as well as the previous findings of Black et al. [12]. Table 6 imply that an additional child significantly reduces the average number of years of schooling in the household by about one tenth of a year.

We now move on to our main set of results. These allow for nonlinear birth order effects and nonlinear family size effects and are reported in Table 7. Specifically, Table 7 includes separate indicators for families with 3, 4, and 5 children while using families with 2 children as the omitted category. In the first-stage we consider the case where birth order effects are also included nonlinearly. Because the second stage includes three endogenous family size effects, we follow the methodology proposed by Angrist et al. [3] and Mogstad and Wiswall [27] and construct multiple instruments using twin births at each birth order.[18]

The IV results are reported in columns 4 and 5 of Table 7. We note again that the Minimum eigenvalue statistic indicates that our instruments are strong. We focus initially on the specification with demographic controls (column 5). The effect of family size is negative and statistically significant for all family sizes. The reference category is $N = 2$. Hence, the point estimates in column 5 of Table 7 implies that an additional child in families of size two reduces average years of schooling by 0.12 years, an additional child in families of size three reduces average years of schooling by an additional 0.17 years, whereas an additional child in families of size four has only limited effect, reducing average years of schooling by by less then 0.01 years. This indicates that family size might impact educational attainment in a nonlinear way. Comparing column 4 and column 5 of Table 7, we note that without demographic controls the effect of family size for families of size three is insignificant. Otherwise, the two columns yield similar results.

Qualitatively, the results reported in Table 7 are in line with those obtained with linear birth order and family size effects. Three clear results stand out. First, birth order effects in

[18]We estimate the family size profile nonparametrically, although we restrict the estimated family size effects to be constant for families with $N \geq 5$. Given that the effect of $N = 1$ and $N = 2$ are normalized to zero, we need three instrumental variables. See the description in section 3, or consult Angrist et al. [3], Mogstad and Wiswall [27] and Mogstad and Wiswall [26], for details.

educational attainment are both economically and statistically significant. This result is in line with the existing literature. Second, our theoretically consistent family size effect is economically and statistically significant; that is, the data strongly suggests the presence of a quantity-quality trade-off. This is at odds with previous results reported in the literature. We have already discussed how this difference is due to previous studies confounding the quantity-quality trade-off (our Proposition 3) with differences in average birth order across families. Third, the effect of family size appears to be non-linear, although monotonic.

Taking a brief look towards Table 8 where family size is restricted to have a linear effect on educational attainment even if birth orders can have nonlinear effects, we see that the estimated family size effects is not strongly affected by the inclusion of the nonlinear birth order effects in the first step of the estimation strategy (compare to Table 6).

Finally, we have experimented with various ways of cutting the data in terms of excluding families with children born before 1960, 1963 and 1965 (i.e. deleting older cohorts from the analysis), or in terms of retaining only the 1962-1971 cohorts or the 1972-1981 cohorts. Our results are broadly robust to these experiments although we lose a bit of precision in the parameter estimates.

Empirical Content of Theoretical Predictions. Up to this point we have used the theoretical model to guide our empirical analysis, but the model in fact imposes restrictions on the data that allow us to test its empirical validity. This allow us to assert whether our empirical findings are in fact consistent with our theoretical framework.

Before turning to a formal test of the model's predictions, consider whether the numbers we have reported above a quantitatively meaningful and how they relate to those reported in the previous literature. This is most easily done by considering the implications of our framework under the assumption of linear family size and birth order effects. In this case, the (biased) individual-level IV estimate of the family size effect is $\hat{b}_{\text{IV}} = \hat{\beta} - \hat{\delta}/2$; see (21) and the discussion in Section 3. Table 5 column 3 shows that $\hat{b}_{\text{IV}} = -0.024$. Our two-step strategy consistently estimates β and δ. These estimates, reported in columns 1 and 5 of Table 6, are $\hat{\beta} = -0.115$ and $\hat{\delta} = -0.178$. Thus we predict $\hat{b}_{\text{IV}} = -0.115 - (-0.178/2) = -0.026$, which is virtually identical to the individual-level IV estimate and the corresponding results reported in Black et al. [12] (on Norwegian data).

Testing Propositions 2 and 3. Our theoretical framework provides a guide to separately identify birth order and family size effects. The theory is actually richer in that it provides a set of testable predictions formulated in Propositions 2 and 3. Proposition 2 states that the human capital profile of small families is always above the human capital profile of large families. Proposition 3 states that the average human capital of small families is larger than the average human capital of large families.

Within the context of our empirical model (18), Proposition 2 implies that

$$\boldsymbol{\iota}_N'(\boldsymbol{\beta} - \bar{\boldsymbol{\delta}}) - \boldsymbol{\iota}_{N+z}'(\boldsymbol{\beta} - \bar{\boldsymbol{\delta}}) > 0, \tag{28}$$

for $N = 2, 3, ..., (N^+ - 1)$ and $z = 1, 2, ..., (N^+ - N)$. With $N^+ = 5$, Proposition 2 thus imposes three restrictions on the parameter space.

Proposition 3 on the other hand, stipulates that

$$\boldsymbol{\iota}_N'\boldsymbol{\beta} - \boldsymbol{\iota}_{N+z}'\boldsymbol{\beta} > 0, \tag{29}$$

for $N = 2, 3, ..., (N^+ - 1)$ and $z = 1, 2, ..., (N^+ - N)$, introducing an additional three (independent) restrictions. Our empirical implementation does not impose these six restrictions. Therefore, we can use our estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, and their variance-covariance matrix, to construct a Wald-type test statistic that informs on the empirical validity of (28) and (29).

The theoretical predictions can be thus characterized by a set of inequality constraints implying that the null (where (28) and (29) holds) is a multivariate composite hypothesis. This impacts the asymptotic distribution of the Wald test statistic. In particular, the limiting distribution is a complicated mixture of χ^2-distributions, making the computation of critical values cumbersome (see e.g. Gourieroux et al. [17]). We overcome this difficulty by applying a relatively simple test procedure developed by Kodde and Palm [22].

The intuition behind the test is as follows. Consider the estimated parameter vector $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}})$ from the unrestricted model (i.e. the parameter estimates reported in Tables 6, 7 and 8). Let $\mathcal{S}$ be the feasible parameter space under the null. The Kodde and Palm test statistic D is the distance between $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}})$ and the "closest" parameter vector admissible under the null. If $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}) \in \mathcal{S}$, then $D = 0$, and if $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}}) \notin \mathcal{S}$, then $D > 0$. Hence, the null is rejected for "large" values of D. Kodde and Palm [22] derive an upper and lower bound, $\overline{D}$ and $\underline{D}$ for the critical value of D for a test of a given size. The decision rule is "reject the null if $D > \overline{D}$", and "do not reject the null if $D < \underline{D}$". If $\underline{D} \leq D \leq \overline{D}$, the exact critical value for D needs to be computed to obtain a conclusive test. Kodde and Palm [22] tabulate upper and lower bound critical values.

Tables 6, 7 and 8 report the results from our tests of Propositions 2 and 3. The propositions are not rejected in any of our multiple specifications. These results provide formal empirical support for our general theory for the analysis of the "quality-quantity trade off" in fertility models.

Within- and between-family variation in human capital. Our empirical strategy exploits the within- and between-family variation in education. We noted in the section 4 that 35 percent of the variation originates within families and 65 percent between families. Using the estimated structural parameters of the model $(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\beta}})$ and the estimated parameters associated with the exogenous controls $(\hat{\boldsymbol{f}}, \hat{\boldsymbol{g}})$ we provide a variance decomposition of the distribution of years of schooling, breaking down the within- and between component. The full decomposition is given as:

$$Var[h^*_{ij}] = \mathbb{E}\{Var[\boldsymbol{\iota}'_i \hat{\boldsymbol{\delta}} + \mathbf{x}'_{ij}\hat{\boldsymbol{g}} + \hat{\varepsilon}_{ij}|j]\} + Var[\mathbb{E}\{\boldsymbol{\iota}'_{N^*_j}\hat{\boldsymbol{\beta}} + \mathbf{x}'_{ij}\hat{\boldsymbol{g}} + \mathbf{w}'_j\hat{\boldsymbol{f}} + \hat{\xi}_j|j\}], \tag{30}$$

where the first term on the right-hand side represents within-family variation and the second term between-family variation.

Table 9 reports the variance decomposition based on our preferred specification with nonlinear birth order effects and nonlinear family size effects (parameter estimates reported in Table 7). The first row in Table 9 is the overall empirical within- and between variance decomposition referred to in section 4. Total variation in years of schooling is 7.954 with the within- and between family components accounting for 2.799 and 5.176.

The second column in Table 9 breaks the within-family variation in years of schooling into a birth order component $\boldsymbol{\iota}'_i\hat{\boldsymbol{\delta}}$, a component coming from age and gender heterogeneity $\mathbf{x}'_{ij}\hat{\boldsymbol{f}}$, and an "unexplained" component $\hat{\varepsilon}_{ij}$, as well as a term reflecting the covariance between birth order and within-family age and gender structure. While birth order effects (and the coefficients on age and gender) came out highly significant in our empirical analysis, we see that the lion's share of within-family variation in years of schooling, almost 98 percent, is "unexplained", or in the language of our model, due to idiosyncratic preference shocks. The within-family variance of birth order effects accounts for 34 percent of the explained within-family variation, with the remainder being accounted for by age and gender heterogeneity, including their covariance with birth order effects.

The third column in Table 9 breaks the between-family variation in years of schooling into a family size component $\boldsymbol{\iota}'_{N_j^*}\hat{\boldsymbol{\beta}}$, a component coming from between-family variation in age and gender structures $\mathbb{E}\{\mathbf{x}'_{ij}\hat{\boldsymbol{g}}|j\}$, a component coming from heterogeneity in parent's age and education $\mathbf{w}'_j\hat{\boldsymbol{g}}$, a "family effect" component $\hat{\xi}_j$, and the covariances amongst these components (some of which are zero by construction).[19] Due to the endogeneity of family size in relation to the family effect $\hat{\xi}_j$, it is difficult to interpret the between-family variance decomposition. However, we note that the variance of the family effect accounts for 77 percent of between-family variation in years of schooling. The second largest contribution comes from the variance in parent's age and education which accounts for 19 percent. In comparison, family size accounts for only 0.2 percent of between-family variation in years of schooling.

6 Conclusions

We study a model of fertility choice in which parents allocate resources differentially according to the order of birth of a child, their *birth order*. The objective of the theory is to provide testable predictions on the effect of family size on the children's distribution of human capital within the household. We then outline an empirical strategy that identifies the effect of family size on the intra-household distribution of human capital separately from the effect that birth order may have on a child's education, and test these predictions using a population-wide comprehensive administrative panel dataset from Denmark. Specifically, we implement a two-step empirical strategy which estimates birth order effects controlling for family-fixed effects (first step), and after netting out the estimated birth order effects from the child's education, allows us to recover the effect of an exogenous change in family size using an instrumental variable approach (second step).

Contrary to previous literature, our results suggest that both birth order and family size affect years of education. Our results are robust to a variety of specifications. Our identification strategy is based on our economic model. In this way, we demonstrate the importance of linking the empirical strategy to economic theory.

[19] We assume that $\mathbf{w}'_j$ is strictly exogenous in relation to ξ_j. This restriction is imposed in the estimation at the family level. The decomposition of between-family variation in Table 9 relates to the family size weighted distribution of family level components. The weighting induces covariance between $\mathbf{w}'_j$ and ξ_j.

References

[1] Anh, T.S., Knodel, J., Lam, D., and Friedman, J. (1998) "Family Size and Children's Education in Vietnam," *Demography*, Vol. 35, 57-70.

[2] Amir, R. (1996) "Sensitivity Analysis of Multisector Optimal Economic Dynamics," *Journal of Mathematical Economics*, Vol. 25(1), 123-141.

[3] Angrist, J., Lavy, V., and Schlosser A. (2010) "Multiple Experiments for the Causal Link between the Quantity and Quality of Children," *Journal of Labor Economics*, Vol. 28, 773-824.

[4] Barro, R. and Becker, G. (1989), "Fertility Choice in a Model of Economic Growth," *Econometrica*, Vol. 57, 481-501.

[5] Becker, G. (1960) "An Economic Analysis of Fertility," in Coale et al. **Demographic and Economic Change in Developed Countries**, Princeton University Press, 209-231.

[6] Becker, G. and Lewis, H.G. (1973) "On the Interaction between the Quantity and Quality of Children," *Journal of Political Economy*, Vol. 81, S279-S288.

[7] Becker, G., Murphy, K.M., and Tamura, R. (1990) "Human Capital, Fertility, and Economic Growth," *Journal of Political Economy*, Vol. 98, S12-S37.

[8] Becker, G. and Tomes, N. (1976) "Child Endowments and the Quantity and Quality of Children,"*Journal of Political Economy,* Vol. 84, Part 2: Essays in Labor Economics in Honor of H. Gregg Lewis**,** S143-S162.

[9] Bedard, K. and Dhuey, B. (2006) "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects," *Quarterly Journal of Economics,* Vol. 121, 1437-1472.

[10] Behrman, J. R., and Taubman, P. (1986) "Birth Order, Schooling, and Earnings," *Journal of Labor Economics*, Vol. 4(3), S121-S145.

[11] Behrman, J.R., Pollak, R.A., and Taubman, P. (1989) "Family Resources, Family Size, and Access to Financing for College Education," *Journal of Political Economy*, Vol. 97, 398-419.

[12] Black, S.E., Devereux, P.J., and Salvanes, K.G. (2005) "The More the Merrier? The Effect of Family Composition on Children's Education," *Quarterly Journal of Economics*, Vol. 120, 669-700.

[13] Brock W. (1971) "Sensitivity of Optimal Growth Paths with Respect to a Change in Target Stocks," in **Contributions to the von Neumann Growth Model**, by Bruckmann, G. and Weber, W. (Eds.), Springer Verlag, 73–89.

[14] Doepke, M. (2004) "Accounting for Fertility Decline During the Transition to Growth," *Journal of Economic Growth*, Vol. 9, 347-383.

[15] Galor, O. and Weil, D. (2000) "Population, Technology and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond," *American Economic Review*, Vol. 90, 806-828.

[16] Glick, P., Marini, A., and Sahn, D.E. (2007) "Estimating the Consequences of Changes in Fertility on Child Health and Education in Romania: An Analysis Using Twins Data," *Oxford Bulletin of Economics and Statistics*, Vol. 69, 667-691.

[17] Gourieroux, C., Holly, A., and Monfort, A. (1982) "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica,* Vol. 50(1), 63-80.

[18] Hanushek, E. (1992) "The Trade-off between Child Quantity and Quality," *Journal of Political Economy*, Vol. 100, 84-117.

[19] Hausman, J. A., and Taylor, W. E. (1981) "Panel Data and Unobservable Individual Effects," *Econometrica,* Vol. 49(6), 1377-1398.

[20] Hestenes, M.R. (1966) **Calculus of Variations and Optimal Control Theory**, Wiley.

[21] Horowitz, J.L. (2001) "The Bootstrap," in: Heckman, J.J., Leamer, E.E. (Eds.), **Handbook of Econometrics**, Vol. 5, 3159-3228.

[22] Kodde, D.A. and Palm, F.C. (1986) "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica* , Vol. 54(5), 1243-1248

[23] Li, H., Zhang, J., and Zhu, J. (2008) "The Quantity-Quality Trade-off of Children in a Developing Country: Identification using Chinese Twins," *Demography*, Vol. 45, 223-243.

[24] Milgrom, P. and Shannon, C. (1994) "Monotone Comparative Statics," *Econometrica*, Vol. 62(1), 157-180.

[25] Millet, D.L. and Wang, L. (2005) "Is the Quantity-Quality Trade-off Really a Trade-off for All?," Unpublished Manuscript, Department of Economics, Southern Methodist University.

[26] Mogstad, M. and Wiswall, M. (2011) "Testing the Quantity-Quality Model of Fertility: Estimation using Unrestricted Family Size Models," Working Paper.

[27] Mogstad, M. and Wiswall, M. (2012) "Instrumental variables estimation with partially missing instruments," *Economics Letters*, Vol. 114, 186-189.

[28] Parish, W. and Willis, R. (1993) "Daughters, Education, and Family Budgets: Taiwan Experiences," *Journal of Human Resources*, Vol. 28, 863-898.

[29] Qian, N. (2006) "Quantity-Quality: The Positive Effect of Family Size on School Enrollment in China," Brown University, Mimeo.

[30] Rosenzweig, M.R., and Wolpin, K.I. (1980) "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, Vol. 48 227-240.

[31] Rosenzweig, M.R., and Zhang, J. (2009). "Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's "One-Child" Policy," *The Review of Economic Studies*, Vol. 76, 1149-1174.

[32] Schultz, T.P. (1997) "Demand for Children in Low Income Countries," in **Handbook of Population and Family Economics**, by Rosenzweig, M. and Stark, O. (Eds.), North Holland, 349-430.

[33] Stock, J.H., and Yogo, M. (2002) "Testing for Weak Instruments in Linear IV Regression," NBER Technical Working Paper 0284, National Bureau of Economic Research, Inc.

[34] Tsuya, N. O. (2003) "Fertility and Family in Nordic Countries, 1960-2000," *Journal of Population Problems*, Vol. 59, no. 1, pp. 49-80.

[35] Vincent, T.L. and Brusch, R.G. (1970) "Optimal Endpoints," *Journal of Optimization Theory and Applications*, Vol. 6(4), 299-319.

Table 1: Descriptive statistics—Analysis data

	Mean	Standard deviation	Minimum	Maximum
Age in 2006	38.2	7.5	25	64
Female	0.48	0.5	0	1
Education	12.9	2.8	7	20
Mother's education	10.2	3.5	7	20
Father's education	11.0	3.5	7	20
Mother's age in 2006	64.4	8.2	42	96
Father's age in 2006	67.3	8.7	40	96
Number of siblings	1.7	0.9	1	17
Twins in family	0.01	0.11	0	1

Note: Descriptive Statistics are from our analysis data consisting of 1,256,031 children from 581,159 families. Twins and single children are excluded from the data.

Table 2: Number of children in the family

Number of children	Frequency	Percentage
1	160,382	21.6
2	371,978	50.2
3	158,861	21.4
4	39,125	5.3
5	8,217	1.1
6+	2,978	0.4

Note: Descriptive statistics are obtained using 741,541 families including single child families but excluding families with only twins.

Table 3: Average education by family size and birth order

	Education	Mother's education	Father's education	Fraction with < 12 years	Fraction with 12 years	Fraction with > 12 years	Observations
				Family size			
1	12.6	10.1	10.8	0.20	0.47	0.33	160,382
2	13.0	10.6	11.3	0.15	0.48	0.37	666,977
3	12.9	10.0	10.9	0.18	0.46	0.36	406,081
4	12.5	9.3	10.1	0.24	0.45	0.32	131,969
5	11.9	8.5	9.3	0.32	0.43	0.26	34,723
6+	11.5	8.0	8.8	0.39	0.40	0.21	16,281
				Birth order			
1	12.9	10.4	11.1	0.17	0.46	0.37	710,459
2	12.8	10.3	11.0	0.17	0.48	0.35	493,616
3	12.6	9.6	10.5	0.20	0.47	0.32	162,243
4	12.3	8.8	9.7	0.26	0.46	0.28	38,274
5	11.9	8.1	8.9	0.32	0.45	0.23	8,462
6+	11.6	7.8	8.6	0.37	0.44	0.19	3,359

Note: Descriptive Statistics are from our analysis data consisting of 1,416,413 children from 741,541 families. Single children are included. Twins are excluded from the data.

Table 4: Family size, birth orders and children's education—OLS regressions

Dependent variable: Child's education	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Family size	−0.268***	−0.186***	−0.204***	−0.173***	−0.040***	−0.066***		
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)		
Birth order		−0.166***			−0.307***		−0.178***	
		(0.004)			(0.004)		(0.006)	
Birth order indicators								
Second			−0.235***			−0.385***		−0.275***
			(0.005)			(0.005)		(0.007)
Third			−0.301***			−0.610***		−0.426***
			(0.008)			(0.009)		(0.014)
Fourth			−0.459***			−0.834***		−0.496***
			(0.014)			(0.016)		(0.022)
Fifth or later			−0.627***			−1.111***		−0.444***
			(0.026)			(0.027)		(0.033)
Demographic controls	No	No	No	Yes	Yes	Yes	Yes	Yes
Family fixed effecs	No	No	No	No	No	No	Yes	Yes
Observations	1,256,031	1,256,031	1,256,031	1,256,031	1,256,031	1,256,031	1,256,031	1,256,031

Note: *** indicates statistical significance at the 1 percent level. Standard errors (in parentheses) are clustered at the family level. All regressions include indicators for age and sex. Demographic controls include indicators for mother's education, mother's age, father's education, and father's age. Single child families are excluded from the sample.

Table 5: Family size, birth orders and children's' education—2SLS regressions

	(1)	(2)	(3)	(4)
	OLS	First stage	2SLS	Obs.
Instrument: Twin at second birth		0.765***		550,077
		(0.007)		
(Sample of first child in families with 2 or more births)				
Family size	−0.090***		−0.024	
	(0.005)		(0.047)	
Instrument: Twin at third birth		0.846***		376,716
		(0.010)		
(Sample of first and second children in families with 3 or more births)				
Family size	−0.184***		−0.095*	
	(0.008)		(0.053)	
Second child	−0.338***		−0.346***	
	(0.009)		(0.010)	
Instrument: Twin at fourth birth		0.908***		132,878
		(0.030)		
(Sample of first, second, and third children in families with 4 or more births)				
Family size	−0.230***		0.016	
	(0.014)		(0.088)	
Second child	−0.283***		−0.301***	
	(0.017)		(0.019)	
Third child	−0.571***		−0.614***	
	(0.023)		(0.027)	

Note: *** indicates statistical significance at the 1 percent level. * indicates statistical significance at the 10 percent level. Standard errors (in parentheses) are clustered at the family level. All regressions include indicators for age and sex. Demographic controls include indicators for mother's education, mother's age, father's education, and father's age. Single child families are excluded from the sample.

Table 6: Two-step estimation with linear family size and birth order profiles

	(1)	(2)	(3)	(4)	(5)
	Step 1: OLS w/ family fixed effects	Step 2: OLS	Step 2: OLS	Step 2: 2SLS	Step 2: 2SLS
Dependent variable:	Child's years of education	Family-level average years of education[1]	Family-level average years of education[1]	Family-level average years of education[1]	Family-level average years of education[1]
Family size		-0.240^{***} (0.003)	-0.163^{***} (0.003)	-0.102^{***} (0.029)	-0.115^{***} (0.026)
Birth order	-0.178^{***} (0.006)				
First stage:					
Minimum eigenvalue statistic[2]				$14,049.4$	$14,573.8$
Demographic controls		No	Yes	No	Yes
Propositions 2 and 3 (test statistic D)[3]		Not rejected [D=0.000]	Not rejected [D=0.000]	Not rejected [D=0.000]	Not rejected [D=0.000]
Observations	1,256,031	581,159	581,159	581,159	581,159

Note: *** indicates statistical significance at the 1 percent level. Standard errors for all regressions are computed by block-bootstrapping at the family level (100 repetitions) and are given in soft brackets. The family fixed effect regression in column (1) includes controls for for age and sex. Demographic controls include indicators for mother's education, mother's age, father's education, and father's age. Single child families are excluded from the analysis data.

[1] The family-level average education in columns (2)-(5) is compute according to equation (25), netting out the effects of age and sex as estimated from the family-fixed effect regression.

[2] See Stock and Yogo [33] for critical values. We clearly reject the null hypothesis of weak instruments.

[3] We test at a 5 percent significance level. The decision rule is: Do not reject if $D < 2.706$, reject if $D > 11.911$. The test is inconclusive if $D \in [2.706, 11.911]$ (see Kodde and Palm [22]).

Table 7: Two-step estimation with nonlinear family size and birth order profiles

	(1)	(2)	(3)	(4)	(5)
	Step 1: OLS w/ family fixed effects	Step 2: OLS	Step 2: OLS	Step 2: 2SLS	Step 2: 2SLS
Dependent variable:	Child's years of education	Family-level average years of education[1]	Family-level average years of education[1]	Family-level average years of education[1]	Family-level average years of education[1]
Family size 3		-0.132^{***} (0.005)	-0.102^{***} (0.005)	-0.060 (0.055)	-0.120^{**} (0.047)
Family size 4		-0.441^{***} (0.009)	-0.301^{***} (0.008)	-0.248^{***} (0.071)	-0.292^{***} (0.063)
Family size 5+		-1.058^{***} (0.015)	-0.720^{***} (0.012)	-0.314^{***} (0.107)	-0.298^{***} (0.094)
Second child	-0.275^{***} (0.007)				
Third child	-0.426^{***} (0.014)				
Fourth child	-0.496^{***} (0.022)				
Fifth child or later	-0.444^{***} (0.032)				
First stage:					
Minimum eigenvalue statistic[2]				$3,378.0$	$3,539.8$
Demographic controls		No	Yes	No	Yes
Propositions 2 and 3 (test statistic D)[3]		Not rejected [D=0.000]	Not rejected [D=0.000]	Not rejected [D=0.440]	Not rejected [D=0.085]
Observations	1,256,031	581,159	581,159	581,159	581,159

Note: *** indicates statistical significance at the 1 percent level. ** indicates statistical significance at the 5 percent level. Standard errors for all regressions are computed by block-bootstrapping at the family level (100 repetitions) and are given in soft brackets. The family fixed effect regression in column (1) includes controls for for age and sex. Demographic controls include indicators for mother's education, mother's age, father's education, and father's age. Single child families are excluded from the analysis data.

[1] The family-level average education in columns (2)-(5) is compute according to equation (25), netting out the effects of age and sex as estimated from the family-fixed effect regression.

[2] See Stock and Yogo [33] for critical values. We clearly reject the null hypothesis of weak instruments.

[3] We test at a 5 percent significance level. The decision rule is: Do not reject if $D < 2.706$, reject if $D > 11.911$. The test is inconclusive if $D \in [2.706, 11.911]$ (see Kodde and Palm [22]).

Table 8: Two-step estimation with linear family size and nonlinear birth order profiles

	(1)	(2)	(3)	(4)	(5)
	Step 1: OLS w/ family fixed effects	Step 2: OLS	Step 2: OLS	Step 2: 2SLS	Step 2: 2SLS
Dependent variable:	Child's years of education	Family-level average years of education[1]	Family-level average years of education[1]	Family-level average years of education[1]	Family-level average years of education[1]
Family size		-0.230^{***} (0.003)	-0.161^{***} (0.003)	-0.111^{***} (0.029)	-0.126^{***} (0.026)
Second child	-0.275^{***} (0.007)				
Third child	-0.426^{***} (0.014)				
Fourth child	-0.496^{***} (0.022)				
Fifth child or later	-0.444^{***} (0.032)				
First stage:					
Minimum eigenvalue statistic[2]				$14,049.7$	$14,573.8$
Demographic controls		No	Yes	No	Yes
Propositions 2 and 3 (test statistic D)[3]		Not rejected [D=0.000]	Not rejected [D=0.000]	Not rejected [D=0.000]	Not rejected [D=0.000]
Observations	1,256,031	581,159	581,159	581,159	581,159

Note: *** indicates statistical significance at the 1 percent level. Standard errors for all regressions are computed by block-bootstrapping at the family level (100 repetitions) and are given in soft brackets. The family fixed effect regression in column (1) includes controls for for age and sex. Demographic controls include indicators for mother's education, mother's age, father's education, and father's age. Single child families are excluded from the analysis data.

[1]The family-level average education in columns (2)-(5) is compute according to equation (25), netting out the effects of age and sex as estimated from the family-fixed effect regression.

[2]See Stock and Yogo [33] for critical values. We clearly reject the null hypothesis of weak instruments.

[3]We test at a 5 percent significance level. The decision rule is: Do not reject if $D < 2.706$, reject if $D > 11.911$. The test is inconclusive if $D \in [2.706, 11.911]$ (see Kodde and Palm [22]).

Table 9: Years of schooling: Within- and between variance decomposition

	Total variance	Within-family	Between-family
Years of schooling h^*_{ij}	7.975	2.799	5.176
Within-family components			
$\mathbb{E}\{Var[\boldsymbol{\iota}'_i\hat{\boldsymbol{\delta}}\|j]\}$		0.023	
$\mathbb{E}\{Var[\mathbf{x}'_{ij}\hat{\mathbf{g}}\|j]\}$		0.072	
$\mathbb{E}\{Var[\hat{\varepsilon}_{ij}\|j]\}$		2.732	
$\mathbb{E}\{2 \times Cov[\boldsymbol{\iota}'_i\hat{\boldsymbol{\delta}}, \mathbf{x}'_{ij}\hat{\mathbf{g}}\|j]\}$		-0.028	
Between-family components			
$Var[\mathbb{E}\{\boldsymbol{\iota}'_{N^*_j}\hat{\boldsymbol{\beta}}\|j\}]$			0.011
$Var[\mathbb{E}\{\mathbf{w}'_j\hat{\mathbf{f}}\|j\}]$			0.971
$Var[\mathbb{E}\{\mathbf{x}'_{ij}\hat{\mathbf{g}}\|j\}]$			0.134
$Var[\mathbb{E}\{\hat{\xi}_j\|j\}]$			3.971
$2 \times Cov[\mathbb{E}\{\boldsymbol{\iota}'_{N^*_j}\hat{\boldsymbol{\beta}}\|j\}, \mathbb{E}\{\mathbf{w}'_j\hat{\mathbf{f}}\|j\}]$			0.014
$2 \times Cov[\mathbb{E}\{\boldsymbol{\iota}'_{N^*_j}\hat{\boldsymbol{\beta}}\|j\}, \mathbb{E}\{\hat{\xi}_j\|j\}]$			0.008
$2 \times Cov[\mathbb{E}\{\mathbf{w}'_j\hat{\mathbf{f}}\|j\}, \mathbb{E}\{\hat{\xi}_j\|j\}]$			0.067

Note: The variance decomposition is done for the estimates reported in Table 7. $\mathbf{x}$ contains indicators for age and sex of the child. $\mathbf{w}$ contains indicators for mother's education, mother's age, father's education, and father's age.

Appendix A: Omitted proofs

Proof of Proposition 3. For convenience, we write $\partial u(h^*(i), i)/\partial h$ simply as $u_h^*(i)$. Likewise, we write the second derivatives evaluated at the optimum as $u_{hi}^*(i)$ and $u_{hh}^*(i)$.

For any value of z, including $z = 0$, optimal choices for $h(i)$ satisfy (4) with a Lagrange multiplier $\lambda(z)$. Differentiating this expression yields

$$u_{hh}^*(i)\frac{\partial h^*(i|N^*+z)}{\partial z} = \frac{\partial\lambda(z)}{\partial z}p(i) = \frac{\partial\lambda(z)}{\partial z}\frac{u_h^*(i)}{\lambda(z)}. \tag{1}$$

Expressions (2) and (4) yield

$$\lambda(z) = \int_0^{N^*+z}\frac{u_h^*(i)h^*(i)}{Y}di. \tag{2}$$

Differentiating (2), and using (1), yields

$$\frac{\partial\lambda(z)}{\partial z} = \left.\frac{u_h^*(i)h^*(i)}{Y}\right|_{i=N^*+z} + \frac{\partial\lambda(z)}{\partial z}\int_0^{N^*+z}\left\{\frac{u_h^*(i)h^*(i)}{\lambda(z)} - \frac{u_h^*(i)^2}{[-u_{hh}^*(i)]}\frac{1}{\lambda(z)}\right\}\frac{di}{Y},$$

which, using (2) and (4), can be written as

$$\frac{\partial\lambda(z)}{\partial z}\int_0^{N^*+z}\frac{u_h^*(i)^2}{[-u_{hh}^*(i)]}di = \lambda(z)u_h^*(N^*+z)h^*(N^*+z). \tag{3}$$

Expression (3) shows that $\partial\lambda(z)/\partial z > 0$, as long as the budget constraint binds (e.g., $\lambda(z) > 0$) and the parents value the human capital of the "additional children." Substituting (3) into (1), yields $\partial h^*(i|N^*+z)/\partial z < 0$ all $i \in [0, N^*+z]$; an alternative proof of Proposition 2.

Next, differentiating $H^*(N+z)$ in (9) yields

$$\frac{\partial H(N^*+z)}{\partial z} = -\frac{1}{N^*+z}\left[H(N^*+z) - h^*(N^*+z) - \int_0^{N^*+z}\frac{\partial h^*(i|N^*+z)}{\partial z}di\right]. \tag{4}$$

Using (3), a necessary condition for (4) to be negative is

$$\int_0^{N^*+z}[u_h^*(i)H(N^*+z) - u_h^*(i)h^*(N^*+z) + u_h^*(N^*+z)h^*(N^*+z)]\frac{u_h^*(i)}{[-u_{hh}^*(i)]}di > 0. \tag{5}$$

If $h^*(i)$ is decreasing in i, then $H(N^*+z) > h^*(N^*+z)$ and (5) is satisfied trivially. This case corresponds to condition (a). Consider next the case in which $h^*(i)$ is increasing in i. Using (4), a sufficient condition for (5) to be satisfied is $u_h(h^*(N^*+z), N^*+z) \geq u_h(h^*(i), i)$. Subtracting $-u_h(h^*(N^*+z), i)$ on both sides, this condition can be written simply as

$$[u_h(h^*(N^*+z), N^*+z) - u_h(h^*(N^*+z), i)] \geq -[u_h(h^*(N^*+z), i) - u_h(h^*(i), i)], \tag{6}$$

or as $u_{hi}(h, i) \geq -u_{hh}(h, i)$ for all (h, i), which is condition (b). ■

Appendix B: General model [not for publication]

Here we consider a more general allocation problem and show that the predictions of the basic model are robust to these extensions.

Let C denote parental consumption. Assume parents value their consumption according to a sub-utility function $U^0(C)$. As in Barro and Becker [4], consider a nonlinear aggregation over the children's sub-utilities. The upper limit of integration would then be given by a function $\alpha(N)$ in (1). We also assume that there is a parental valuation for family size that is independent of human capital. This utility component is $v(N)$. Further, as in Behrman et al. [10], we consider a series of CES aggregator for the sub-utilities of parents and children.

The general parental utility function is

$$U(C,N,h) \equiv \left\{U^0(C)^\varphi + U^1(N,h)^\varphi\right\}^{1/\varphi}, \tag{7}$$

where

$$U^1(N,h) \equiv \left\{ v(N)^\eta + \left[\int_0^{\alpha(N)} u(h(i),i)^\rho di\right]^{\eta/\rho} \right\}^{1/\eta}.$$

Substitution takes place along multiple dimensions: $-\infty < \rho \leq 1$ measures the degree of substitutability between parental valuations across children at different birth orders, $-\infty < \eta \leq 1$ measures the degree of substitutability between parental investments in human capital and family size, and $-\infty < \varphi \leq 1$ measures the degree of substitutability between the parental utility from consumption and the parental utility from investments in children. The function $u(h(i),i)$ can be interpreted as in the text.

Parents maximize the utility function $U(C,N,h)$ subject to the following budget constraint

$$qC + \pi N + \int_0^N p(i)h(i)di \leq I, \tag{8}$$

where I is total parental income, q is the cost of parental consumption, and π is a fertility cost independent of human capital.

Let λ denote the Lagrange multiplier of (8). The first order condition for parental consumption C is

$$\left(\frac{U_C^0(C)}{U(C,N,h)}\right)^{\varphi-1} = \lambda q, \tag{9}$$

the first order condition for human capital $h(i)$ is

$$\left(\frac{U^1(N,h)}{U(C,N,h)}\right)^{\varphi-1} \left(\frac{[U^1(N,h)^\eta - v(N)^\eta]^{(\eta-\rho)/\eta}}{U^1(N,h)^{\eta-1}}\right) u(h(i),i)^{\rho-1} u_h(h(i),i) = \lambda p(i), \tag{10}$$

and the first order condition for family size N is

$$\left(\frac{U^1(N,h)}{U(C,N,h)}\right)^{\varphi-1} \left(\frac{v(N)}{U^1(N,h)}\right)^{\eta-1} \left[v_N(N) + \frac{u(h(\alpha(N)),\alpha(N))}{\eta v(N)^{\eta-1}}\right] \alpha_N(N) = \lambda p(N)h(N). \tag{11}$$

Notice that the model in the text assumes $\rho = \eta = \varphi = 1$, $U^0(C) = v(N) = \pi = q = 0$, and $\alpha(N) = N$. Because there are multiple margins of substitution, the economic interpretation of the first order conditions in the general model is much more involved than those in the basic model. In (10), for example, it is necessary to recognize the substitution between parental utility from their own consumption and the utility from their children, as well as the substitution between family size and the overall value given to human capital investments.

Let N^* denote the optimal family size and let $h^*(i)$ denote the human capital profile in the general model. Despite the additional margins of substitution in the general model, the analysis of this problem can be undertaken in essentially the same manner as in the basic model. First, we characterize the human capital profiles treating N^* as a parameter, and then we examine the response to exogenous changes in N^*. From, (10) we can write

$$\frac{\partial h^*(i)}{\partial i} = \tilde{\Delta}(h^*(i), i), \text{ for all } i \in [0, N^*], \tag{12}$$

with

$$\tilde{\Delta}(h^*(i), i) \equiv \frac{u_h^*(i)\,(p_i(i)/p(i)) - [(\rho-1)u^*(i)u_h^*(i)u_i^*(i) + u_{hi}^*(i)]}{(\rho-1)u^*(i)u_h^*(i)^2 + u_{hh}^*(i)},$$

such that when $\rho = 1$, we have $\tilde{\Delta}(h, i) = \Delta(h, i)$ given in the basic model. Expression (12) also represents a non-autonomous differential equation.

It is possible to see that we can obtain a characterization analogous to Proposition 1. That, is the human capital profiles in the general model can have any general shape. An exogenous change in family size can also be studied as in the basic model. From a parental perspective, such an increase is equivalent to a reduction in income. Notice that $\pi(N^* + z) - \pi N^*$ represents a lower bound of the additional expenses due to an increase in fertility from N^* to $N^* + z$. If parental consumption is unchanged, an exogenous increase in family size must lower total spending in children. If parents lower their consumption in response to an increase in fertility, it is possible to see an increase in parental spending following an exogenous increase in family size. For this case to take place, the income elasticity of parental consumption must be larger than one. In other words, parental consumption C must be a luxury.

Notice also that if $p(i) = p$ for all i, then we can write the budget constraint (8) as $qC + \pi N + pNH \leq I$, which is the same budget constraint as in Becker and Lewis [6].

Appendix C: Detailed data description [not for publication]

This appendix documents the construction of the raw IDA data. For added readability, and to keep this appendix self-contained, some of the information provided here is also included in the main text. This appendix also contains a brief description of the mapping of Statistic Denmark's education codes into years of education.

IDA. Our empirical analysis utilizes the person files of IDA, a population-wide comprehensive Danish administrative panel dataset. IDA is constructed and maintained by Statistics Denmark and contains annual observations on all individuals (identified via a unique and time-invariant person ID) aged 15-70 and residing in Denmark with a valid social security number. For cohorts born in 1960 or later, we are able to link parents and children through a unique person ID.[20] Furthermore, IDA contains detailed individual-level information on an array of relevant economic and socioeconomic characteristics. These include gender, educational attainment, and information on the person ID of an individual's cohabiting partner (if any) as well as a variable measuring the number of children aged 0-14 in the household where the individual is residing. Information on educational attainment comes in the form of information on highest completed education. We transform this measure into years of education using the first two digits of an eight-digit classification code of Danish educations.

Years of education is the main outcome variable of this paper and we offer some details on the construction of this variable here. The IDA person files contains information on each individual's highest completed education. This information is contained in an official 8-digit classification code of Danish educations, denoted `HFFSP`.[21] The first two digits of the `HFFSP` code identifies the level of each education within the Danish educational system. Table A.1 present the mapping of 2-digit `HFFSP` codes to education lengths, including a general description of each of the education levels.

At the onset, the population-wide annual IDA person files consists of 101,720,769 observations on 5,808,600 individuals (aged 15-70 at some point during 1980-2006). We supplement IDA with data on the precise date of birth obtained from Statistics Denmark (usually the IDA person files carries only information on the year of birth). Precise information on birth data is essential for correctly identifying twin births.[22]

The construction of the analysis sample goes over three steps. First, we identify potential mothers and fathers from the population dataset and record relevant information. Second, we identify children, whose outcomes in terms of education is the main object of interest in this study, and merge information on parents onto the children dataset. Third, we impose a number of selection criteria on the children dataset.

[20] Up until 1978 it was common practice to delete information linking a person with his/her parents in the family register when an individual turned 18. From 1978 onwards, the practice changed, and effectively, almost all individuals born in 1960 or later have a pointer to their parents' person ID, *source*: Statistics Denmark. The data confirms that this is in fact the case, with almost every individual born in 1960 or later having valid pointers to their mother's person ID, whereas for individuals born prior to 1960, the fraction with valid information on the ID of their mothers is much lower and declining in age.

[21] Detailed documentation for this variable (in Danish) can be found on Statistic Denmark's website.

[22] The merging of the IDA person files and the date-of-birth data resulted in 69.40 percent of all individuals in the IDA person files being assigned a precise date of birth. The unmatched individuals predominantly belong to old cohorts, which is unproblematic for our analysis as it relies on identifying twin births among the younger cohorts (whose parents are also present in the IDA data). We retain both matched and unmatched individuals in the data as the unmatched observations (on older individuals) may contain information on parents.

Mothers and fathers.– The set of potential mothers contains all women in the population data. Likewise, the set of potential fathers is made up of all men in the population data. For each potential mother we record the following information: age, years of education (measured in the last year the mother appears in the population data), the number of children aged 0-14 in her household in 2006 (the year in which we measure outcomes), the mother's number of siblings (if available), and her income in each of the years she appears in the population data. We also retain information on the person ID of cohabiting partners in each of the years the mother appears in the population data.[23] We record the same information for potential fathers, except that we do not retain information on the number of children aged 0-14 in the father's household in 2006, and do not retain information on the father's cohabiting partner.[24]

Children.– The set of children contains all individuals in the population data that can be merged with information on their mothers (i.e., it requires a non-missing pointer to the mother's person ID). In total 46,724,171 observations on 2,600,878 individuals distributed over 1,308,906 families are selected. We delete 8,783 families (i.e., sets of siblings born to the same mother) where at least one of the children have invalid information on their precise date of birth. Likewise, information on fathers is merged onto the panel of children. Notice, however, that we retain all children independently of whether they are matched to their father.

At this stage we compute family size defined as the number of siblings born to the same mother. Using information on the precise date of birth we also identify twins as well as the birth order of each child within their family. Formally, we use multiple births as an instrument, but denote it twin births.[25] Finally, we retain only the 2006 cross section of children (i.e., all outcome measurements are conducted in 2006). Parts of our analysis makes use of various parental income measures and we trim the distribution of each of these measures by recoding observations in the 1st and 99th percentiles as missing while retaining the observation in the dataset. Before selection of the analysis data, the panel of children contains 2,453,843 individuals and 1,272,874 families.[26]

Analysis data. A brief description of the selection criteria and the resulting analysis data is included in Section 4. Starting from the set of children in IDA, defined as the 2,453,843 individuals (distributed over 1,272,874 families) who can be matched with a mother in IDA,[27] we impose the following selection criteria to arrive at our analysis data:

- We exclude 6,104 families where the mother was below 17 or above 49 when giving birth.[28]
- We exclude 188,954 families in which at least one member (a child or one of the parents) has missing education data.

[23] This allow us to later condition the analysis on the father being the mother's cohabiting partner.

[24] We define families by the mother's identity, and so, it is the composition of the mother's household that is the relevant conditioning variable.

[25] At this stage of the sample selection process, the data contained 24,408 multiple births: 24,186 sets of twins, 217 sets of triplets, and 5 sets of quadruplets.

[26] Our data is roughly consistent with official statistics related to cohort sizes and the frequency of twin births available on Statistic Denmark's website `http://www.statistikbanken.dk/` (in Danish) .

[27] In other words, the set of children in IDA consist of all individuals in IDA with a valid pointer to the mother's ID, who's mother was alive, aged 15-70 and present in IDA at some point during 1980-2006. There are 2,453,843 such individuals.

[28] The age variable used for sample selection indicates age X in the calendar year a person turns X. The youngest possible mother in our sample would be a woman who turns 17 on Dec. 31st in year X and who gives birth on January 1st year X (at age 16). The woman would thus have been above the age of consent at the time of conception.

- We exclude 199 families in which at least one pair of non-twin siblings are recorded as being born less than seven months apart.

- We exclude 87,633 families where the father is unknown for at least one of the siblings or where siblings have different fathers.

- We exclude 118,065 families with children born after 1991 (aged 0-14 in 2006).[29] Below we condition on children being 25 or above. IDs of parents are only sporadically available for cohorts born before 1960.[30] Hence, children born close to 1960 may have older siblings that are not linked to their parents in IDA, inducing measurement errors in both family size and birth order. Our results are robust to discarding families with children born before 1960, 1963 and 1965.

- Only children aged 25 or above in 2006 are retained. We discard 308,893 individuals in this step.

- We exclude 3,904 observations on twins, only retaining an indicator for a twin birth in the family and the order of the twin birth (we lose 2,559 families in this step).

We are left with all individuals in 2006 aged 25-41 who's parents were both alive and present in IDA at some point during 1980-2006, and who satisfies the additional regularity conditions listed above. The analysis data contains 1,256,031 individuals distributed over 581,159 families. The unit of observation is an individual (in 2006). Note that some selection criteria are imposed at the family level while others are imposed at the individual level. Family size and birth order are recorded before the individual level selection criteria are imposed.

[29] Technically, since children are not included in IDA until the year they turn 15, we discard families where at least one child age 0-14 resides on the mother's address in 2006.

[30] Up until 1978 it was common practice in Denmark to delete information linking a person with his/her parents when an individual turned 18. From 1978 onwards, the practice changed, and effectively, almost all individuals born in 1960 or later have a pointer to their parents' person ID (source: Statistics Denmark).

Table A.1: Mapping of highest completed education (`HFFSP`) into years of education

`HFFSP` (2-digit)	Description	Years of education
10 and born 1958 or earlier	Primary education	7 years
10 and born 1959 or later	Primary education	9 years
15,17	Preparatory educations for highschool or equivalent	10 years
20	Highschool (traditional math or language track)	12 years
25	Highschool (technical or business track)	12 years
30	Introductory part of vocational education	10 years
35	Vocational education	12 years
40	Short further education	14 years
50	Medium-length further education	16 years
60	Bachelor degree	15 years
65	Master degree	18 years
70	Ph.D. degree	20 years
90	Unknown $^{(a)}$	missing

The text provides a further description of the different education levels including the distinction between primary school education for individuals born 1958 or earlier and individuals born 1959 or later. (a): Immigrants are overrepresented in this category as education obtained in their home country is seldom converted to Danish standards and recorded.