Education, Birth Order, and Family Size*

Jesper Bagger†
Royal Holloway, University of London and CAP

Javier A. Birchenall‡
University of California at Santa Barbara

Hani Mansour§
University of Colorado Denver and IZA

Sergio Urzúa¶
University of Maryland, NBER and IZA

November 14, 2013

Abstract

We assess, theoretically and empirically, the effect of family size on a child's education when parents have birth order predispositions. Our framework delivers theoretically-consistent birth order and family size effects. We show that existing empirical strategies confound these effects and do not identify a "quantity-quality trade-off," even when the endogeneity of both birth order and family size are correctly accounted for. Guided by the theory, we develop an empirical strategy to separately identify these effects. Danish administrative data does not reject the theoretical implications of the model. We find significant birth order and family size effects on a child's education.

Keywords: Quantity-quality trade-off, fertility models, fixed-effects, instrumental variables
JEL classification: E20; E24; D52.

*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, Kjell Salvanes, and Daniel Hamermesh.

†Department of Economics, Royal Holloway, University of London, Egham Hill, Surrey TW20 0EX, United Kingdom. E-mail: jesper.bagger@rhul.ac.uk. Bagger gratefully acknowledges financial support from The Cycles, Adjustment, and Policy research unit, CAP, Department of Economics and Business, Aarhus University.

‡Department of Economics, 2127 North Hall, University of California, Santa Barbara CA 93106. Phone/fax: (805) 893 5275. E-mail: jabirche@econ.ucsb.edu.

§Department of Economics, University of Colorado Denver. Address: Campus Box 181, 1380 Lawrence Street, Suite 460, P.O. Box 173364, Denver, CO 80217-3364. E-mail: hani.mansour@ucdenver.edu.

¶Department of Economics, University of Maryland. Address: 3505 Tydings Hall Room 3115C, College Park, MD, 20742. E-mail: urzua@econ.umd.edu. Sergio Urzúa 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.

1 Introduction

Understanding the relationship between family size, family structure, and a child's education has been an ongoing concern for social scientists. The theoretical framework of Becker and Lewis (1973) paved the way for economic studies of the "quantity-quality trade-off." Their framework, under the assumption that parents treat all children equally, predicts that an increase in family size reduces a child's education. This prediction has been applied and/or tested by numerous studies.[1] Recent empirical work has refined the scope of the relationship between education and family size by recognizing that Becker and Lewis' equal treatment assumption overly simplifies how education is distributed within the family, and that birth order effects in educational attainment are empirically important.[2] This refinement on the empirical front has not been guided by a theoretical framework that recognizes within-family differences in education motivated by birth order predispositions. Our first contribution is to provide such a framework.

Family size and birth order are jointly determined. An increase in family size necessarily introduces new (e.g., high) birth orders into the family. Thus, under birth order predispositions, changes in family size manifest themselves into changes in birth order. We define the structural (or causal) family size effect as the difference in average educational attainment between otherwise identical families of different sizes. Changes in this household average represent changes in the education of a "typical" child in the family and are consistent with largely unrestricted birth order predispositions. We present conditions on parental preferences and the cost of human capital under which a "quantity-quality trade-off," i.e., a negative family size effect, arises. Likewise, we define structural birth order effects as the difference in the educational attainment

[1]The literature on family size and education is too large to provide a detailed account here. Some of the empirical studies that examine the relationship between education and family size include Ahn, Knodel, Lam, and Friedman (1998), Angrist, Lavy, and Schlosser (2010), Behrman, Pollak, and Taubman (1989), Glick, Marini, and Sahn (2007), Hanushek (1992), Li, Zhang, and Zhu (2008), Millet and Wang (2011), Parish and Willis (1993), Qian (2009), Schultz (1997), and Juhn, Rubinstein, and Zuppann (2013). These studies typically abstract from birth order effects. From a theoretical perspective, Barro and Becker (1989), Becker, Murphy, and Tamura (1990), Becker and Tomes (1976), Doepke (2004), and Galor and Weil (2000) study family size and education in models of economic growth.

[2]Black, Devereux, and Salvanes (2005) is one of the most detailed economic studies of birth order effects. Many previous studies lack large representative data and this has undermined their results; see, e.g, Booth and Kee (2009), Behrman and Taubman (1986), Hanushek (1992), Hauser and Sewell (1985), and Iacovou (2008) . These previous studies provide a wide range of estimates for the effect of birth order, mostly imprecisely estimated. They are unable to control for family size indicators, indicators for children's cohorts or for parental cohorts. An exception is Iacovou (2008) but the sample is small and the estimates are subject to considerable attrition bias. There is also a large birth order literature in psychology (for a review see Eckstein, Aycock, Sperber, McDonald, Wiesner, Watts, and Ginsburg (2010))

of children within the same family who differ only by birth order.

The existing empirical literature does not separately identify these structural family size and birth order effects. Black, Devereux, and Salvanes (2005), for example, used an instrumental variable strategy to estimate the effect of family size while controlling for birth order. We find that this strategy confounds the structural family size and birth order effects. Specifically, if parents favor children with low birth orders, 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 structural family size effect.[3] This false rejection of the "quantity-quality trade-off" occurs even when the endogeneity of both birth order and family size are properly accounted for.

Guided by the theory, we introduce and implement a simple two-step empirical strategy that identifies the structural family size effect separately from the effect of birth order. This is the second contribution of the paper. In the first step, the structural birth order effects can be consistently estimated using within-family variation in educational attainment. In the second-step, the structural family size effect is consistently estimated by instrumental variable techniques using between-family variation in average educational attainment. We use the event of a twin birth as an instrumental variable for family size.

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, an additional birth order reduces years of schooling by little less than one-fifth of a year. Findings are similar in magnitude when we allow for nonlinear birth order effects. Family size, once carefully analyzed in the light of the theory, has a strong negative effect on the average education in the household. 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, our findings support the existence of a trade-off between the quality and the quantity of children even when children in the household are not equally treated. This is our third contribution to the literature.

[3]In the special case in which parents treat their children identically and so there are no birth order effects, our framework, and hence the structural family size effect, coincides with that of Becker and Lewis (1973).

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 their time allocation). Some general remarks are presented in Appendix A. In order to facilitate the analysis, we assume that family size is a continuous variable and that parents make all decisions in a single stage of choice.[4] Children are born sequentially but within this single stage. The analysis is carried out at the family level.

Let C represent the parents' consumption. Parents value their own consumption according to $V(C)$. Let N represent *family size*. Parents value family size directly according to $U(N)$. 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, $h(i)$, and parents value the human capital at each birth order. In the empirical analysis we measure $h(i)$ by years of education. The parents' utility function is

$$W(C, N, h) \equiv (1 - \theta)V(C) + \varphi U(N) + \theta \int_0^N u(h(i), i)di, \tag{1}$$

with θ and φ as weighting factors, and with $u(h(i), i)$ as the parents' utility of a child of birth order i with human capital $h(i)$.

The parents' budget constraint is

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

with Y representing the parents' total income and $p(i) > 0$ the (given) cost per unit of human capital for a child born at order i. For simplicity, the baseline model assumes no "fixed" cost in terms of family size. We discuss more general cases in Appendix A.

The cost $p(i)$ captures pure pecuniary advantages for some birth orders. The utility function $u(h(i), i)$ captures birth order differences in child endowments, intellectual ability, and parental preferences.[5] Human capital will differ across birth orders as long as $\partial^2 u(h, i)/\partial h \partial i \neq 0$ or

[4] 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 modelled as a continuous or a discrete variable. We later illustrate this in footnote 8.

[5] The literature on the intrahousehold allocation of resources usually relies on an earnings production function

$\partial p(i)/\partial i \neq 0$; see Proposition 1(iii) below. We assume that $(V, U) \in \mathcal{C}^2$ and are strictly concave. Further, $u \in \mathcal{C}^2$ 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.[6] Parental preferences are gender-neutral and there is no child mortality.

Parents maximize (1) subject to (2). Optimal decisions for (N, h) 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* $h^* \in \mathcal{C}^1$. Family size satisfies

$$\varphi \frac{\partial U(N^*)}{\partial N} + \theta u(h^*(N^*), N^*) = \lambda p(N^*) h^*(N^*), \tag{3}$$

and human capital satisfies

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

for all $i \in [0, N^*]$, where λ is the Lagrange multiplier associated with the budget constraint (2). Parental consumption C^* satisfies a standard first-order condition, $(1-\theta)\partial V(C^*)/\partial C = \lambda$.

The first-order conditions (3) and (4) are fairly intuitive. In expression (3), an additional child increases parental utility directly by $\varphi \partial U(N^*)/\partial N$ and indirectly by adding $\theta u(h^*(N^*), N^*)$ to the sum of utilities of the other children. Increasing N, however, requires additional spending; parents must invest $h^*(N^*)$ at a cost $p(N^*)$. In expression (4), the utility gains from human capital investments at birth order i must, at the margin, equal their cost. Combining (3) and (4) yields

$$\frac{1}{p(N^*)} \left[\frac{\varphi}{\theta h(N^*)} \frac{\partial U(N^*)}{\partial N} + \frac{u(h^*(N^*), N^*)}{h^*(N^*)} \right] = \frac{1}{p(i)} \frac{\partial u(h^*(i), i)}{\partial h(i)}, \tag{5}$$

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

$f(h(i), e(i), i)$, where $e(i)$ is a child endowment; see, e.g., Behrman and Taubman (1986). 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, ability, and preferences.

[6] From a variational point of view, the maximization of (1) subject to (2) can be seen as an *isoperimetric* problem; see Hestenes (1966). 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 that $\varphi \partial^2 U(N)/\partial N^2 + \theta \partial u(h,i)/\partial i - h(\partial u(h,i)/\partial h)(\partial p(i)/\partial i)/p(i) < 0$ for all (h,i). This second order condition is standard in variational problems with optimal endpoints; see Vincent and Brusch (1970, Theorem 3.1). As remarked in Appendix A, the analysis can be extended by assuming that parents aggregate the sub-utilities of their children nonlinearly, as in Barro and Becker (1989). Appendix A also strengthens the concavity of $V(C)$ by assuming that its elasticity of substitution is bounded away from unity.

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 regarding (N, h) can be broken into two independent parts. First, family size N^* is determined by (3), and, second, human capital satisfies expression (6) with terminal condition $(h^*(N^*), N^*)$. The following proposition describes the human capital profile.

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 simply 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. As the proof makes clear, however, 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 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 (1973) formulation.

Proposition 1 does not describe the effects of exogenous changes in family size on a child's human capital. The entire empirical literature analyzing the "quantity-quality trade-off" revolves around identifying the effects of exogenous changes in family size. Since family size is endogenous, twin births are commonly used as sources of exogenous variation (i.e., instruments) for family size; see, e.g., Rosenzweig and Wolpin (1980). The conceptual experiment behind this logic is as follows. Choosing family size involves the possibility of multiple births. Suppose that twins are unanticipated and undiversifiable, and denote by $Z \in \{0, z\}$ the occurrence of a twin birth: if there are no twin births, $N^*(Z) = N^*$; if there are twins, $N^*(Z) = N^* + z$ with $z > 0$ representing the exogenous "additional children." Parents plan for their children's human capital satisfy (6); parents plan for family size satisfy (3) but only for families without twins.

Let $\{h^*(i|N^*) : i \in [0, N^*]\}$ be the solution of (6) when family size is N^*. Proposition 2 characterizes how the human capital of 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 is above) the human capital profile of families of size $N^* + z$. That is, $h^*(i|N^*) > h^*(i|N^* + z)$, for all $i \in [0, N^*]$.*

Proof. See Appendix A. ■

Proposition 2 is a direct consequence of the decline in available resources per child due to the increase in family size. To present its main idea, let $C^*(z)$ and $C^*(0)$ denote the optimal parental consumption for parents with and without twins, respectively. Suppose that $C^*(z) = C^*(0)$ though no such assumption is needed for this proposition to hold. Since parental spending in children is constant, the budget constraint (2) implies

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

with the right-hand side representing the value of the human capital of the additional children. Any two human capital profiles should not cross.[7] Thus (8) implies that $h^*(i|N^*) > h^*(i|N^*+z)$ for all existing children $i \in [0, N^*]$. This result, as Appendix A shows, is maintained even though parental consumption declines in response to an exogenous increase in family size.

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 the

[7] To show that the profiles should not cross, recall that 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, 1966, Appendix, Theorem 3.1). Accordingly, $h^*(i|N^*)$ and $h^*(i|N^*+z)$ should not cross (Hestenes, 1966, Appendix, Theorem 4.1). These arguments are similar in spirit to Brock (1971)'s analysis of changes in the planning horizon in the neoclassical growth model. A general proof based on standard comparative statics is given in the Appendix. Our results might be proved by *monotone comparative statics* methods which have generalized Brock's Brock (1971) results; see, e.g., Amir (1996) and Milgrom and Shannon (1994). These methods may be useful to obtain more general results than those presented here.

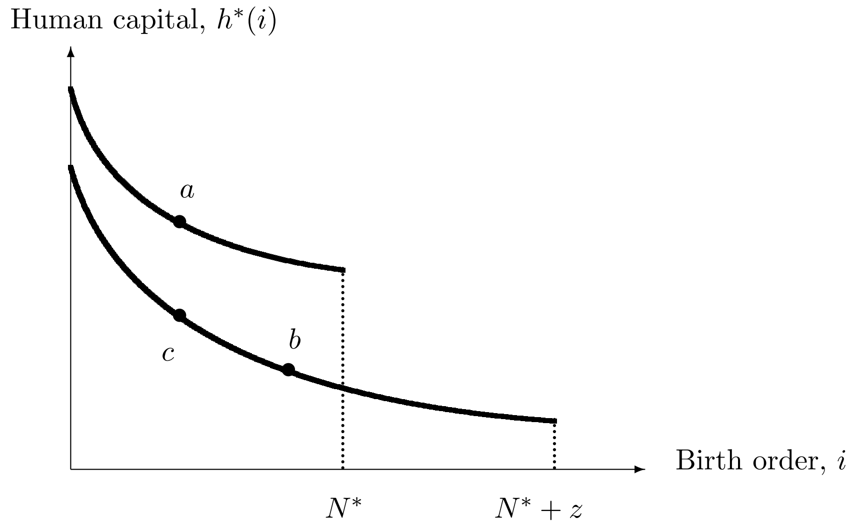

Figure 1: Human capital profiles under negative birth order effects.

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 because the preference and cost differentials subsumed in $\Delta(h, i)$ are independent of the presence of twins. That is, the additional children z 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 parity 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 this proposition 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 increase in family size on the household's average human capital. Let $H(N^*)$ denote 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 next states our general version of the "quantity-quality trade-off" in the presence of birth order effects.

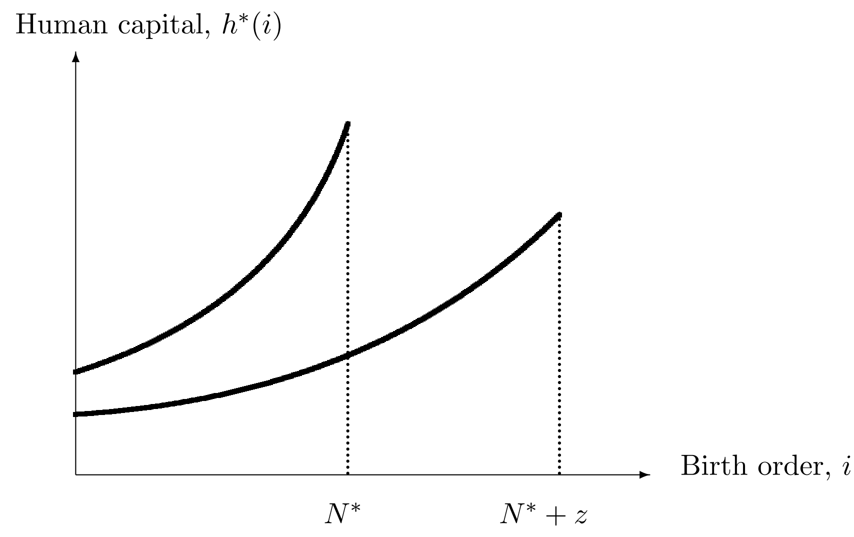

Figure 2: Human capital profiles under positive birth order effects.

Proposition 3 *Consider the following conditions:*

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

(b) $\Delta(h, i) > 0$ *and* $\partial^2 u(h, i)/\partial h \partial i \geq -\partial^2 u(h, i)/\partial h^2 + \bar{c}$*, for some* $\bar{c} > 0$ *and 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 A. ■

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) guarantees that the substitution channel is weaker than the income channel. This condition 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 illustration of our "quantity-quality trade-off" formulation does not require either condition (a) or condition (b) to hold. It relies, however, on special assumptions about the parental utility for consumption and the unit cost of human capital.

Proposition 4 *Suppose that parents are unwilling to substitute their consumption for their children's human capital. Suppose also that $p(i) = p$ for all i. Then, as family size increases exogenously, average human capital decreases nonlinearly with family size.*

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

$$C + pNH \leq Y. \tag{10}$$

The parents' unwillingness to substitute their consumption implies that $V(C) = \min\{C, C^*\}$, with $C^* = C^*(z) = C^*(0)$. Thus (10) implies that $H(N^*) = (Y - C^*)/pN^* > (Y - C^*)/p(N^* + z) = H(N^* + z)$. ■

The parents' unwillingness to reduce their consumption strengthens the income channel. The assumption that all children face the same human capital cost eliminates the substitution channel and forces N and H to enter *multiplicatively* in the budget constraint (10). This product between N and H is central in Becker and Lewis (1973)'s formulation of the "quantity-quality trade-off." The "quantity-quality trade-off" in Propositions 3 and 4, however, is more general than the conventional framework in Becker and Lewis (1973). 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. Therefore, the "quantity-quality trade-off" in Propositions 3 and 4 allows for a general analysis of birth order effects, which, as we shall see in the next section, are empirically relevant.

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. In Appendix A, we consider the more general aggregators used by Behrman and Taubman (1986) and Barro and Becker (1989). Appendix A also provides some remarks that demonstrate the robustness of the basic theoretical predictions to other 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 (measured as years of education).

To understand the empirical implications of the theory, assume that the parents' utility function $u(h(i), i)$ in (1) takes the single index form

$$u_j(h_{ij}, i) = u_j(h_{ij} - \boldsymbol{\iota}_i'\boldsymbol{\delta} - v_j - \varepsilon_{ij}), \tag{11}$$

where $\boldsymbol{\iota}_i$ is a vector whose i-th entry equals 1 and all other entries equal 0. The dimension of $\boldsymbol{\iota}_i$ is N^+ and $\boldsymbol{\delta} \equiv (\delta_1, ..., \delta_{N+})$ is a vector of birth order preference parameters. We shall refer to $\boldsymbol{\delta}$ as the structural birth order effects, or simply as birth order effects. v_j is a family specific effect in preferences, 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. v_j in (11) is related to the structural family size effects 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):

$$h_{ij}^* = \boldsymbol{\iota}_i'\boldsymbol{\delta} + \mu_{ij} + v_j + \varepsilon_{ij}, \tag{12}$$

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

In order to endow (12) with empirical content, we need to impose further restrictions on μ_{ij}.

Particularly, if $p_{ij} = p_j$ (compare to Proposition 4), 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. The family fixed effect now represents family specific preferences, expenditures, and prices; the human capital equation reads

$$h^*_{ij} = \boldsymbol{\iota}'_i \boldsymbol{\delta} + v_j + \varepsilon_{ij}. \tag{13}$$

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

$$v_j = H(N^*_j) - \boldsymbol{\iota}'_{N^*_j} \bar{\boldsymbol{\delta}}, \tag{14}$$

where $H(N^*_j)$ is family j's average human capital and $\bar{\boldsymbol{\delta}} \equiv (\bar{\delta}_1, ..., \bar{\delta}_{N^+})$ is a vector of the average birth order effects in families of size $1, ..., N^+$. That is, for generic family size N,

$$\bar{\delta}_N = \frac{1}{N} \sum\nolimits_{i=1}^{N} \delta_i, \tag{15}$$

so $\boldsymbol{\iota}'_{N^*_j} \bar{\boldsymbol{\delta}}$ is the average birth order effect in a family of size N^*_j.

The theory is consistent with a nonlinear relationship between H_j and N^*_j; see Proposition 4. Therefore we adopt the following flexible relationship in the empirical analysis,

$$H(N^*_j) = \alpha + \boldsymbol{\iota}'_{N^*_j} \boldsymbol{\beta} + \xi_j, \tag{16}$$

where $\boldsymbol{\beta} \equiv (\beta_1, ..., \beta_{N^+})$ is a vector of structural family size effects. ξ_j is a function of parental spending and preferences. Notice that N^*_j is endogenous in relation to ξ_j. Substituting (14)-(16) into (13) yields

$$h^*_{ij} = \alpha + \boldsymbol{\iota}'_i \boldsymbol{\delta} + \boldsymbol{\iota}'_{N^*_j} (\boldsymbol{\beta} - \bar{\boldsymbol{\delta}}) + \xi_j + \varepsilon_{ij}, \tag{17}$$

for $i = 1, ..., N^*_j$.

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, $H(N^*_j + z) - H(N^*_j) = \beta_{N^*_j+z} - \beta_{N^*_j}$. As we remarked after Proposition 3 (and 4), this effect is associated with our general "quantity-quality trade-off." 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 paper, 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 $\beta_{N_j^*} = \beta N_j^*$. 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 the Becker and Lewis (1973) formulation: $\delta_i = 0$ for all i. In this case, $\bar{\delta}_{N_j^*} = 0$ and only the family size effect is relevant.

The parameters of interest in (17) are the structural birth order effects $\boldsymbol{\delta}$ and the structural family size effects $\boldsymbol{\beta}$. Our empirical work is based on the following regression model:

$$h_{ij}^* = a + \boldsymbol{\iota}_i'\mathbf{d} + \boldsymbol{\iota}_{N_j^*}'\mathbf{b} + \epsilon_{ij}, \tag{18}$$

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 $\mathbf{b} = \boldsymbol{\beta} - \bar{\boldsymbol{\delta}}$, (18) is simply a reparameterization of (17).

3.1 Existing empirical strategies

The overriding issue in an empirical analysis of education, birth order, and family size is a dual endogeneity problem. Our theoretical framework implies that family size depends on parental preferences, spending, and the cost of human capital, all of which are part of the composite family fixed effect. A family's birth order configuration is jointly determined with family size, and thus also correlated with the family fixed effect. It is therefore clear that an Ordinary Least Squares (OLS) estimator of (18) does not yield a consistent estimation of the parameters of interest $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$.

[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 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 later on and in the empirical analysis.

Black, Devereux, and Salvanes (2005) proposed an empirical strategy for dealing with the dual endogeneity problem. Mogstad and Wiswall (2012b) extend the strategy to allow for nonlinear effect of family size. The strategy relies on extracting subsets of the data where birth order is exogenous.[10] Consider the subsample consisting of the first child in families with 2 or more children. The birth order configuration is identical for all families in the subsample, and thus independent of the family fixed effect. Stated formally in terms of (18), $\epsilon_{ij} \perp \boldsymbol{\iota}'_i | (i = 1, N^*_j \geq 2)$. Family size remains endogenous and is dealt with using instrumental variables regression techniques.[11] Denote the resulting estimator by $\widehat{\mathbf{b}}_{\text{IV}}$.

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

$$\operatorname{plim} \widehat{\mathbf{b}}_{\text{IV}} = \mathbf{b} = \boldsymbol{\beta} - \bar{\boldsymbol{\delta}}. \tag{19}$$

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 $b_2 = \beta_2 - \bar{\delta}_2 = 0$; see footnote 9). Hence, $\widehat{\mathbf{b}}_{\text{IV}}$ is a consistent estimator of the causal effect of family size on individual human capital conditional on birth order i.

However, the causal effect of family size on mean individual human capital conditional on birth order i, confounds two separate effects, $b_N = \{\mathbb{E}[H_j|N] - \mathbb{E}[H_j|N = 2]\} + \{\mathbb{E}[h^*_{ij} - H_j|i, N] - \mathbb{E}[h^*_{ij} - H_j|i, N]\} = \beta_N - \bar{\delta}_N$. The first effect β_N is the structual family size effect. It 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 $N = 2$. Our theoretical analysis characterized conditions under which $\beta_N < 0$, i.e., where a "quantity-quality trade-off" exists. 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 because, as discussed in the context of our model, (19) does not take into account that family size and birth order are jointly determined.

Figure 1 illustrates these two separate effects. Point a in Figure 1 depicts the human capital of the "typical" child in families of size N^*_j; point b does the same for families of size $N^*_j + z$. The

[10] In this setup, birth order effects are in fact nuisance parameters in relation to estimating the coefficient on family size. Black, Devereux, and Salvanes (2005) subsequently conduct a comprehensive empirical analysis of birth order effects.

[11] The procedure can of course be implemented for subsamples containing the first and second child in families with 3 or more children, the first, second and third child in families with 4 or more children, and so on.

family size effect is the difference in the human capital of the "typical" children: the vertical distance between a and b. The instrumental variable estimator of Black, Devereux, and Salvanes (2005) evaluates the impact of an increase in family size using points a and c. That is, it does not net out the average birth order effect $\bar{\delta}_{N_j^*}$ for "non-typical" children. Indeed, when birth order effects are negative, $\widehat{\mathbf{b}}_{\mathrm{IV}}$ tends toward zero because the instrumental variable estimator compares the "typical" child in small families (families of size N_j^*) to a relatively educated child in large families (families of size $N_j^* + z$).

Disentangling the structural family size effects $\boldsymbol{\beta}$ from the structural birth order effects $\boldsymbol{\delta}$ is important for several reasons. From a theoretical point of view, the framework we propose has predictions regarding $\boldsymbol{\beta}$ (cf. Proposition 3), and testing these predictions requires an estimator of $\boldsymbol{\beta}$. From a more applied point of view, consider a policymaker with an policy to, say, lower fertility. The effect of such a policy on average human capital in the family depends on $\boldsymbol{\beta}$, but not on $\boldsymbol{\delta}$. The effect of the policy on human capital inequality depends on both $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$. In fact, within-family inequality in human capital depends on $\boldsymbol{\delta}$, but not on $\boldsymbol{\beta}$, whereas between-family inequality depends on $\boldsymbol{\beta}$, but not on $\boldsymbol{\delta}$. We return to this issue below.

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 $\boldsymbol{\beta}$. Our strategy draws its motivation from the decomposition of the variance of individual human capital. In particular, using expression (17), the variance of individual human capital can be written as

$$Var[h_{ij}^*] = \mathbb{E}\{Var[\boldsymbol{\iota}_i'\boldsymbol{\delta} + \varepsilon_{ij}|j]\} + Var[\boldsymbol{\iota}_{N_j^*}'\boldsymbol{\beta} + \xi_j]. \tag{20}$$

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.[12] We

[12]The use of within and between sources of variation is loosely reminiscent of Hausman and Taylor (1981). In a linear panel data model, Hausman and Taylor (1981) propose an instrumental variable estimator that separately exploits 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

next apply this simple logic to overcome the dual endogeneity problem of birth order and family size.

In the empirical analysis we augment (18) 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 (18). The within-family estimator is equivalent to the family fixed effect estimator of the birth order effects. Applying the within-family transformation to (18) produces the following estimation equation:

$$h_{ij}^* - H(N_j^*) = \left[\boldsymbol{\iota}_i - \frac{1}{N_j^*}\sum_{i=1}^{N_j^*} \boldsymbol{\iota}_i\right]' \mathbf{d} + \left[\mathbf{x}_{ij} - \frac{1}{N_j^*}\sum_{i=1}^{N_j^*} \mathbf{x}_{ij}\right]' \mathbf{g} + [\epsilon_{ij} - \xi_j], \qquad (21)$$

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

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 (18) at the family level yields

$$H(N_j^*) = a + \left[\frac{1}{N_j^*}\sum_{i=1}^{N_j^*} \boldsymbol{\iota}_i\right]' \mathbf{d} + \boldsymbol{\iota}_{N_j^*}' \boldsymbol{b} + \left[\frac{1}{N_j^*}\sum_{i=1}^{N_j^*} \mathbf{x}_{ij}\right]' \mathbf{g} + \mathbf{w}_j' \mathbf{f} + \xi_j, \qquad (22)$$

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

variables.

"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 $\bar{\boldsymbol{d}} \equiv (\bar{d}_1, ..., \bar{d}_{N^+})$ with generic element $\bar{d}_N = (1/N)\sum_{i=1}^{N} d_i$, and that an estimate of $\mathbf{g}$ is available from (21). Rewrite (22) as

$$\left(H(N_j^*) - \left[\frac{1}{N_j^*}\sum_{i=1}^{N_j^*}\mathbf{x}_{ij}\right]'\hat{\mathbf{g}}\right) = a + \boldsymbol{\iota}'_{N_j^*}\boldsymbol{c} + \mathbf{w}'_j\mathbf{f} + \xi_j, \tag{23}$$

where $\boldsymbol{c} \equiv \bar{\boldsymbol{d}} + \boldsymbol{b}$. The left hand side of (23) is family level average human capital net of the effect of (strictly exogenous) individual level controls. On the right hand side, family size N_j^* is endogenous in relation to the family fixed effect ξ_j.[13] We overcome this problem by employing an instrumental variable estimator. Specifically, following the literature, we use twins as a source of exogenous variation in family size.[14] Denote the resulting estimator of $\boldsymbol{c}$ by $\hat{\mathbf{c}}_{\text{Between}}$. Using (17) and (23), it is standard to show that

$$\text{plim}\,\hat{\mathbf{c}}_{\text{Between}} = \boldsymbol{\beta}. \tag{24}$$

That is, the between-family instrumental variable 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., $\boldsymbol{\iota}'_{N^*}\boldsymbol{\beta} = \beta N^*$, 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^*.

In the general case of a nonlinear family size profile, (23) 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 (2012b) let $\tilde{z}_j^k$ take the value 1 if family j experienced a twin-birth in the k-th birth and the value 0 otherwise. Since $\tilde{z}_j^k = 0$

[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 (23).

[14] As we noted in our discussion following Proposition 2, we can think of a twin birth 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 (2009) 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 Denmark from which our data originates.

when information on twin birth at parity k is missing due to truncation at N_j^*, $\tilde{z}_j^k$ is correlated with the family fixed effect ξ_j. Formally, $\mathbb{E}[\tilde{z}_j^k \xi_j | \mathbf{w}_j] \neq 0$. 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 $\boldsymbol{\iota}_{N^*}$. $\tilde{z}_j^2$ is in fact a valid instrument, as we exclude single child families (no truncation before the second birth).

Angrist, Lavy, and Schlosser (2010) and Mogstad and Wiswall (2012a) show how this problem of partially missing information can be circumvented. Let

$$z_j^k = \left(\tilde{z}_j^k - \mathbb{E}[\tilde{z}_j^k | \mathbf{w}_j, N_j^* \geq k] \right) \mathbf{1}(N_j^* \geq k), \tag{25}$$

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 (25) by regressing $\tilde{z}_j^k$ onto all variables in $\mathbf{w}_j$ as well as their interactions within the relevant subsamples defined by $N_j^* \geq k$.

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, which are detailed in Appendix B.[15] We supplement IDA with information on the precise date of birth, which allows us to identify twinning (or

[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 completed 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. Appendix B contains further details of the data.

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.

rather multiple births). Information on date of birth is obtained from Statistics Denmark; the merging procedure is documented in the Appendix B. 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. Forty eight percent are females and, on average, their mothers and fathers have completed 10.2 and 11 years of schooling, respectively. An individual has 1.7 siblings, on average.

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. 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

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.

to between-family variation. These values imply that understanding differences in education within the family is important for understanding the overall variation in education.

5 Findings

OLS findings. We start by estimating the impact of family size on a child's education, as specified in expression (18), by OLS. The first column of Table 4 reports findings 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. Consistent with previous findings reported in the literature, 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 findings markedly.[16] The impact of family size is −0.066, smaller but comparable to the coefficient (−0.013) reported in Black, Devereux, and Salvanes (2005) using Norwegian data.

[16] 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.

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.

Birth order effects, whether included linearly or nonlinearly, are negative, large and highly significant. They suggest, for example, that a third child in a family has, on average, 0.61 fewer years of education (column 6 in Table 4). As we discussed earlier, the negative birth order effects could simply reflect family-specific unobservable factors. In columns 7 and 8 of Table 4, we report findings 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 findings are similar in magnitude to the results reported elsewhere in the literature.

IV findings. 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 of Black, Devereux, and Salvanes (2005). 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 findings 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, but insignificant. The coefficient on family size, when using a sample of the first two children in families with three or more children, is the exception. It is negative, sizable (-0.095), and significant at the 5 percent level. Overall, these results are consistent with those reported in Black, Devereux, and Salvanes (2005).

Two-step estimation findings. 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 be misspecified, but we include it because it is a common specification in the literature and thus facilitates 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. 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 (22), 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

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 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 computed according to equation (23), netting out the effects of age and sex as estimated from the family-fixed effect regression.

[2]See Stock and Yogo (2002) 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, 1986).

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. Table 6 implies 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, Lavy, and Schlosser (2010) and Mogstad and Wiswall (2012a) and construct multiple instruments using twin births at each birth order.[17]

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 imply that an additional child in families of size three reduces average years of schooling by 0.12 years, an additional child in families of size four reduces average years of schooling by an additional 0.17 years, whereas an additional child in families of size five or bigger reduces average years of schooling by 0.30 years, all compared to children in families of size two. 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,

[17]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, Lavy, and Schlosser (2010), Mogstad and Wiswall (2012a), or Mogstad and Wiswall (2012b) for details.

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 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 computed according to equation (23), netting out the effects of age and sex as estimated from the family-fixed effect regression.

[2]See Stock and Yogo (2002) 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, 1986).

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 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 nonlinear, 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 (compared 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 some 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 are quantitatively meaningful and how they relate to those reported in 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 (19) 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

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 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 computed according to equation (23), netting out the effects of age and sex as estimated from the family-fixed effect regression.

[2]See Stock and Yogo (2002) 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, 1986).

$\hat{\delta} = -0.178$. Thus we predict $\hat{b}_{\text{IV}} = -0.115 - (-0.178/2) = -0.026$, which is virtually identical to the individual-level IV estimate.

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 (17), Proposition 2 implies that

$$\boldsymbol{\iota}'_N(\boldsymbol{\beta} - \bar{\boldsymbol{\delta}}) - \boldsymbol{\iota}'_{N+z}(\boldsymbol{\beta} - \bar{\boldsymbol{\delta}}) > 0, \tag{26}$$

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

$$\boldsymbol{\iota}'_N\boldsymbol{\beta} - \boldsymbol{\iota}'_{N+z}\boldsymbol{\beta} > 0, \tag{27}$$

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 (26) and (27).

The theoretical predictions can be thus characterized by a set of inequality constraints implying that the null (where (26) and (27) 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, Holly, and Monfort, 1982). We overcome this difficulty by applying a relatively simple test procedure developed in Kodde and Palm (1986).

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 (1986) 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 (1986) 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 from within families and 65 percent from 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- components. The full decomposition is given as:

$$Var[h_{ij}^*] = \mathbb{E}\{Var[\boldsymbol{\iota}_i'\hat{\boldsymbol{\delta}} + \mathbf{x}_{ij}'\hat{\boldsymbol{g}} + \hat{\varepsilon}_{ij}|j]\} + Var[\mathbb{E}\{\boldsymbol{\iota}_{N_j^*}'\hat{\boldsymbol{\beta}} + \mathbf{x}_{ij}'\hat{\boldsymbol{g}} + \mathbf{w}_j'\hat{\boldsymbol{f}} + \hat{\xi}_j|j\}], \qquad (28)$$

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-family 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

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. **x** contains indicators for age and sex of the child. **w** contains indicators for mother's education, mother's age, father's education, and father's age.

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 parents' 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).[18] 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. We note, however, 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 parents' 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

This paper assesses, theoretically and empirically, the relationship between family size, birth order, and children's education. The first contribution of the paper is to provide a theoretical framework to study the relationship between family size and children's education, while recognizing that parents often allocate resources differentially according to a child's birth order.

Guided by the theory, we show that exiting empirical strategies confound the effects of family size and birth order leading to a false rejection of the "quantity-quality trade-off," even when changes in family size are exogenous. The second contribution of the paper is to introduce and implement an empirical strategy that separately identifies the causal effect of family size and birth order. We use within-family variation in educational attainment to estimate the effect of birth order on a child's education, while the family size effect is estimated using between-family variation in educational attainment. Here we use the event of a twin birth as an instrument to circumvent the endogeneity of family size.

[18] 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.

We test our model's predictions using a population-wide comprehensive panel data set from Denmark. Consistent with previous literature, we find that higher birth orders reduce educational attainment. In contrast to previous studies that rejected the presence of a "quantity-quality trade-off" when birth order effects are present, our results suggest that an increase in family size reduces the average education in the family. In summary, our findings support the presence of a "quantity-quality trader-off" even when parents do not treat their children equally. This is our third contribution to the literature.

References

Ahn, T. S., J. Knodel, D. Lam, and J. Friedman (1998): "Family Size and Children's Education in Vietnam," *Demography*, 35(1), 57–70.

Amir, R. (1996): "Sensitivity Analysis of Multisector Optimal Economic Dynamics," *Journal of Mathematical Economics*, 25(1), 123–141.

Angrist, J., V. Lavy, and A. Schlosser (2010): "Multiple Experiments for the Causal Link between the Quantity and Quality of Children," *Journal of Labor Economics*, 28(4), 773–824.

Barro, R., and G. Becker (1989): "Fertility Choice in a Model of Economic Growth," *Econometrica*, 57(2), 481–501.

Becker, G., and H. G. Lewis (1973): "On the Interaction between the Quantity and Quality of Children," *Journal of Political Economy*, 81(2), 279–288.

Becker, G., K. M. Murphy, and R. Tamura (1990): "Human Capital, Fertility, and Economic Growth," *Journal of Political Economy*, 98(5), 12–37.

Becker, G., and N. Tomes (1976): "Child Endowments and the Quantity and Quality of Children," *Journal of Political Economy*, 84(4), 143–162.

Behrman, J. R., R. A. Pollak, and P. Taubman (1989): "Family Resources, Family Size, and Access to Financing for College Education," *Journal of Political Economy*, 97(2), 398–419.

Behrman, J. R., and P. Taubman (1986): "Birth Order, Schooling, and Earnings," *Journal of Labor Economics*, 4(3), 121–145.

Black, S. E., P. Devereux, and K. G. Salvanes (2005): "The More the Merrier? The Effect of Family Composition on Children's Education," *Quarterly Journal of Economics*, 120(2), 669–700.

Booth, A. L., and H. J. Kee (2009): "Birth Order Matters: The Effect of Family Size and Birth Order on Educational Attainment," *Journal of Population Economics*, 22(2), 367–397.

Brock, W. (1971): *Contributions to the von Neumann Growth Model*chap. Sensitivity of Optimal Growth Paths with Respect to a Change in Target Stocks, pp. 73–89. Springer Verlag.

Doepke, M. (2004): "Accounting for Fertility Decline During the Transition to Growth," *Journal of Economic Growth*, 9(3), 347–383.

Eckstein, D., K. J. Aycock, M. A. Sperber, J. McDonald, V. V. Wiesner, R. E. Watts, and P. Ginsburg (2010): "A review of 200 birth-order studies: Lifestyle characteristics," *The Journal of Individual Psychology*, 66(4), 408–434.

Galor, O., and D. Weil (2000): "Population, Technology and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond," *American Economic Review*, 90(4), 806–828.

Glick, P., A. Marini, and D. E. Sahn (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*, 69(5), 667–691.

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

Hanushek, E. (1992): "The Trade-off between Child Quantity and Quality," *Journal of Political Economy*, 100(1), 84–117.

Hauser, R. M., and W. H. Sewell (1985): "Birth Order and Educational Attainment in Full Sibships," *American Educational Research Journal*, 22(1), 1–23.

Hausman, J. A., and W. E. Taylor (1981): "Panel Data and Unobservable Individual Effects," *Econometrica*, 49(6), 1377–1398.

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

Iacovou, M. (2008): "Family size, Birth Order and Educational Attainment," *Marriage and Family Review*, 42(3), 35–57.

Juhn, C., Y. Rubinstein, and A. Zuppann (2013): "The Quantity-Quality Tradeoff and the Formation of Cognitive and Non-Cognitive Skills," Discussion paper, University of Houston.

Kodde, D. A., and F. C. Palm (1986): "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica*, 54(5), 1243–1248.

Li, H., J. Zhang, and J. Zhu (2008): "The Quantity-Quality Trade-off of Children in a Developing Country: Identification using Chinese Twins," *Demography*, 45(1), 223–243.

Milgrom, P., and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157–180.

Millet, D. L., and L. Wang (2011): "Is the Quantity-Quality Trade-off Really a Trade-off for All?," *Economic Development and Cultural Change*, 60(1), 155–195.

Mogstad, M., and M. Wiswall (2012a): "Instrumental variables estimation with partially missing instruments," *Economics Letters*, 114(2), 186–189.

——— (2012b): "Testing the Quantity-Quality Model of Fertility: Linearity, Marginal Effects, and Total Effects," Discussion paper, University College London.

Parish, W., and R. Willis (1993): "Daughters, Education, and Family Budgets: Taiwan Experiences," *Journal of Human Resources*, 28(4), 863–898.

Qian, N. (2009): "Quantity-Quality: The Positive Effect of Family Size on School Enrollment in China," NBER Working Papers 14973, National Bureau of Economic Research, Inc.

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

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

Schultz, T. (1997): "Demand for Children in Low Income Countries," in *Handbook of Population and Family Economics*, ed. by M. Rosenzweig, and O. Stark, Handbook of Population and Family Economics. North-Holland.

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

Vincent, T. L., and R. G. Brusch (1970): "Optimal Endpoints," *Journal of Optimization Theory and Applications*, 6(4), 299–319.

Appendix A: Omitted proofs and remarks

This Appendix collects proofs that are not directly obtainable from the equations available in the text. It also presents some general remarks and supplemental results. 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 $u_{hh}^*(i)$. Similar notation applies to the parental valuation of consumption.

Proof of Proposition 2. The proof is in two parts. First, in Lemma 1, we show that an exogenous increase in family size tightens the budget constraint, i.e., it increases the Lagrange multiplier of (2). Second, using this result and the first-order condition (4), we show that the human capital of existing children declines as family size increases exogenously.

Lemma 1 *An exogenous increase in family size increases the marginal utility of wealth. That is,* $\partial\lambda(z)/\partial z > 0$.

Proof. 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

$$\theta u_{hh}^*(i)\frac{\partial h^*(i|N^*+z)}{\partial z} = \frac{\partial\lambda(z)}{\partial z}p(i) = \frac{\partial\lambda(z)}{\partial z}\frac{\theta u_h^*(i)}{\lambda(z)}. \tag{A1}$$

Likewise, the first-order condition of $C^*(z)$ is

$$(1-\theta)V_C(C^*(z)) = \lambda(z), \text{ with} \tag{A2}$$

$$(1-\theta)V_{CC}(C^*(z))\frac{\partial C^*(z)}{\partial z} = \frac{\partial\lambda(z)}{\partial z}. \tag{A3}$$

The budget constraint (2) and the first-order conditions, (4) and (A2), yield

$$\lambda(z) = \frac{(1-\theta)V_C(C^*(z))C^*(z)}{Y} + \int_0^{N^*+z}\frac{\theta u_h^*(i)h^*(i)}{Y}di. \tag{A4}$$

Differentiating (A4) using Leibnitz's rule, and using (A1) and (A3), yields

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

which can be written simply as

$$\frac{\partial\lambda(z)}{\partial z}\Phi(z) = \left.\frac{\theta u_h^*(i)h^*(i)}{Y}\right|_{i=N^*+z}, \text{ with}$$

$$\Phi(z) \equiv 1 - \left\{ C^*(z) - \frac{V_C(C^*(z))}{[-V_{CC}(C^*(z)]} \right\} \frac{1}{Y} - \int_0^{N^*+z} \left\{ p(i)h^*(i) - \frac{u_h^*(i)}{[-u_{hh}^*(i)]} p(i) \right\} \frac{di}{Y}.$$

Since parents value the human capital of the "additional children," to show that $\partial\lambda(z)/\partial z > 0$ it is then enough to show that $\Phi(z) > 0$. To that end, use (4) in $\Phi(z)$ to obtain

$$\Phi(z) = 1 - \left\{ C^*(z) + \int_0^{N^*+z} p(i)h^*(i)di \right\} \frac{1}{Y} \qquad \text{(A5)}$$

$$+ \frac{V_C(C^*(z))}{[-V_{CC}(C^*(z))]} \frac{1}{Y} + \int_0^{N^*+z} \left\{ \frac{u_h^*(i)}{[-u_{hh}^*(i)]} \right\} \frac{p(i)di}{Y}.$$

The first line in (A5) cancels out; see (2). The second line is positive by the concavity of the utility functions. This shows that $\partial\lambda(z)/\partial z > 0$ as needed. ■

To complete the proof of Proposition 2, notice that (A1) implies

$$\frac{\partial h^*(i|N^* + z)}{\partial z} = -\frac{\partial\lambda(z)}{\partial z} \frac{p(i)}{[-\theta u_{hh}^*(i)]}. \qquad \text{(A6)}$$

Lemma 1 and the strict concavity of the utility function yield $\partial h^*(i|N^* + z)/\partial z < 0$ for all $i \in [0, N^* + z]$. This establishes the stated ranking for the existing children. ■

Remarks. (i) Since family size changes exogenously, there is no use for the first-order condition with respect to N^* in the previous proof. The proof treats $N^* + z$ as a parameter and examines the changes in consumption and human capital in response to changes in z. Using (A3) and Lemma 1 it is possible, for example, to show that parental consumption declines when family size increases exogenously,

$$\frac{\partial C^*(z)}{\partial z} = -\frac{1}{[-(1-\theta)V_{CC}(C^*(z))]} \frac{\partial\lambda(z)}{\partial z} < 0.$$

Therefore, the decline in the existing children's human capital discussed in Proposition 2 takes place even after parents adjust their consumption in response to the presence of twins.

(ii) Since the first-order condition (3) is redundant to study the effects of exogenous changes in family size, our main conclusions will likely not change in environments with more general preferences for family size. For example, Barro and Becker (1989) consider a nonlinear aggregation of the children's utilities, as in

$$(1-\theta)V(C) + \varphi U(N) + \theta\alpha(N) \int_0^N u(h(i), i)di.$$

The total utility from parental investments is now discounted by the term $\alpha(N)$. Introducing this term changes the first-order condition (3). These changes, however, are irrelevant for understanding the effects of exogenous changes in family size as these generalizations influence N^* but not the response in C^* and $h^*(i)$ to *exogenous* changes in family size. For example, the first-order condition with respect to C remains unchanged. The first-order condition with respect to $h(i)$ becomes $\theta\alpha(N)u_h^*(i) = \lambda(z)p(i)$. Since the term $\alpha(N)$ multiplies the marginal utilities of all children, the comparative statics with respect to z remain unchanged.[19] Thus (A1) and (A6) continue to hold.

(iii) Likewise, suppose, as in Behrman and Taubman (1986), that the sub-utilities of children are aggregated using a CES aggregator with an elasticity parameter $1-\rho$, as in

$$\left[\int_0^N u(h(i),i)^{1-\rho}di\right]^{\frac{1}{1-\rho}}. \tag{A7}$$

Despite this additional margin of substitution, the analysis of this problem can be undertaken in essentially the same manner as in the basic model. In particular, since (C^*, N^*) are household-level variables, we can write

$$\frac{\partial h^*(i)}{\partial i} = \tilde{\Delta}(h^*(i),i), \text{ for all } i \in [0,N^*], \text{ with} \tag{A8}$$

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

so that when $\rho = 1$, we have $\tilde{\Delta}(h,i) = \Delta(h,i)$ given in the basic model. We can therefore obtain a characterization analogous to Proposition 1. An exogenous increase in family size can also be studied as in the basic model because, from a parental perspective, such an increase is equivalent to a reduction in income.

(iv) The budget constraint can also be generalized without affecting the main qualitative conclusions of the analysis. For example, suppose that there is a "fixed" cost $P > 0$ for family size as well as a time cost $\psi \leq 1/N^+$ proportional to each child. The generalized budget constraint is

$$C + PN + \int_0^N p(i)h(i)di \leq Y(1-\psi N), \tag{A9}$$

where the time worked is $(1-\psi N) \in (0,1)$. These generalizations are relevant to determine optimal fertility N^* but they do not alter the main conclusions regarding the effects of exogenous changes in family size. In fact, an exogenous increase in family size reduces parental resources even more compared to our baseline case. In the baseline case, total income is constant whereas

[19]Other changes in parental preferences, however, are likely to invalidate Lemma 1 and Proposition 2. For example, if parental consumption and family size are not separable, as in $(1-\theta)V(C,N)$, parents might reduce their consumption more than in our baseline case. Thus parents with additional children will be able to invest more than (or at least as much as) parents without additional children.

in (A9) income actually declines as family size exogenously increases.

(v) The proof of Lemma 1 uses the fact that $\Phi(z) > 0$. Simple substitutions show that $\Phi(z)$ is a weighted average of the elasticities of substitution of $V(C)$ and $u(h,i)$, with weights given by the budget shares,

$$\Phi(z) = \frac{V_C(C^*(z))}{[-V_{CC}(C^*(z))C^*(z)]}\frac{C^*(z)}{Y} + \int_0^{N^*+z}\left\{\frac{u_h^*(i)}{[-u_{hh}^*(i)h^*(i)]}\right\}\frac{p(i)h^*(i)}{Y}di. \tag{A10}$$

Suppose, as a special case, that $V(C) = C$ so that parents have an infinite willingness to substitute their consumption in response to an exogenous increase in family size (provided that $C^*(z) > 0$). As the first-order condition (A2) suggests, Lemma 1 does not hold because the marginal utility of wealth is constant. In this case, the children's human capital profiles are unaffected by exogenous changes in family size; see, e.g., (A6). As another special case, suppose that parents are unwilling to substitute their consumption for their children's human capital, i.e., $V_C(C)/[-V_{CC}(C)C] \to 0$. This case was illustrated by expression (8). Finally, suppose that the elasticity of substitution of $V(C)$ is bounded away from unity, or

$$\frac{V_C(C)}{[-V_{CC}(C)]C} \leq 1. \tag{A11}$$

Then (A10) and (A11) imply that

$$\int_0^{N^*+z}\left\{\frac{u_h^*(i)}{[-u_{hh}^*(i)h^*(i)]}\right\}\frac{p(i)h^*(i)}{Y}di \geq \Phi(z) - 1. \tag{A12}$$

The following proof makes use of the previous inequality.

Proof of Proposition 3. Differentiating $H^*(N+z)$ in (9) yields

$$\frac{\partial H(N^*+z)}{\partial z} = -\frac{1}{N^*+z}\left[H(N^*+z) - h^*(N^*+z) - \int_0^{N^*+z}\frac{\partial h^*(i|N^*+z)}{\partial z}di\right]. \tag{A13}$$

If $h^*(i)$ is decreasing in i, then $H(N^*+z) > h^*(N^*+z)$ and $\partial H(N^*+z)/\partial z < 0$. This case corresponds to condition (a).

Consider next the case in which $h^*(i)$ is increasing in i. Using (A6) and (A10) yields

$$\begin{aligned} -\int_0^{N^*+z}\frac{\partial h^*(i|N^*+z)}{\partial z}di &= \frac{\theta u_h^*(N^*+z)h^*(N^*+z)}{\Phi(z)}\int_0^{N^*+z}\frac{p(i)}{[-\theta u_{hh}^*(i)]}\frac{1}{Y}di, \\ &= \frac{h^*(N^*+z)}{\Phi(z)}\int_0^{N^*+z}\frac{u_h^*(N^*+z)}{[-u_{hh}^*(i)h^*(i)]}\frac{p(i)h^*(i)}{Y}di. \end{aligned} \tag{A14}$$

A sufficient condition for a negative sign in (A13) is

$$u_h(h^*(N^*+z), N^*+z)/\Phi(z) \geq u_h(h^*(i), i)/(\Phi(z)-1). \tag{A15}$$

To see this, notice that (A14) and (A15) imply

$$-\int_0^{N^*+z} \frac{\partial h^*(i|N^*+z)}{\partial z} di \geq \frac{h^*(N^*+z)}{\Phi(z)-1} \int_0^{N^*+z} \frac{u_h^*(i)}{[-u_{hh}^*(i)h^*(i)]} \frac{p(i)h^*(i)}{Y} di \geq h^*(N^*+z),$$

where the second inequality follows from (A12).

Next, to show that (A15) corresponds to condition (b) in Proposition 3, rewrite (A15) as

$$u_h(h^*(N^*+z), N^*+z) - u_h(h^*(i), i) \geq \frac{u_h(h^*(N^*+z), N^*+z)}{\Phi(z)}.$$

Using the fact that $h^*(i)$ is increasing in i, and substracting and adding $u_h(h^*(N^*+z), i)$ on the left-hand-side of the previous expression, it is possible to write (A15) simply as $u_{hi}(h,i) + u_{hh}(h,i) \geq \bar{c}$ for all (h,i). This is condition (b) with $\bar{c} \equiv u_h(h^*(N^*+z), N^*+z)/\Phi(z)$. ■

Appendix B: 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 B.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]

[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

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.

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

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.

[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.

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.

- 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

[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.

[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).

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.

Table B.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	missing

Note: Immigrants are overrepresented in the Unknown category as education obtained in their home country is seldom converted to Danish standards and recorded.