Quantitative Economics 7 (2016), 157–192 1759-7331/20160157

Testing the quantity–quality model of fertility: Estimation using unrestricted family size models

Magne Mogstad
Department of Economics, University of Chicago and Statistics Norway

Matthew Wiswall
Department of Economics, Arizona State University

We examine the relationship between child quantity and quality. Motivated by the theoretical ambiguity regarding the sign of the marginal effects of additional siblings on children's outcomes, our empirical model allows for an unrestricted relationship between family size and child outcomes. We find that the conclusion in Black, Devereux, and Salvanes (2005) of no family size effect does not hold after relaxing their linear specification in family size. We find nonzero effects of family size in ordinary least squares estimation with controls for confounding characteristics like birth order and in instrumental variables estimation instrumenting family size with twin births. Estimation using a unrestricted specification for the quality–quantity relationship yields substantial family size effects. This finding suggests that social policies that provide incentives for fertility should account for spillover effects on existing children.

Keywords. Quantity–quality model of fertility, family size, birth order, nonlinearity, instrumental variables.

JEL classification. C26, C31, J13.

1. Introduction

Most developed countries have a range of policies affecting fertility decisions, including cash transfers (e.g., family allowances, single-parent benefits, and family tax credits) and in-kind transfers (e.g., child care subsidies). In fact, families with children receive special treatment under the tax and transfer provisions in 28 of the 30 Organization for Economic Development and Cooperation (OECD) countries (OECD (2002)). Many of these policies are designed such that they reduce the cost of having a single child more than the cost of having two or more children, in effect promoting smaller families. For example, welfare benefits or tax credits are, in many cases, reduced or even cut off after reaching a certain number of children (see, e.g., Feyrer, Sacerdote, and

Magne Mogstad: magne.mogstad@gmail.com
Matthew Wiswall: matt.wiswall@gmail.com
Thanks to two anonymous referees, the editor, and a number of seminar and conference participants for useful comments and suggestions. The Norwegian Research Council has provided financial support for this project.

DOI: 10.3982/QE322

Stern (2008), Del Boca and Wetzels (2008)). Motivated by the seminal quantity–quality (QQ) model of fertility by Becker and Lewis (1973), the advocates of policies promoting smaller families argue that large families invest less in the quality of each child and thereby keep human capital and living standards low (see, e.g., Galor and Weil (2000), Moav (2005)).

A large and growing body of empirical research tests the QQ model by estimating the relationship between family size and children's outcomes, and by examining the spillover on existing children when a family has additional children. Much of the early literature found that larger families reduce child quality, such as educational attainment (e.g., Rosenzweig and Wolpin (1980), Hanushek (1992)). However, recent studies from several developed countries, using large data sets, controls for confounding characteristics such as birth order, and instrumental variables for family size, have challenged this conclusion and argued that family size has no causal effect on children's outcomes. Black, Devereux, and Salvanes (2005, p. 697) conclude that "there is little if any family size effect on child education; this is true when we estimate the relationship with controls for birth order [OLS] or instrument family size with twin births." Using data from the United States, Caceres-Delpiano (2006) comes to a similar conclusion. Other recent studies reporting no effect of family size include Angrist, Lavy, and Schlosser (2010) using data from Israel and Aaslund and Grønquist (2010) using data from Sweden.

Although these studies represent a significant step forward, a concern is that the evidence for no quantity–quality trade-off is based on a model that is linear in family size, imposing constant marginal effects of additional siblings across family sizes. Motivated by the theoretical ambiguity about the sign of the marginal effects on child quality of additional siblings (Rosenzweig and Wolpin (1980)), we estimate a model that allows for an unrestricted relationship between family size and child outcomes. We estimate unrestricted models in family size using both empirical strategies employed to test the QQ model: ordinary least squares (OLS) estimation with controls for confounding characteristics like birth order, and instrumental variable (IV) estimation using instruments for family size. To rule out differences simply because of different data sources, we use the same data as Black, Devereux, and Salvanes (2005), administrative registers for the entire population of Norway. In addition, we follow their study in using twin births as the instrument for family size.

Our study proceeds by first showing theoretically that the QQ model does not suggest that the marginal family size effects are constant or even negative at all margins. On the contrary, even with no heterogeneity in the production function of child quality, there can be nonlinearities in child quality from changes in family size. This is because parental preferences mediate exogenous changes in family size: With different levels of complementarity in parental preferences between quantity and quality of children, an increase in the number of children can have negative or positive effects on existing children. Because the relationship between family size and child outcome may be nonlinear and even nonmonotonic, imposing a linear specification in family size when testing the QQ model is worrisome.

Using the Norwegian data, we next show empirically that although the OLS and IV estimates of the linear model indicate little if any effect of family size on children's education, this finding does not hold up when we relax the linearity restriction in family size.

The OLS estimates of an unrestricted model in family size reveal an inverse U-shaped pattern, with statistically significant and sizable marginal family size effects. Turning to IV estimates, we find large and statistically significant family size effects. For first born children, we estimate a nonmonotonic relationship between their education and number of siblings. Although a third child added to a 2 child family increases the educational attainment of first born children, additional children have a negative marginal effect. The negative effects of family size at higher parities actually exceed the marginal birth order effects. We find similarly large negative effects of additional siblings for later born children. In comparison, the restricted linear IV estimates are close to zero for children of all birth orders. By comparing the results from the linear and unrestricted models in family size, we see the important role of the linearity restriction in masking the family size effects.

To understand why the linear model yields a misleading picture of the relationship between family size and children's education, we estimate the weights attached to the marginal family size effects for the linear OLS and IV estimators. In general, the weights each estimator places on different family size margins differ across estimators, and in the case of the IV estimator, by instrument. The reasons for the almost zero effect of family size in both OLS and IV estimation of the linear model are that negative and positive marginal effects at different parities are weighted in such way that these effects offset, creating an estimate close to zero.

Our results are important both on the policy side and in terms of assessing the empirical content of the QQ model. Accepting the recent findings of no effect of family size suggests that there is no need to be concerned with the spillovers on existing children when designing policies to change fertility rates. Our findings run counter to this conclusion. The evidence of an inverse U-shaped pattern suggests that an efficient policy might be to target incentives for higher fertility to small families and to discourage larger families from having additional children. In terms of the QQ model, the estimated marginal effects can be interpreted as suggesting a substitution between quantity and quality in large families, and complementarities between quantity and quality in small families.

The paper unfolds as follows. Section 2 outlines the QQ model, focusing on the theoretical ambiguity about the magnitude and sign of the marginal effects on child quality of additional siblings. Section 3 describes our data. Section 4 discusses the empirical models and compares OLS estimates of the linear and unrestricted models in family size. Section 5 describes the IV methods and presents IV results from the linear and unrestricted family size model. Section 6 summarizes and concludes with a discussion of policy implications.

2. Family size and child quality in the QQ model

In the seminal QQ model of fertility introduced by Becker and Lewis (1973), a unitary household is assumed to choose the number of children and expenditure on child specific goods to maximize a utility function $U(N, Q, C)$, with number of children N, the quality per child Q, and parental consumption C as arguments. Parents are endowed with I in income from which they can finance their own consumption and purchase child specific goods. For simplicity, we ignore price differences in child specific and parental specific goods. There is assumed to be an underlying homogeneous production function that relates expenditure on child specific goods per child, e, to child quality: $Q = q(e)$.

The QQ model assumes that child quality and quantity are jointly determined. For a given number of children N, the optimal expenditure per child on child specific goods can be defined as

$$e^*(N) = \arg\max_e U(N, Q, C)$$
$$\text{s.t.} \quad I = Ne + C \quad \text{and} \quad Q = q(e).$$

The level of quality for each child in a family with N total children is then given by $q(e^*(N))$.

Since the seminal work of Rosenzweig and Wolpin (1980), major empirical interest has centered on testing the QQ model (Black, Devereux, and Salvanes (2005), Caceres-Delpiano (2006), Angrist, Lavy, and Schlosser (2010), Aaslund and Grønquist (2010)). The identification problem posed by the joint determination of N and Q has typically been addressed by using the randomness of twin birth as a source of exogenous variation in N, and/or controlling for confounding characteristics like birth order and parental age and education when regressing family size on child outcomes.

The difference in the quality of a child from an exogenous increase in family size from $N - 1$ to N children is given by

$$\Delta(N, N-1) = q\big(e^*(N)\big) - q\big(e^*(N-1)\big), \tag{1}$$

where (1) defines the *marginal* family size effect for a given child at the $N - 1$ family size margin for any $N > 1$. For example, $\Delta(3, 2) = q(e^*(3)) - q(e^*(2))$ is the marginal effect of another sibling for a child from a 2 child family. The family size effects are *linear* if the marginal effects are constant: $\Delta(N + 1, N) = \Delta(N, N - 1)$ for all $N > 1$. Speaking to the unrestricted specification of family size in the OLS estimation in Black, Devereux, and Salvanes (2005), the *total* effect of family size relative to 1 child families is given by

$$\Omega(N, 1) = q\big(e^*(N)\big) - q\big(e^*(1)\big). \tag{2}$$

Although $\Omega(2, 1) = \Delta(2, 1)$, the marginal effect and total effects will generally differ, even if there is a linear relationship between family size and child outcome.

Although the QQ model assumes that family size and child quality is jointly determined, there is nothing in the theory that suggests that the marginal family size effects are constant or even negative at all margins. This was pointed out by Rosenzweig and Wolpin (1980) but has received little attention in the subsequent empirical literature. To illustrate this point, we use a parameterized version of the QQ model, assuming a nested constant elasticity of substitution (CES) structure for preferences. We emphasize that this parameterization is merely to illustrate the *possibility* that the QQ model allows for nonlinear and positive marginal effects of additional children. Importantly, we do not impose this parameterization in the empirical estimation.

Assume preferences and technology take the form

$$U(N, Q, C) = U_1(N, Q)^{\nu} C^{1-\nu},$$

where

$$U_1(N, Q) = \left[\alpha N^{\sigma} + (1-\alpha) Q^{\sigma}\right]^{1/\sigma},$$

and child quality production technology takes the form

$$q(e) = e^{\gamma}, \quad \text{with } \gamma > 0,$$

where $\nu \in (0, 1)$, $\sigma \in (-\infty, \infty)$, $\alpha \in (0, 1)$, and $\gamma \in (0, \infty)$. In this specification, child quality and quantity form a CES aggregate $U_1(N, Q)$, with elasticity of substitution between quantity and quality of $1/(1-\sigma)$. Parents are then assumed to have Cobb–Douglas preferences over the quantity and quality child aggregate $U_1(N, Q)$ and parental consumption C with parameter ν.

Figure 1 provides an illustration of how the marginal effects (1) and total effects (2) vary as we change the substitution elasticities between quality and quantity. In this figure, parental income, the child quality technology γ, and preferences for parental consumption ν are kept constant. We vary only the substitution elasticity $1/(1-\sigma)$ from a low value of 0.1 to a high value of 2. The vertical axis measures the total effect of family size (the level of child quality) relative to 1 child families, $\Omega(N, 1)$, whereas the slopes for each of the curves provide the marginal effects, $\Delta(N, N-1)$. We immediately see that even with no heterogeneity in the production function of child quality, there can be

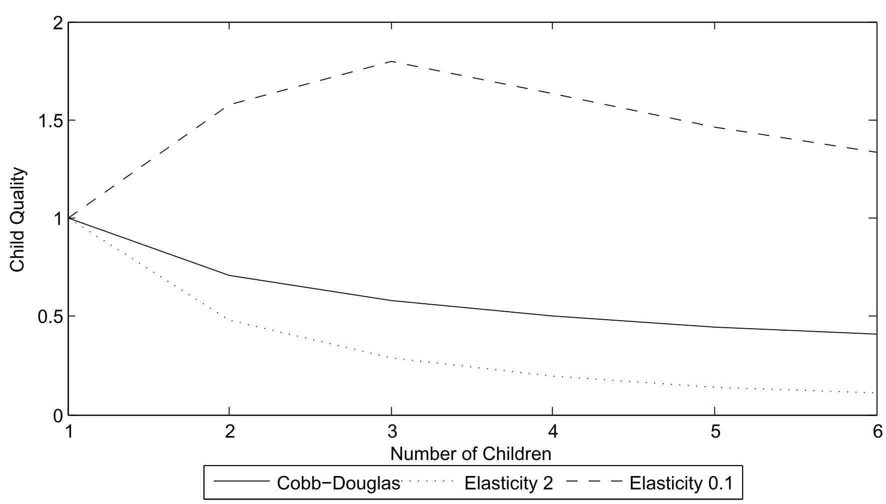

FIGURE 1. Family size effects in the quantity–quality model with different substitution elasticities between the quantity and quality of children. *Note*: The other model parameters are set at $\alpha = 0.5$, $\gamma = 0.5$, and $\theta = 0.5$.

nonlinearities in the effects on child quality from changes in family size. This is because parental preferences mediate exogenous changes in family size. Given different levels of complementarity between quantity and quality of children, an increase in the number of children can have large or small and negative or positive effects on existing children.

Figure 1 also illustrates that the marginal effects—not the total effects—provide the appropriate test for a quantity–quality trade-off. For instance, for families with more than 2 children, the total effect of family size can be zero even if all the marginal effects are nonzero. This occurs, for example, for a 3 child family if the marginal effect from 1 to 2 children offsets the marginal effect from 2 to 3 children. In this case, $\Omega(N, 1) = \Delta(3, 2) + \Delta(2, 1) = 0$, although both $\Delta(3, 2)$ and $\Delta(2, 1)$ are nonzero. This is not merely a theoretical peculiarity, but is exactly what occurs for the OLS estimates for first born children in Norway, as shown in Figure 2 and discussed below. In terms of the QQ model, such an inverse U-shaped pattern suggests a trade-off between quantity and quality in large families, and strong complementarities between quantity and quality in small families.

It should finally be noted that nonlinearities and positive marginal family size effects could come from several sources outside the QQ model. In particular, additional siblings might benefit existing children if they stabilize parental relationship (see, e.g., Becker (1998)), if they make maternal employment less likely (see, e.g., Ruhm (2008)), or if there are positive direct spillover effects among siblings (see, e.g., Bandura (1977)).

3. Data

As in Black, Devereux, and Salvanes (2005), our data are based on administrative registers from Statistics Norway that cover the entire resident population of Norway who were between 16 and 74 years of age at some point during the period 1986–2000. The family and demographic files are merged by unique individual identifiers with detailed information about educational attainment reported annually by Norwegian educational establishments. The data also contain family identifiers that allow us to match parents to their children. As we observe each child's date of birth, we are able to construct birth order indicators for every child in each family.

The Norwegian data have several advantages over those available in most other countries. First, Norway offers a population panel data set over an extended period of time; as a result, we are able to overcome many limitations of earlier research resulting from small sample sizes or limited information on children's long-run outcomes. Second, we observe a large number of twin births, which generates plausible exogenous variation in family size; importantly, the children we consider were born before the introduction of fertility drugs and other treatments (that strongly increase the odds of twinning).[1]

To the best of our knowledge, we use the same sample selection as Black, Devereux, and Salvanes (2005). We restrict the sample to children who were aged at least 25 in 2000

[1]The children we consider were also born before the legalization of abortion and the introduction of birth control pills. See Black, Devereux, and Salvanes (2005) for a more detailed discussion of institutional details.

Table 1. Descriptive statistics.

	Mean	Std. Dev.
Age in 2000	38.5	8.6
Female	0.48	0.50
Education	12.1	2.6
Mother's education	9.9	1.3
Father's education	10.3	2.2
Mother's age in 2000	65.8	10.6
Father's age in 2000	67.3	10.3
Number of children	2.9	1.2
Twins in family	0.014	0.12

Note: Descriptive statistics are for 1,429,126 children from 625,068 families with no more than 6 children (98% of the full sample). All children are aged at least 25 in 2000. Twins are excluded from the sample. All children and parents are aged between 16 and 74 years at some point between 1986 and 2000. Source: Administrative registers from Statistics Norway.

to make it likely that most individuals in our sample have completed their education. Twins are excluded from the estimation sample because of the difficulty of assigning birth order to these children. To increase the chances of our measure of family size being completed family size, we drop families with children aged less than 16 in 2000. We also exclude a handful of families where the mother had a birth before she was aged 16 or after she was 49. In addition, we exclude a small number of children where their own or their mother's education is missing. Rather than dropping the larger number of observations where information on fathers is missing, we include a separate category of missing for father's education and father's age.

The only difference between our sample selection and that in Black, Devereux, and Salvanes (2005) is that we exclude a small number of families with more than 6 children. The final sample includes 1,429,126 children from 625,068 families (98 percent of the full sample of all families). Table 1 displays the basic descriptive statistics for this sample. In all respects, there are only minor differences between our sample and that of Black, Devereux, and Salvanes (2005). Moreover, we cannot detect any difference between the characteristics of the full sample and our sample of families with 6 or fewer children. About 48 percent of the children in the sample are female and a twin birth occurs in about 1.4 percent of families. The ages of the child, the mother, and the father are measured in year 2000. The child's education is also collected from year 2000, and the education of the parents is measured at age 16 of the child. Fathers are, on average, slightly older and more educated than mothers.

The measure of family size is the number of children born to each mother. In the sample of families with 6 or fewer children, the average family size is 2.9 children. Table 2 provides the distribution of family sizes. Nearly 8 percent of the sample were only children, 33 percent were from 2 child families, and 32 percent were from 3 child families. The remaining 27 percent of the sample consist of children born to families with 4, 5, or 6 children.

Table 2. Distribution of family sizes by children.

Family Size	Number	Fraction
1	111,064	0.078
2	477,633	0.334
3	459,831	0.322
4	239,840	0.168
5	99,940	0.070
6	40,818	0.029

Note: Descriptive statistics are for 1,429,126 children from 625,068 families with no more than 6 children (98% of the full sample). All children are aged at least 25 in 2000. Twins are excluded from the sample. All children and parents are aged between 16 and 74 years at some point between 1986 and 2000. Source: Administrative registers from Statistics Norway.

4. Empirical models and OLS estimates

This section focuses on the first of the two empirical strategies employed by the previous literature to estimate the effects of family size: OLS estimation with controls with confounding characteristics such as birth order.

4.1 *Linear versus unrestricted models in family size*

The main empirical model used in the family size literature specifies outcomes for children as a function of their family size and a vector of other covariates X_i. For child i, we denote her number of siblings using $s_i \in \{0, 1, \dots, \bar{s}\}$. When convenient, we also refer to the effect of family size defined as the total number of children in the family: c_i. The linear model in the family size model is specified as

$$y_i = \beta s_i + X_i' \delta + \varepsilon_i, \tag{3}$$

where X_i always includes a constant and, in some specification, a set of controls for child i's birth order and other characteristics.

Motivated by the theoretical ambiguity in the functional form of the relationship between family size and child quality, our point of departure is to generalize (3) and specify an unrestricted model in family size by including dummy variables for each number of siblings,

$$y_i = \gamma_1 d_{1i} + \cdots + \gamma_{\bar{s}} d_{\bar{s}i} + X_i' \delta + \varepsilon_i, \tag{4}$$

where $d_{si} = 1\{s_i \geq s\}$. This dummy variables construction implies that the γ_s coefficients provide the *marginal* effect of having s siblings rather than $s-1$ siblings. The linear model (3) restricts the marginal effects to be constant at $\gamma_s = \beta$ for all s. With respect to the QQ model discussed above, γ_s from (4) is the "reduced form" analog of $\Delta(N, N-1)$ from (1), where the $N-1$ to N margin is the same as the s sibling margin.

4.2 *OLS estimates with controls for birth order*

Table 3 reports the OLS estimates of family size on children's education. The first column of Table 3 reports the OLS estimate from model (3), suggesting that each additional sibling reduces the average education of the children in the family by as much as 0.2 years. The second column of Table 3 report the OLS estimates from model (4). The estimates show the marginal effects of increasing family size by one additional sibling, indicating a nonmonotonic relationship between family size and children's education. Moving from a 1 child family to a 2 child family is estimated to increase education by 0.37 years. In contrast, the marginal effects of additional siblings at higher birth parities are negative.

The next four columns of Table 3 add control variables (the same as Black, Devereux, and Salvanes (2005)) to models (3) and (4). Columns 3 and 4 add dummy variables for gender, child's birth cohort, mother's birth cohort, father's birth cohort, mother's education, and father's education. Including these variables reduces (in absolute value) both the linear and the unrestricted estimates of the effect of family size on children's education, suggesting that OLS estimation could be biased because child quality and quantity is jointly determined.

The drop in family size effects from adding control variables highlight the concern about omitted variables bias in the OLS estimation. To address this concern, previous research using OLS has controlled for birth order, hoping that any remaining bias is small. Columns 5 and 6 add controls for birth order to models (3) and (4). To provide a direct comparison to the marginal family size effects, we construct the five dummy variables for birth order as marginal effects: The first dummy variable is equal to 1 if the child was born second or higher in the birth order (and is 0 otherwise), the second dummy variable is equal to 1 if the child was born third or higher in the birth order (and is 0 otherwise), and so on. In this specification of model (4), the supports of the birth order and family size variables are fully saturated, with the reference category specified as first born children in families with 1 child (only children). The estimates then indicate the marginal effect of increasing family size by 1 child (e.g., from a 1 child family with 0 siblings to a 2 child family with 1 sibling) or of being born one birth parity later in the birth order (e.g., from first to second born).

We find that the effect of family size in the linear model that controls for birth order and other demographic variables is very small, around −0.01. However, when comparing the results from column 5 to those from column 6, we see that relaxing the linearity assumption in family size reveals always significant and mostly sizable marginal family size effects. Controlling for birth order actually sharpens the picture of an inverse U-shaped pattern in family size. In particular, the inclusion of birth order controls boosts the only child penalty, as the marginal effect of moving from 0 to 1 sibling increases from 0.042 to 0.224 additional years of education. In comparison, the marginal effect of moving from 1 to 2 siblings is estimated to be small and positive at 0.02. However, the marginal effects of additional siblings at higher parities are between −0.073 and −0.089.[2]

[2] We have performed two sensitivity checks of the sample used in Table 3. First, we have included families with more than 6 children. Second, we have excluded one child families. In both cases, the estimates of the marginal effects of family size barely move.

TABLE 3. OLS estimates of marginal effects from linear and unrestricted models.

	Marginal Effects						Total Effects From (6)
	(1)	(2)	(3)	(4)	(5)	(6)	
Linear family size	−0.198		−0.112		−0.008		
	(0.003)		(0.002)		(0.003)		
							Total Effect vs. 0 Siblings
Siblings ≥ 1		0.370		0.042		0.224	0.224
		(0.009)		(0.008)		(0.001)	(0.001)
Siblings ≥ 2		−0.148		−0.099		0.020	0.244
		(0.007)		(0.006)		(0.006)	(0.009)
Siblings ≥ 3		−0.352		−0.157		−0.073	0.171
		(0.009)		(0.007)		(0.008)	(0.011)
Siblings ≥ 4		−0.348		−0.146		−0.089	0.082
		(0.014)		(0.012)		(0.012)	(0.014)
Siblings = 5		−0.281		−0.131		−0.084	−0.002
		(0.023)		(0.019)		(0.019)	(0.020)
							Total Effect vs. First Born
Birth order ≥ 2					−0.332	−0.373	−0.373
					(0.005)	(0.005)	(0.005)
Birth order ≥ 3					−0.222	−0.219	−0.591
					(0.006)	(0.006)	(0.007)
Birth order ≥ 4					−0.157	−0.100	−0.691
					(0.009)	(0.009)	(0.011)
Birth order ≥ 5					−0.106	−0.040	−0.731
					(0.015)	(0.015)	(0.017)
Birth order ≥ 6					−0.117	−0.063	−0.791
					(0.029)	(0.029)	(0.029)
Control variables	No	No	Yes	Yes	Yes	Yes	
R-squared	0.008	0.012	0.204	0.204	0.208	0.209	

Note: Each column is a separate regression. Columns 1–6 provide *marginal effects* of family size and birth order. Siblings ≥ 1 is the marginal effect from moving from 0 to 1 siblings, siblings > 2 is the marginal effect from moving from 1 to 2 siblings, and so on. The last column reports the *total effects* and standard errors for the marginal effect estimates from column 6. Standard errors in parentheses are robust to within family clustering and heteroskedasticity. Control variables include dummy variables for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Source: Administrative registers from Statistics Norway.

As shown in Angrist and Krueger (1999), the linear OLS estimator can be decomposed into weighted averages of the marginal effects. Panel A in Table 4 reports the weights for the linear OLS estimator of family size. Given the distribution of family sizes in Norway, where most families have between 2 and 3 children, the OLS estimator places much more weight on the marginal effects of moving from 1 to 2 siblings and 2 to 3 siblings than on other margins. The nonmonotonic distribution of marginal family size effects and these particular OLS weights yield the near zero linear OLS estimate.

TABLE 4. Linear OLS and IV weights on marginal family size effects.

	Sibling Margin				
	0–1	1–2	2–3	3–4	4–5
Panel A sample:					
All families					
All birth orders					
OLS weight	0.110	0.336	0.313	0.175	0.066
Panel B sample:					
Family size ≥ 2					
First birth					
OLS weight	–	0.444	0.344	0.154	0.053
IV (instr.: twin2) weight	–	0.763	0.170	0.050	0.016
Panel C sample:					
Family size ≥ 3					
2nd birth					
OLS weight	–	–	0.547	0.333	0.119
IV (instr.: twin3) weight	–	–	0.851	0.122	0.027
Panel D sample:					
Family size ≥ 4					
3rd birth					
OLS weight	–	–	–	0.678	0.322
IV (instr.: twin4) weight	–	–	–	0.882	0.117

Note: This table reports the weights for the linear OLS and IV estimator, for simplicity, with no controls. The formula for the linear OLS and IV weights is given in Angrist and Krueger (1999) and Angrist and Imbens (1995). Panel A computes the weights of the marginal effects for the linear OLS estimate reported Table 3. Similarly, panels B, C, and D compute the weights of the linear OLS estimates reported in columns 2, 3, and 4 of panel I in Table 5, and the weights of the linear IV estimates reported in column 1 of panel I in Tables 6, 7, and 8. Source: Administrative registers from Statistics Norway.

4.3 *Total effects versus marginal effects*

The last column of Table 3 reports results from the unrestricted model in family size that is used by Black, Devereux, and Salvanes (2005) in their OLS estimation.[3] This model replaces the marginal effects specification with a total effects specification for birth order and family size,

$$y_i = \psi_1 D_{1i} + \cdots + \psi_{\bar{s}} D_{\bar{s}i} + X_i'\delta + \varepsilon_i, \tag{5}$$

where $D_{si} = 1\{s_i = s\}$. This dummy variables construction implies that the ψ_s coefficients provide the *total* effect of family size from having s siblings rather than 0. We think of ψ_s as the "reduced form" analog of $\Omega(N, 1)$ from (2) in the QQ model discussed above.

Although both dummy variable specifications (4) and (5) are unrestricted in family size as they fully saturate the support of the family size variable, the difference in con-

[3]This replicates the OLS results of Black, Devereux, and Salvanes (2005, p. 679, Table IV, column 6) when they use their unrestricted total effects model in family size, including controls for birth order and other confounding characteristics.

struction is important for interpretation.[4] We find that the total effects relative to only children are generally positive and declining as the number of siblings increases: The total effects decline from 0.224 for 1 vs. 0 siblings to −0.002 for 5 vs. 0 siblings. From these results, Black, Devereux, and Salvanes (2005) conclude that the almost zero effect of family size from the linear OLS estimate is strengthened by the small coefficients on the family size dummy variables, many of which are now statistically insignificant. However, it is the marginal effects—which are always significant and mostly sizable—that provide the relevant comparison to the linear estimate and the appropriate test of the QQ model. Although the linear estimate suggests no effect of family size on child outcome, the marginal family size effects estimates can be interpreted as suggesting a trade-off between quantity and quality in large families, and potential complementarities between quantity and quality in small families

4.4 *Relative importance of birth order versus family size*

We next investigate the relative importance of birth vs. family size by examining the marginal birth order effects. An important and much cited finding of Black, Devereux, and Salvanes (2005) is the large birth order effects on children's education. Because the marginal birth order effects are monotonically negative, total effect specification overstates the effect of being one birth parity later in the birth order compared to increasing family size by 1 child. As is clear from column 6 of Table 3, if birth order matters, so does family size. For example, the effect of having 2 children in the family instead of being the only child (0.224) actually exceeds every marginal birth order effect, except for the effect of being second instead of first born (−0.373). Moreover, the effect of having 4 instead of 3 children in the family (−0.073) is only slightly lower than the effect of being born fourth rather than third (−0.100), although the effect of having 5 instead of 4 children in the family (−0.089) is actually more than twice as large as the effect of being fifth instead of fourth born (−0.040).

Table 3 also shows that the conclusion that birth order effects appear to drive the observed negative relationship between family size and child education does not hold once the linear specification in family size is relaxed. In fact, including the birth order effects actually boosts the positive effect of having 2 children in the family instead of being the only child.

4.5 *Results by birth order*

Table 5 reports results from the linear family size model (3) and the unrestricted model in family size (4) when estimated separately by birth order. Every model estimated in this table includes the full set of demographic controls. The top panel of Table 5 estimates the linear family size model, whereas the bottom panel estimates the unrestricted model in family size. Contrasting the estimates from the two types of models for each birth order

[4] Estimates of marginal effects that are numerically equivalent as those from model (4) can of course be deduced by differencing the total effect estimates: $\gamma_s = \psi_s - \psi_{s-1}$ for all $s > 0$ and $\gamma_1 = \psi_1$.

TABLE 5. OLS estimates of marginal effects by birth order for linear and unrestricted models.

	Birth Order				
	1	2	3	4	5
	Panel I: Linear Estimates of Marginal Effects				
No. of children	0.0001	−0.020	−0.037	−0.037	−0.006
	(0.003)	(0.004)	(0.007)	(0.013)	(0.033)
	Panel II: Unrestricted Estimates of Marginal Effects				
Siblings ≥ 1	0.245				
	(0.009)				
Siblings ≥ 2	−0.021	0.081			
	(0.007)	(0.008)			
Siblings ≥ 3	−0.086	−0.096	−0.010		
	(0.011)	(0.010)	(0.012)		
Siblings ≥ 4	−0.157	−0.091	−0.055	−0.010	
	(0.019)	(0.019)	(0.018)	(0.020)	
Siblings ≥ 5	−0.107	−0.072	−0.102	−0.091	−0.006
	(0.033)	(0.032)	(0.0301)	(0.031)	(0.033)

Note: Each column of each panel is a separate regression. Siblings ≥ 1 is the marginal effect from moving from 0 to 1 siblings, siblings > 2 is the marginal effect from moving from 1 to 2 siblings, and so on. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors in parentheses are heteroskedastic robust, but clustering is not necessary given that regression includes only 1 child from each family. Source: Administrative registers from Statistics Norway.

indicates the extent to which the linear model approximates the underlying relationship between family size and child education. Figures 2 and 3 graph the predicted average child education from the models using the regression estimates reported in Table 5. The figures present educational attainment relative to only children, whose average educational attainment is normalized to 0.

For each of the birth order subsamples, the coefficients on the main diagonal of Table 5 indicate the marginal effect of the first sibling on the youngest child in the family (e.g., the marginal effect on the first born child moving from 1 to 2 children, the marginal effect on the second born from moving from 2 to 3 children, and so on). The OLS estimates indicate that this marginal next child has a positive effect on first and second born children and a small negative (but insignificant) effect for later born children.[5] For each of the birth orders, the linear family size specification underestimates the negative effect of additional children beyond the marginal next child. Examining Figure 2, it is clear that the contrast between the linear and unrestricted specifications is particularly stark for the subsample of first born children. Although the linear OLS specification predicts that additional children have a zero impact on first born children, the unrestricted specification predicts significant negative effects of having more than 1 sibling. Adding

[5]One interpretation of this result for first and second born children is that the birth of an additional child benefits the existing youngest child because this child learns from interacting with or teaching the younger sibling. Another interpretation is that parents are uncertain about the quality of their children and the realization of a high quality child induces them to have an additional child.

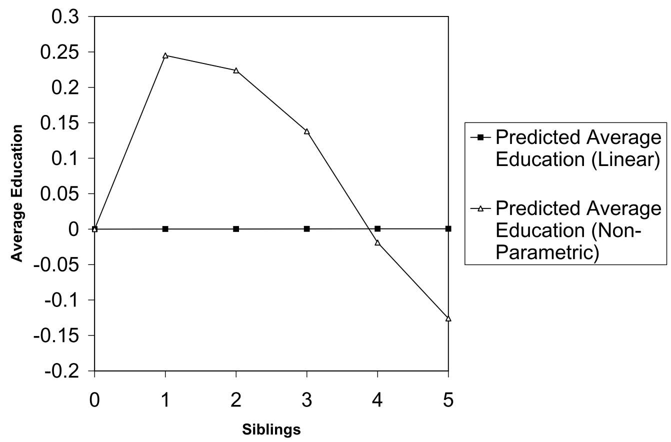

FIGURE 2. Average educational attainment for first born children by number of siblings (relative to only children). *Note*: This figure graphs the linear and unrestricted in family size predictions from OLS regressions (from Table 5). These values are graphed relative to only children (0 siblings), that is, the education of only children is normalized to 0. The slopes in this figure provide the estimated marginal family size effects at each margin, where the linear model imposes constant slopes, whereas the unrestricted model allows nonconstant slopes. The marginal effect estimate from the linear model is close to zero (represented by a flat line), although the unrestricted estimate of the marginal effects indicates that they are nonmonotonic. The linear prediction of total effect is $\hat{y} = \hat{\beta} * s$ for $s = 0, 1, \dots, 5$, where s is number of siblings and $\hat{\beta}$ is the OLS estimate from the first panel of Table 5. The unrestricted prediction is $\hat{y} = \hat{\gamma}_1 * 1\{s \geq 1\} + \cdots + \hat{\gamma}_5 * 1\{s = 6\}$, where $\hat{\gamma}_s$ are the OLS estimates from the second panel of Table 5. Source: Administrative registers from Statistics Norway.

a third sibling is estimated to reduce educational attainment of first born children by 0.086 years, adding a fourth sibling reduces education an additional 0.16 years, and a fifth sibling reduces education an additional 0.11 years.

5. IV ESTIMATES

This section focuses on the second of the two empirical strategies employed by the previous literature to estimate the effects of family size on child outcome: IV estimation using twin births as instruments for family size.[6]

[6]Rosenzweig and Zhang (2009) argue that twin birth affects child spacing and child health in a manner that seems likely to accentuate any negative effect of family size. Angrist, Lavy, and Schlosser (2010) examine carefully this threat to instrument validity, finding no evidence suggesting biases in IV estimates using twin instruments. In any case, the argument of Rosenzweig and Zhang (2009) does not explain our findings of a nonmonotonic relationship between family size and child outcome; neither does it rationalize our evidence of the positive effect of family size in small families. An alternative instrument is sex composition among siblings, but recent literature suggests that it might have direct effects on children's outcomes (see, e.g., Dahl and Moretti (2008)).

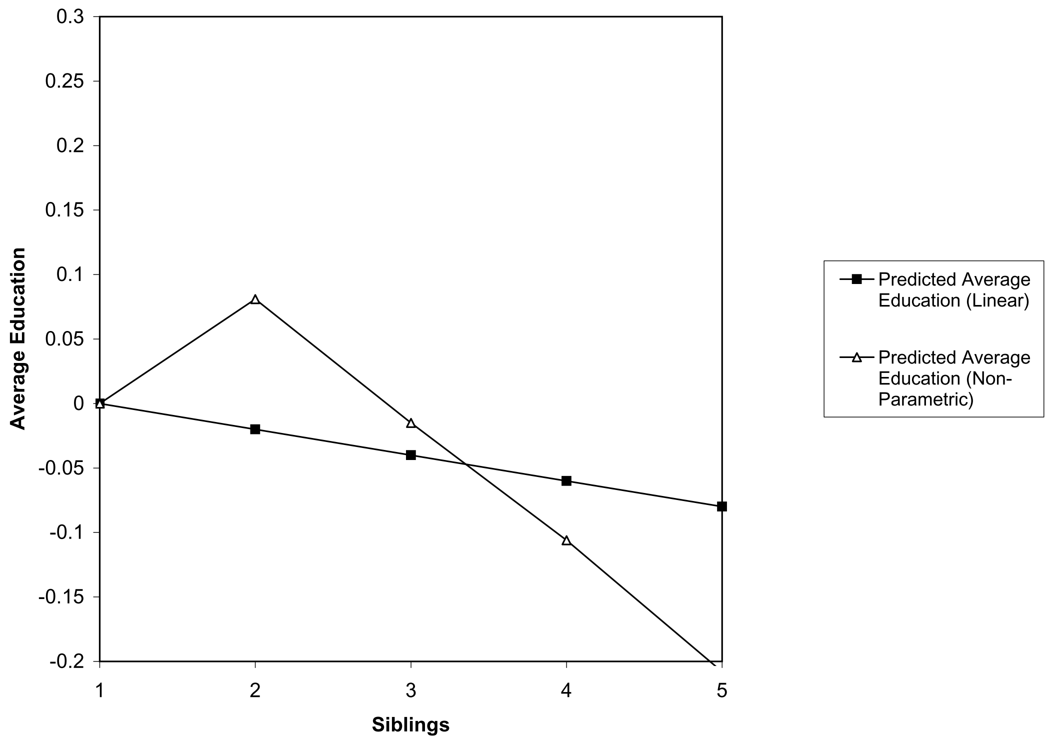

FIGURE 3. Average educational attainment for second born children by number of siblings (relative to only children). *Note*: This figure graphs the linear and unrestricted in family size predictions from OLS regressions (from Table 5). These values are graphed relative to second born children in 2 child families (1 sibling), that is, the education of second born children in 2 child families is normalized to 0. The slopes in this figure provide the estimated marginal family size effects at each margin, where the linear model imposes constant slopes, whereas the unrestricted model allows nonconstant slopes. The linear prediction is $\hat{y} = \hat{\beta} * (s-1)$ for $s = 1, 2, \dots, 5$, where s is number of siblings and $\hat{\beta}$ is the OLS estimate from the first panel of Table 5. The unrestricted prediction is $\hat{y} = \hat{\gamma}_1 * 1\{s \geq 2\} + \cdots + \hat{\gamma}_5 * 1\{s = 5\}$, where $\hat{\gamma}_s$ are the OLS estimates from the second panel of Table 5. Source: Administrative registers from Statistics Norway.

5.1 *Linear IV models*

We follow the previous literature in restricting the sample to children born before the twin birth so as to avoid including the endogenously selected outcomes of children born after the twin birth as well as of twins themselves. We estimate the following linear IV models:

Model 1. Sample of first born children from families with 2 or more children ($c_i \geq 2$):

$$y_i = \beta s_i + X_i'\delta + \varepsilon_i \quad \text{(second stage)},$$
$$s_i = \lambda \text{twin}_{2i} + X_i'\rho + \eta_i \quad \text{(first stage I)},$$

where twin_{2i} is a dummy variable for whether the second birth was a twin birth (implying that second and third born children are twins).

Model 2. Sample of second born children from families with 3 or more children ($c_i \geq 3$):

$$y_i = \beta s_i + X_i'\delta + \varepsilon_i \quad \text{(second stage)},$$
$$s_i = \lambda \text{twin}_{3i} + X_i'\rho + \eta_i \quad \text{(first stage I)},$$

where twin_{3i} is a dummy variable for whether the third birth was a twin birth (implying that third and fourth born children are twins).

Model 3. Sample of third born children from families with 4 or more children ($c_i \geq 4$):

$$y_i = \beta s_i + X_i'\delta + \varepsilon_i \quad \text{(second stage)},$$
$$s_i = \lambda \text{twin}_{4i} + X_i'\rho + \eta_i \quad \text{(first stage I)},$$

where twin_{4i} is a dummy variable for whether the fourth birth was a twin birth (implying that fourth and fifth born children are twins).

In addition to standard regularity conditions and the existence of a first stage, there are two sufficient conditions for consistent IV estimation of the β parameter. The first assumption states that the regression error is mean-independent of the covariates, so that s_i is the only potentially endogenous variable. The second assumption implies that twin birth is conditionally random and affects existing children only through changes in family size. When considering the sample with $\tilde{c}$ or more children, these assumptions can be expressed as

$$E[\varepsilon_i | X_i, c_i \geq \tilde{c}] = E[\varepsilon_i | c_i \geq \tilde{c}] = 0, \tag{6}$$
$$E[\varepsilon_i | X_i, c_i \geq c, \text{twin}_{ci}] = E[\varepsilon_i | X_i, c_i \geq c] \quad \text{for all } c \geq \tilde{c}, \tag{7}$$

where $E[\varepsilon_i | c_i \geq \tilde{c}] = 0$ follows from the standard mean zero normalization of the ε_i error for each of the regression models.[7]

Because we follow the existing literature in restricting the sample to children born before the twin birth, we do not have a twin birth instrument for the 0 to 1 sibling margin. Alternative instruments that induce families to increase family size from 1 to 2 children could be used to instrument for the 0 to 1 sibling margin. For example, Qian (2008) uses the nonuniform application of the one child policy in China to study the effects of having a sibling on child outcome. Interestingly, she finds a positive effect on first born children of an increase in family size from 1 to 2 children, which conforms with our OLS results.

5.2 *What linear IV identifies*

All of the above IV models impose a linear relationship between family size and child outcome. For example, Model 1 restricts the marginal effect of increasing family size from 2 to 3 siblings to be the same as the marginal effect of increasing family size from 3 to 4 siblings, and so on. As shown in Angrist and Imbens (1995), the linear IV estimator

[7]Note that mean independence of covariates is not necessary for nonparametric IV estimation of family size effects (i.e., an IV estimate of β for each $X_i = x$ cell). However, in the empirical literature, which we follow here, the main model (second stage) is specified as an additively separable function of a vector of X_i covariates.

can be decomposed into a weighted average of underlying marginal effects, where the linear IV estimator assigns more weight on the marginal effects, where the cumulative distribution function of family size is more affected by the particular instrument chosen. For instance, IV estimation of Model 1 identifies the marginal effect of moving from 2 to 3 children if a twin on second birth (twin_{2i}) only affects the probability of having 3 instead of 2 children (d_{2i}).

To interpret the linear IV estimates, it is therefore useful to know how twin births shift the fertility distribution. The rationale for using twins as instruments is that for some families, twin births increase the number of siblings beyond the desired family size. A particular feature of the twin instrument is that there are no "never-takers" at the parity of occurrence; everyone who has a twin at the nth parity will at least have $n + 1$ children. However, it is not given that twin birth only affects family size at the parity of occurrence. For example, twin birth at the second parity may exogenously shift the probability that a family has 4 or more children due to complementarities in preferences or economies of scale in the child quality production.

Ultimately, how twin birth shifts the fertility distribution is an empirical question. As shown by Angrist and Imbens (1995), this question can be answered by computing the proportion of families who, because of the twin instrument, change their family size from less than n children to n or more children. Panels B, C, and D in Table 4 calculate the IV weights for β in the IV estimation of Models 1–3. As expected, using twins at second birth as the instrument weights the 2 to 3 children margin most heavily (76 percent), but also places considerable weight on the marginal effects at higher parities (24 percent). A similar pattern is evident for the other twin birth instruments. Consequently, the linear IV estimators of Models 1–3 are weighted averages of several marginal effects, and the estimators differ both in terms of which marginal effects they capture and how much weight they assign to a given marginal effect.

5.3 *Unrestricted IV model*

The sensitivity of the OLS results to the choice between a linear and an unrestricted model in family size underscores that we need to be cautious in using the linear IV models to test the QQ model. We therefore depart from the previous literature in relaxing the assumption of constant marginal effects of family size in the IV estimation. Note that we do not estimate outcomes as a fully nonparametric function in both covariates X_i and family size s_i. We build on the current empirical literature by relaxing the linear restriction for the main variable of interest: family size. We specify unrestricted models in family size by replacing the linear family size variables in the second stages of Models 1–3 with a set of dummy variables for each number of siblings.

Model 4. Sample of first born children from families with 2 or more children ($c_i \geq 2$):

$$y_i = \gamma_2 d_{2i} + \gamma_3 d_{3i} + \gamma_4 d_{4i} + \gamma_5 d_{5i} + X_i'\delta + \varepsilon_i \quad \text{(second stage)}.$$

Model 5. Sample of second born children from families with 3 or more children ($c_i \geq 3$):

$$y_i = \gamma_3 d_{3i} + \gamma_4 d_{4i} + \gamma_5 d_{5i} + X_i'\delta + \varepsilon_i \quad \text{(second stage)}.$$

Model 6. Sample of third born children from families with 4 or more children ($c_i \geq 4$):

$$y_i = \gamma_4 d_{4i} + \gamma_5 d_{5i} + X_i'\delta + \varepsilon_i \quad \text{(second stage)}.$$

In these unrestricted models in family size there are several endogenous explanatory variables that need to be instrumented for. This raises two issues with regard to the specification of the first stages.[8]

First, identification of the unrestricted models in family size requires at least as many instruments as endogenous family size dummy variables. In Model 4, for example, our strategy is to identify $\gamma_2, \dots, \gamma_5$ using the full set of twin birth instruments, $\text{twin}_{2i}, \dots, \text{twin}_{5i}$. However, because of the nature of the twin birth instrument, the full set of instruments is not observed for the entire sample. In particular, twin_{ci} is only defined for families with at least c children. For example, for children from families with only 2 children, whether the family experienced a twin birth on the third (or higher) birth is simply not defined. By using a linear IV estimator, previous studies sidestep this problem of partially missing instruments, because the linear models are identified from a single instrument that is observed for the entire sample. For example, twin_{2i} is sufficient to identify β in Model 1, given the linearity restriction. As discussed below, we address the issue of partially missing instruments by using the strategy proposed by Angrist, Lavy, and Schlosser (2010) and extended in Mogstad and Wiswall (2012). The method allows us to construct valid instruments defined for the entire sample under the same assumptions as used in the linear IV estimation of Models 1–3.

Second, by restricting the number of endogenous explanatory variables that need to be instrumented for, Black, Devereux, and Salvanes (2005) produce sufficiently precise linear IV estimates to conclude that they can rule out large negative effects of family size. However, when relaxing the linearity restriction and performing IV estimation of Models 4–6, we can no longer rule out large effects of family size at conventional significance levels. Instead of settling for the inconclusive evidence from these imprecise IV estimates, we try to gain precision by exploiting the binary nature of the family size dummy variables as well as the unequivocal effect a twin birth has on adding another child to the family. As discussed below, imposing this structure generates sufficient precision in the IV estimation of Models 4–6 and, moreover, this alternative IV strategy produces estimates of the family size effects that are consistent under the same assumptions as the linear IV estimation of Models 1–3.

5.4 *Using the full set of instruments*

Consider using twin births on the second through fifth births $\text{twin}_{2i}, \dots, \text{twin}_{5i}$ as instruments for the four endogenous explanatory variables $d_{2i}, \dots, d_{5i}$ in Model 4. For $s = 2, 3, 4, 5$, the first stages would then be given by

$$d_{si} = \lambda_{s2}\text{twin}_{2i} + \lambda_{s3}\text{twin}_{3i} + \lambda_{s4}\text{twin}_{4i} + \lambda_{s5}\text{twin}_{5i} + X_i'\rho_s + \eta_{si}. \tag{8}$$

[8]It should be noted that one of the standard issues with nonparametric IV is avoided in our family size application. A considerable literature discusses how to nonparametrically estimate a model $y_i = f(s_i)$ using IV, where $f(\cdot)$ is an unknown function of the endogenous variables (e.g., Horowitz (2009), Newey and Powell (2003)). In our family size application, however, the support of s_i is discrete with only a few values and, hence, we can specify a known nonparametric $f(\cdot)$ function without any loss of generality.

However, (8) is not feasible because the twin birth instruments twin_{3i}, twin_{4i}, and twin_{5i} are "undefined" or "missing" for some families. For example, for children from families with only 2 children, whether the family experienced twins on the third birth is not defined.

A naive approach to deal with the problem of partially missing instruments would be to "fill in" the missing twin instruments with zeros (or any arbitrary constant). Suppose we construct instruments defined for the entire sample as

$$z_{ci} = \begin{cases} 0, & \text{if } c_i < c, \\ \text{twin}_{ci}, & \text{if } c_i \geq c. \end{cases}$$

For $s = 2, 3, 4, 5$, the infeasible first stages defined by (8) can then be replaced with the feasible first stages

$$d_{si} = \lambda_{s2}\text{twin}_{2i} + \lambda_{s3} z_{3i} + \lambda_{s4} z_{4i} + \lambda_{s5} z_{5i} + X_i'\rho_s + \eta_{si}.$$

However, this IV strategy would not produce consistent estimates of $\gamma_2, \dots, \gamma_5$ because the constructed instruments are functions of the endogenous family size variables. To see this, note that these instruments can be written as $z_{ci} = 1\{c_i \geq c\}\text{twin}_{ci}$ for $c = 3, 4, 5$.

Angrist, Lavy, and Schlosser (2010) propose an IV estimator in which instruments are formed from a linear projection of the twin instruments on included covariates X. As shown in Mogstad and Wiswall (2012), this estimator is not consistent unless strong auxiliary assumptions are met. They propose an alternative estimator that is robust to violations of the assumptions. Following Mogstad and Wiswall (2012), we construct instruments defined for the full sample as

$$\text{twin}^*_{ci} = \begin{cases} 0, & \text{if } c_i < c, \\ \text{twin}_{ci} - \hat{E}[\text{twin}_{ci} | X_i, c_i \geq c], & \text{if } c_i \geq c, \end{cases}$$

where twin^*_{ci}, as define above, is distinct from the actual twin birth indicator $\text{twin}_{ci} \in \{0, 1\}$. The term $\hat{E}[\text{twin}_{ci}|X_i, c_i \geq c]$ is a initial stage nonparametric estimator for the conditional mean of the instrument (probability of twin birth) in the subsample where it is nonmissing. In Appendix A, we show that the twin^*_{ci} instruments are valid under the same assumption as in Models 1–3.

To be specific, we use twin^*_{ci} as instruments to construct the following first stage specifications for Models 4–6.

Model 4. Sample of first born children from families with 2 or more children ($c_i \geq 2$), for which twin_{2i} is nonmissing, whereas twin_{3i}, twin_{4i}, and twin_{5i} are missing:

$$\begin{aligned} d_{si} = {} & \lambda_{s2}\text{twin}_{2i} + \lambda_{s3}\text{twin}^*_{3i} + \lambda_{s4}\text{twin}^*_{4i} \\ & + \lambda_{s5}\text{twin}^*_{5i} + X_i'\rho_s + \eta_{si}, \quad s = 2, 3, 4, 5 \text{ (first stages I)}. \end{aligned}$$

Model 5. Sample of second born children from families with 3 or more children ($c_i \geq 3$), for which twin_{3i} is nonmissing, whereas twin_{4i} and twin_{5i} are missing:

$$d_{si} = \lambda_{s3}\,\text{twin}_{3i} + \lambda_{s4}\text{twin}^*_{4i} + \lambda_{s5}\text{twin}^*_{5i} + X_i'\rho_s + \eta_{si}, \quad s = 3, 4, 5 \text{ (first stages I)}.$$

Model 6. Sample of third born children from families with 4 or more children ($c_i \geq 4$), for which twin_{4i} is nonmissing, whereas twin_{5i} is missing:

$$d_{si} = \lambda_{s4}\text{twin}_{4i} + \lambda_{s5}\text{twin}^*_{5i} + X_i'\rho_s + \eta_{si}, \quad s = 4, 5 \text{ (first stages I)}.$$

In general, $E[\text{twin}_{ci}|X_i, c_i \geq c]$ is an unknown nonlinear function that needs to be estimated. We estimate the conditional mean using a polynomial function of X_i. The regression includes all of the variables in X_i (a full set of dummy variables for child's birth cohort, mother's and father's age, mother's and father's education, child gender), along with interactions of all parental education levels with parental age and parental age squared. We argue that this rich specification, which includes nearly 200 covariates, provides a reasonable polynomial approximation of the conditional mean function. Given that the main predictor of twinning probabilities is the mother's age at birth, this approximation is particularly well suited to our application because we allow for an unrestricted relationship between mother's age and twinning probabilities. In fact, the additional interaction terms between parental age and education barely move the estimate of $E[\text{twin}_{ci}|X_i, c_i \geq c]$. We also provide a simulation exercise in Appendix B that shows that instruments constructed in this way perform well. Standard errors for the IV estimates are calculated using a clustered (with respect to families) bootstrap procedure to take account of this first stage estimation of the conditional mean function, as described below.

5.5 *Efficient instruments*

Relaxing the linearity restrictions in family size means that we need to instrument for several endogenous family size dummy variables, which turns out to exacerbate the imprecision in the IV estimates. We therefore draw on some well known econometric results on optimal instruments in an attempt to construct more efficient IV estimators. Assumptions (6) and (7) imply that we can use any function of twin_{ci} and X_i to form valid instruments. The optimal (lowest asymptotic variance) instruments are, in general, an unknown function of twin_{ci} and X_i. Newey (1990, 1993) discusses a number of nonparametric estimators for optimal instruments. As an alternative, we impose a particular functional form when constructing our "efficient instruments," so as to address the concern that a higher level of small sample bias may be introduced if we use nonparametric methods and implicitly impose more overidentifying restrictions.

It is important, however, to emphasize that IV estimators using these efficient instruments will be robust to misspecification of the functional form (see, e.g., Newey (1990, 1993)). In particular, our approach is not a parametric control function approach, like the Heckman two-stage method. If the functional form is correct, our efficient instruments are the optimal instruments, and if the functional form is misspecified, our efficient instruments are still consistent under assumptions (6) and (7).

The way we define the efficient instruments exploits two particular features of our family size application: (i) twin births unequivocally increase family size by at least one child and (ii) the endogenous family size dummy variables are binary in nature. In contrast, using the twin birth instruments directly in the first stage specifications, as above, ignores this inherent structure, which may generate a loss in efficiency. Although previous studies of family size and children's outcome have not imposed such structure in the IV estimation, it should be noted that our approach is not novel. Wooldridge (2002) and Carneiro, Heckman, and Vytlacil (2011) provide examples of IV estimation using efficient instruments constructed as we have here. In both applications, they find a substantial improvement in the precision of the IV estimates using the efficient instruments over the IV estimates using the instrument directly. As emphasized by Wooldridge (2002, p. 625), in the case of a binary endogenous variable, as with the family size dummy variables we instrument for here, constructed instruments are "a nice way to way to exploit the binary nature of the endogenous explanatory variable."

To be specific, consider Model 4, where the sample consists of first born children from families with 2 or more children. We define the efficient instrument for the 1 to 2 sibling margin as the predicted probability of having 2 or more siblings, which is given by

$$\hat{p}_{2i} = \begin{cases} 1, & \text{if } \text{twin}_{2i} = 1, \\ f_2(X_i, \hat{\theta}_2), & \text{if } \text{twin}_{2i} = 0. \end{cases}$$

This functional form recognizes that if there are twins on the second birth, then the probability of having at least 2 siblings is, by definition, 1. For a child from a family with a singleton on the second birth, the predicted probability that he or she has 2 or more siblings is specified as a nonlinear function of the included covariates, with an appropriate range restriction to the unit interval: $f_2(X_i, \hat{\theta}_2) \in [0, 1]$, where $\hat{\theta}_2$ are estimates of the unknown parameters of this function. We use the Normal cumulative distribution function (CDF) to restrict the range of the probability and, therefore, estimate the $f_2(X_i, \hat{\theta}_2)$ using a probit model.

In a similar way, we define the efficient instruments for the 2 to 3, the 3 to 4, and the 4 to 5 sibling margins as the predicted probability of having 3 or more, 4 or more, and 5 or more siblings, which is given by

$$\begin{aligned} \hat{p}_{3i} &= f_3(X_i, \text{twin}^*_{3i}, \hat{\theta}_3), \\ \hat{p}_{4i} &= f_4(X_i, \text{twin}^*_{4i}, \hat{\theta}_4), \\ \hat{p}_{5i} &= f_5(X_i, \text{twin}_{2i}, \text{twin}^*_{5i}, \hat{\theta}_5), \end{aligned}$$

where $f_s(\cdot)$ for $s = 3, 4, 5$ includes a linear function of each of the constructed twin instruments that occur after the first birth, in addition to the same covariates as in f_2. Note that we include twin_{2i} in the specification $\hat{p}_{5i}$, given the possibility of "twin exhaustion" in which twin births on the second birth affect future fertility decisions on the fifth birth.

Next, we replace the instruments twin_{2i}, twin^*_{3i}, twin^*_{4i}, and twin^*_{5i} with the efficient instruments $\hat{p}_{2i}$, $\hat{p}_{3i}$, $\hat{p}_{4i}$, and $\hat{p}_{5i}$ in the first stage specifications of Model 4, before applying standard two-stage least squares (2SLS) to estimate the model. In the same way, we construct efficient instruments for Models 5 and 6. This gives us the following, alternative first stage specifications for Models 4–6.

Model 4. Sample of first born children from families with 2 or more children ($c_i \geq 2$):

$$d_{si} = \lambda_{s2}\hat{p}_{2i} + \lambda_{s3}\hat{p}_{3i} + \lambda_{s4}\hat{p}_{4i} + \lambda_{s5}\hat{p}_{5i} + X_i'\rho_s + \eta_{si}, \quad s = 2, 3, 4, 5 \text{ (first stages II)}.$$

Model 5. Sample of second born children from families with 3 or more children ($c_i \geq 3$):

$$d_{si} = \lambda_{s3}\hat{p}_{3i} + \lambda_{s4}\hat{p}_{4i} + \lambda_{s5}\hat{p}_{5i} + X_i'\rho_s + \eta_{si}, \quad s = 3, 4, 5 \text{ (first stages II)}.$$

Model 6. Sample of third born children from families with 4 or more children ($c_i \geq 4$):

$$d_{si} = \lambda_{s4}\hat{p}_{4i} + \lambda_{s5}\hat{p}_{5i} + X_i'\rho_s + \eta_{si}, \quad s = 4, 5 \text{ (first stages II)}.$$

The difference between using the twin birth instruments directly, as in first stages I, and the efficient instruments, as in first stages II, is embedded in the implicit model used to predict the endogenous family size variables d_{si}. To see this, consider Model 4 and note that using first stages I is equivalent to using the first stages

$$d_{si} = \delta_2\tilde{p}_{2i} + \cdots + \delta_5\tilde{p}_{5i} + X_i'\rho + \eta_i, \quad s = 2, 3, 4, 5,$$

where

$$\tilde{p}_{si} = \hat{\kappa}_s \text{twin}_{ci} + X_i'\hat{\omega}_s,$$

and $\hat{\kappa}_s$ and $\hat{\omega}_s$ are the OLS estimate from the OLS regression of d_{si} on twin_{ci}, and X_i in the subsample of children from families with at least c children. This illustrates that when using the twin birth instruments directly, a linear probability model is used to predict the endogenous family size variables d_{si}. In contrast, the IV estimator based on the efficient instruments uses a nonlinear model to predict the endogenous family size variables. This has the advantages of appropriately restricting the range to the unit interval, in addition to taking into account that twin births unequivocally increase family size by 1 child. In doing so, the efficient instruments may be more strongly correlated with the endogenous family size dummy variables, which will improve the efficiency in the IV estimation.

To provide a direct comparison between the results from the linear and unrestricted family size models when using the same set of efficient instruments, we will also use the following first stage specifications for Models 1–3.

Model 1. Sample of first born children from families with 2 or more children ($c_i \geq 2$):

$$s_i = \lambda\hat{p}_{2i} + X_i'\rho + \eta_i \quad \text{(first stage II)}.$$

Model 2. Sample of second born children from families with 3 or more children ($c_i \geq 3$):

$$s_i = \lambda\hat{p}_{3i} + X_i'\rho + \eta_i \quad \text{(first stage II)}.$$

Model 3. Sample of third born children from families with 4 or more children ($c_i \geq 4$):

$$s_i = \lambda \hat{p}_{4i} + X_i' \rho + \eta_i \quad \text{(first stage II)}.$$

In general, the consistency of the IV estimator is unaffected by misspecification of the functional form of the instrument and the asymptotic variance of the IV estimator is unaffected by the initial estimation of θ_s. However, the small sample properties of the IV estimator may depend on whether we use the efficient instruments or the twin instruments directly (see the discussion in Newey (1990, 1993)). Like Angrist, Lavy, and Schlosser (2010), who interact the twin birth instruments with covariates in their study of family size effects, the efficient instruments generate an overidentified IV estimator, which may exacerbate any small sample bias in the IV estimation. We therefore choose a parsimonious specification of the covariates in $f_s(\cdot)$. Specifically, we include (i) linear and quadratic in child's own age, mother's age, and father's age, (ii) 6 intercepts for each level of father's education and 6 intercepts for each level of mother's education, (iii) an intercept for missing father's age, and (iv) an intercept for child's sex. Adding the common intercept, this specification includes 21 unknown parameters.[9]

Given our large samples and first stage results, the literature on small sample bias of the IV estimator suggests that this number of overidentifying restrictions should be of little concern (e.g., Staiger and Stock (1997)). Our simulation exercise reported in Appendix B supports this conjecture. The simulation results show that the small sample bias and small sample variance of the IV estimator using the efficient instruments is smaller than that for the IV estimator using the twin birth instruments directly. That we achieve lower small sample bias in these simulations despite estimating the instruments in a first step and using additional overidentifying restrictions is suggestive that this procedure does not increase the small sample bias of the IV estimator.

5.6 *IV estimates*

Tables 6–8 present IV results for the linear models in family size in panel I (Models 1–3) and the unrestricted models in family size in panel II (Models 4–6). The first stage results are reported in Appendix C. For each model, we present results using the twin birth instruments directly as specified in first stages I (labeled "Standard 2SLS"), and when employing the efficient instruments as specified in first stages II (labeled "Efficient IV").[10]

[9]We have also estimated nonparametric optimal instruments, as suggested by Newey (1993). Specifically, we estimated $E[d_{si}|X_i, \text{twin}_{ci}]$ for each permissible X_i and twin_{ci} cells (both X_i and twin_{ci} have discrete supports). Using the estimated $E[d_{si}|X_i, \text{twin}_{ci}]$ instruments generated precise IV estimates of the nonparametric model in family size, with coefficient estimates similar to those for the nonparametric OLS. However, we are reluctant to report these results, because the very large number of cells implies that this procedure uses many overidentifying restrictions, which could increase the small sample bias of the IV estimation. Our approach here of using a particular nonlinear model and a parsimonious parametric function of the X_i variables is intended to achieve a more reasonable trade-off between bias and variance of the IV estimator. For an in-depth discussion of this issue, see Donald and Newey (2001).

[10]When interpreting the first stage results, there are two things to keep in mind. First, the efficient instruments are estimated functions of the twin instruments and the covariates. The first stage coefficients associated with the efficient instruments may therefore exceed 1: Only in the case where the first stage is

Evidence for efficiency gains in using the efficient IV is found by examining the first stage R-squared values. In all cases, the efficient IV have higher R-squared in the first stage than the standard 2SLS. This demonstrates that the efficient instruments are more strongly correlated with the endogenous family size variables. The gains in R-squared are modest for some IV estimators, but are particularly large for the small probability events, which are probably most affected by the implicit linear probability model used by the standard 2SLS first stage. For instance, in Model 1, the R-squared for the endogenous variable of having more than 4 children is 0.0656 for standard 2SLS but 0.0774 for the efficient IV. This is a gain of nearly 20 percent in explanatory power. The R-squared for the even rarer event of having 5 or more siblings is 0.0423 for standard 2SLS compared to 0.0544 for efficient IV, a gain of nearly 29 percent. Similar efficiency gains are found across the models for small probability events. Given these gains in first stage fit, we would expect the second stage results using the efficient IV to have smaller standard errors (relative to the point estimates), as compared to those based on the standard 2SLS.

Table 6. IV estimates of marginal effects in linear and unrestricted models for first born children in families with 2 or more children.

Panel I: Linear Estimates of Marginal Effects		
Instrument	Standard 2SLS	Efficient IV
No. of children	0.053	−0.0036
	(0.0495)	(0.0460)
Panel II: Unrestricted Estimates of Marginal Effects		
Instrument(s)	Standard 2SLS	Efficient IV
Siblings ≥ 2	0.079	0.153
	(0.067)	(0.063)
Siblings ≥ 3	−0.044	−0.474
	(0.073)	(0.079)
Siblings ≥ 4	0.023	−0.800
	(0.100)	(0.129)
Siblings ≥ 5	−0.051	−0.787
	(0.181)	(0.247)

Note: Each column is a separate regression. Siblings ≥ 1 is the marginal effect from moving from 0 to 1 siblings, siblings > 2 is the marginal effect from moving from 1 to 2 siblings, and so on. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors are calculated using a cluster bootstrap, sampling each family (and all children in this family) with replacement 50 times. For each bootstrap repetition, we recompute the instruments defined for the full sample, the constructed efficient instruments, and the 2SLS estimators that use these instruments.

simply $d_{si} = \delta \text{twin}_{ci} + \eta_i$, where both $d_{si} \in \{0, 1\}$ and $\text{twin}_{ci} \in \{0, 1\}$, would the range of δ be expected to be [0, 1]. Second, recall that the unrestricted model consists of dummy variables for having 2 or more children, 3 or more children, etcetera. Therefore, instruments derived from having a twin on a third birth or higher may very well have a significant effect on the probability of having 2 or more children.

TABLE 7. IV estimates of marginal effects in linear and unrestricted models for second born children in families with 3 or more children.

Panel I: Linear Estimates of Marginal Effects		
Instrument	Standard 2SLS	Efficient IV
No. of children	−0.051	−0.171
	(0.053)	(0.051)
Panel II: Unrestricted Estimates of Marginal Effects		
Instrument(s)	Standard 2SLS	Efficient IV
Siblings ≥ 3	−0.058	−0.090
	(0.068)	(0.068)
Siblings ≥ 4	−0.054	−0.586
	(0.093)	(0.110)
Siblings ≥ 5	0.138	−0.504
	(0.224)	(0.205)

Note: Each column is a separate regression. Siblings ≥ 1 is the marginal effect from moving from 0 to 1 siblings, siblings > 2 is the marginal effect from moving from 1 to 2 siblings, and so on. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors are calculated using a cluster bootstrap, sampling each family (and all children in this family) with replacement 50 times. For each bootstrap repetition, we recompute the instruments defined for the full sample, the constructed efficient instruments, and the 2SLS estimators that use these instruments.

TABLE 8. IV estimates of marginal effects in linear and unrestricted models for third born children in families with 4 or more children.

Panel I: Linear Estimates of Marginal Effects		
Instrument	Standard 2SLS	Efficient IV
No. of children	−0.107	−0.191
	(0.075)	(0.074)
Panel II: Unrestricted Estimates of Marginal Effects		
Instrument(s)	Standard 2SLS	Efficient IV
Siblings ≥ 4	−0.096	−0.145
	(0.090)	(0.093)
Siblings ≥ 5	−0.178	−0.520
	(0.163)	(0.176)

Note: Each column is a separate regression. Siblings ≥ 4 is the marginal effect from moving from 3 to 4 siblings, siblings $= 5$ is the marginal effect from moving from 1 to 2 siblings, and so on. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors are calculated using a cluster bootstrap, sampling each family (and all children in this family) with replacement 50 times. For each bootstrap repetition, we recompute the instruments defined for the full sample, the constructed efficient instruments, and the 2SLS estimators that use these instruments.

Table 6 shows efficient IV and standard 2SLS estimates of the effects of changes in family size on the first born child from families with 2 or more children. As in the previous literature, the first column of panel I shows a linear effect of family size of 0.053, with the lower bound of the 95 percent confidence interval not greater in absolute value than −0.05.[11] Large positive effects of family size cannot be ruled out, as the upper bound of the 95 percent confidence interval is as large as 0.15. As shown above, in terms of the QQ model, positive effects of family size changes can result from complementarities between child quantity and quality.

Relaxing the linearity restriction in family size, the results reported in the first column of panel II in Table 6 reveal that we can no longer reject the hypothesis of larger negative effects of family size from the standard 2SLS estimates. The lower bound of the 95 percent confidence interval is −0.18 for the marginal effect from 2 to 3 siblings, −0.20 for the 3 to 4 margin, and −0.42 for 4 to 5 margin. Note that the lower bounds on these marginal effects are several times larger than the corresponding marginal birth order effects estimated in Table 3. We therefore conclude that the small negative effect of family size found in the linear restricted estimation is not robust to relaxing the linear specification in the IV estimation.

Turning to the efficient IV results of the unrestricted model in family size, reported in the second column of panel II in Table 6, the main finding is that there are significant and large marginal family size effects on children's education. Furthermore, the results indicate a nonmonotonic causal relationship between family size and children's education. For the first born in families with at least 2 children, a third child is estimated to increase completed education by 0.15 years. This estimate is within the 95 percent confidence interval for the corresponding IV estimate reported in first column of panel II. We can also see that changes in family size are estimated to reduce children's education by 0.47 years for a fourth child, another 0.8 years for a fifth child, and an additional 0.79 years for a sixth child. These estimates are several times larger than the corresponding OLS estimates and are outside the lower bound of the 95 percent confidence intervals for the IV estimates reported in the first column of panel II. It should also be noted that these marginal family size effects exceed the marginal birth order effects. In terms of the QQ model, the efficient IV estimates of the unrestricted model in family size indicate a trade-off between quantity and quality in large families, and complementarities between quantity and quality in small families.

Comparing the efficient IV results from the linear and unrestricted models in family size reported in the first column of Table 6, we immediately see the role of the linearity restriction in masking the marginal family size effects. In line with the standard 2SLS estimate of the linear family size model, the linear IV estimate using the efficient instrument is close to zero and imprecise. In contrast, the efficient IV estimates of the unrestricted model—using the *same* type of instruments—are larger and statistically significant at the 95 percent level. Hence, we can conclude that for a given set of instruments, the second stage restriction in family size plays an important role in the conclusion about the effects of family size on child outcome.

[11] Standard errors are calculated using a cluster bootstrap, sampling each family (and all children in this family) with replacement 50 times. For each bootstrap repetition, we recompute the instrument defined for the full sample, the constructed efficient instruments, and the 2SLS estimators that use these instruments.

For later born children, the IV estimates in Table 7 and Table 8 reveal a similar pattern as for the results for first born children. First, relaxing the linearity restriction, we see from the standard 2SLS estimates that we cannot reject the hypothesis of large negative effects of family size for second and third born children. In fact, the 95 percent confidence intervals of the 2SLS estimates of the unrestricted model cover the sizable OLS estimates of the marginal family size effects and are considerably larger than the marginal birth order effects. Second, the efficient IV estimates suggest sizable and significant negative effects of family size for second and third born children. This is true both for the linear and unrestricted models in family size. For example, the linear estimate of the effect of family size on second born children suggests that having another sibling reduces their educational attainment by −0.171. Our interpretation is that this linear estimate reflects the weighted average of the relatively small marginal family size effect of having 4 instead of 3 children (−0.09), and the much larger negative family size effect of having 5 instead of 4 children (−0.59) and 6 instead of 5 children (−0.50).

In general, we find larger point estimates of marginal effects in the unrestricted model using the efficient instruments as compared to those using the twin birth instruments directly. A possible explanation is population heterogeneity in the effects of family size. As emphasized by Heckman, Urzua, and Vytlacil (2006) and others, different valid instruments will, in general, identify different local average treatment effect (LATE). The interactions between covariates and the twin instrument will change the weights assigned to different groups of the population. If there is heterogeneity in effects across these groups, then the 2SLS estimates should differ from the efficient IV estimates. Brinch, Mogstad, and Wiswall (2014) investigate heterogeneity in the effects of family size. Their findings suggest that family size effects vary both in sign and magnitude across individuals. Brinch et al. also show how population heterogeneity may help in explaining the differences in IV estimates that use twin births, sex composition, or their interaction with covariates as instruments for family size.

6. Conclusions

Motivated by the seminal QQ model of fertility by Becker and Lewis (1973), a large and growing body of empirical research has tested the QQ model by examining the relationship between family size and children's outcome. Given the theoretical ambiguity about the magnitude and sign of the marginal effects on child quality of additional siblings, we have explored the implications of allowing for a nonlinear relationship between family size and child outcome when testing the QQ model. We find that the conclusion of no effect of family size in previous studies does not hold up if we relax their linear specification in family size. This is true when we perform OLS estimation with controls for confounding characteristics like birth order and when instrumenting family size with twin births. When estimating models that are unrestricted in family size, we find a nonmonotonic relationship with statistically significant and sizable marginal effects. In terms of the QQ model, this inverse U-shaped pattern suggests a trade-off between quantity and quality in large families and (strong) complementarities in small families.

An understanding of the relationship between family size and children's outcomes can be important from a policy perspective. Most developed countries have a range of policies affecting fertility decisions. Many of these policies are designed such that they reduce the cost of having a single child more than the cost of having two or more children, in effect promoting smaller families. If a policy goal is to slow or reverse the unprecedented fertility decline most developed countries have experienced over the last 30 years, the effects of family size on children's outcomes become ever more important. Accepting the recent findings of no effect of family size suggests there is no need to be concerned with the externalities on the human capital development of existing children when designing policies promoting larger families. Our findings run counter to this conclusion. In fact, the evidence of an inverse U-shaped pattern suggests that an efficient policy is to target incentives for higher fertility to small families, and discourage larger families from having additional children.

However, caution is in order. The IV estimates are only informative about the average causal effect of the instrument induced change in family size. In general, families induced to have another child because of a twin birth may differ from families induced to have another birth by a policy change. In particular, tax and transfer policies would typically affect the household's budget constraint, and households would optimally choose family size considering any number of factors. We therefore need to be cautious in extrapolating the IV estimates to the population at large. Indeed, recent work suggests considerable heterogeneity in the effect of another sibling on existing children, not only at different family size margins as we demonstrate here, but also among families at a particular margin (Brinch, Mogstad, and Wiswall (2014)). Determining the mechanisms through which this heterogeneity arises is important to design effective policies and to understand the breadth and nature of the relationship between family size and child quality.

Appendix A: Deriving the full sample instruments

Below, we show how to derive instruments defined for the full sample. More details are provided in Mogstad and Wiswall (2012).

We consider a simple example, where we have the linear model

$$y_i = \beta s_i + X_i'\delta + \varepsilon_i.$$

Suppose we want to estimate this model for first born children in families with at least 2 children ($c_i \geq 2$). Assume that (6) holds, that is, $E[\varepsilon_i|X_i, c_i \geq 2] = 0$, implying that s_i is the only potentially endogenous regressor.

Consider using twin_{3i} (twin on third birth) as the instrument for s_i. This instrument is partially missing, because twin_{3i} is defined only for the subsample with at least 3 children ($c_i \geq 3$), and is missing for the subsample of children from 2 child families $c_i = 2$. Assume that (7) holds, so that $E[\varepsilon_i|X_i, c_i \geq 3, \text{twin}_{3i}] = E[\varepsilon_i|X_i, c_i \geq 3]$.

The naive "fill in" IV method forms an instrument for full sample as $z_i = 1\{c_i \geq 3\}\text{twin}_{3i}$. This instrument is invalid because

$$\begin{aligned} &E[\varepsilon_i z_i|X_i, c_i \geq 2] \\ &\quad = E[\varepsilon_i \text{twin}_{3i}|X_i, c_i \geq 3]\,\text{pr}(c_i \geq 3|X_i) \\ &\quad = E[\varepsilon_i|X_i, c_i \geq 3, \text{twin}_{3i} = 1]\,\text{pr}(\text{twin}_{3i} = 1|X_i, c_i \geq 3)\,\text{pr}(c_i \geq 3|X_i). \end{aligned}$$

Note that $E[\varepsilon_i|X_i, c_i \geq 3, \text{twin}_{3i}] = E[\varepsilon_i|X_i, c_i \geq 3]$ does *not* imply that this first term is zero. In general,

$$E[\varepsilon_i|X_i, c_i \geq 3, \text{twin}_{3i} = 1] \neq E[\varepsilon_i|X_i, c_i \geq 2] = 0.$$

With fertility endogenously determined, the mean of ε_i will in general be different for the sample of children from 2 child families compared to children from 3 child families.

Our missing IV robust strategy first "demeans" twin_{3i} by subtracting its conditional mean. Define the new instrument twin^*_{3i} as

$$\text{twin}^*_{3i} = 1\{c_i \geq 3\}\big(\text{twin}_{3i} - E[\text{twin}_{3i}|X_i, c_i \geq 3]\big).$$

This instrument is valid because

$$\begin{aligned} &E\big[\varepsilon_i \text{twin}^*_{3i}|X_i, c_i \geq 2\big] \\ &\quad = E\big[\varepsilon_i\big(\text{twin}_{3i} - E[\text{twin}_{3i}|X_i, c_i \geq 3]\big)\big|X_i, c_i \geq 3\big] \\ &\quad = E[\varepsilon_i \text{twin}_{3i}|X_i, c_i \geq 3] - E\big[\varepsilon_i E[\text{twin}_{3i}|X_i, c_i \geq 3]\big|X_i, c_i \geq 3\big]. \end{aligned}$$

Given $E[\varepsilon_i|X_i, c_i \geq 3, \text{twin}_{3i}] = E[\varepsilon_i|X_i, c_i \geq 3]$, we have

$$E[\varepsilon_i \text{twin}_{3i}|X_i, c_i \geq 3] = E[\varepsilon_i|X_i, c_i \geq 3]E[\text{twin}_{3i}|X_i, c_i \geq 3].$$

Substituting,

$$\begin{aligned} &E\big[\varepsilon_i \text{twin}^*_{3i}|X_i\big] \\ &\quad = E[\varepsilon_i|X_i, c_i \geq 3]E[\text{twin}_{3i}|c_i \geq 3] - E\big[\varepsilon_i E[\text{twin}_{3i}|c_i \geq 3, X_i]\big|X_i, c_i \geq 3\big] \\ &\quad = E[\varepsilon_i|X_i, c_i \geq 3]\big\{E[\text{twin}_{3i}|X_i, c_i \geq 3] - E[\text{twin}_{3i}|X_i, c_i \geq 3]\big\} \\ &\quad = 0. \end{aligned}$$

This shows that instruments constructed in this fashion are valid under the same assumptions as the linear IV models. Extension to the unrestricted model in family size is straightforward. The simulation presented below suggests that these IV estimators perform well even in small samples.

Appendix B: Simulation of IV estimator

We use a simulation exercise to examine the small sample properties of the IV estimators using the efficient instruments. We focus on first born children with 1 to 3 siblings (2 to 4 total children). For each first born child, the data consist of a number of siblings $s_i \in \{1, 2, 3\}$, one scalar exogenous covariate x_i (e.g., mother's education), two twin birth instruments twin_{2i} (twin on second birth) and twin_{3i} (twin on third birth), and an observed outcome for the first born child y_i.

Table B.1. Simulation results.

Distributional Assumption:	$\varepsilon_i \sim N(0,1)$		$\varepsilon_i \sim G(1,2)$	
True Parameters:	$\gamma_2 = 1$	$\gamma_3 = -1$	$\gamma_2 = 1$	$\gamma_3 = -1$
(i) OLS				
Mean absolute value of bias	1.19	1.32	0.96	1.93
(ii) IV using twin instruments directly				
Mean absolute value of bias	0.072	0.12	0.25	0.32
Standard deviation of estimates	0.09	0.15	0.32	0.39
Mean squared error	0.016	0.047	0.20	0.31
(iii) IV using efficient instruments				
Mean absolute value of bias	0.046	0.057	0.086	0.10
Standard deviation of estimates	0.063	0.063	0.11	0.13
Mean squared error	0.0064	0.0089	0.023	0.033

Note: Simulation results from 500 replications of the data generating process described.

We specify the following data generating process. In the absence of twin births, the choice of family size takes an ordered choice form with latent utility from children given by $u_i = \alpha x_i + \varepsilon_i$. The number of siblings is selected as $s_i = 1$ if $u_i < \pi_2$, $s_i = 2$ if $\pi_2 \leq u_i < \pi_3$, and $s_i = 3$ if $u_i \geq \pi_3$. The twin birth instruments exogenously increase siblings by one child: $s_i = 2$ if $\text{twin}_{2i} = 1$ and $s_i = 3$ if $\text{twin}_{3i} = 1$. The observed outcome is then $y_i = \gamma_2 d_{2i} + \gamma_3 d_{3i} + \rho x_i + \varepsilon_i$, where $d_{2i} = 1\{s_i \geq 2\}$ and $d_{3i} = 1\{s_i \geq 3\}$. Random variables are distributed $x_i \sim N(1,1)$ and $\text{twin}_{ci} = 1$ with probability 0.05, for $c = 2, 3$. The remaining parameters are set at $\pi_2 = 1$, $\pi_3 = 1.5$, $\alpha = 1$, $\gamma_2 = 1$, $\gamma_3 = -1$, and $\rho = 1$. In this data generating process, the marginal effects of family size are homogeneous across families but nonconstant across margins ($\gamma_2 \neq \gamma_3$).

Table B.1 presents the simulation results for 500 replications. For each replication, we draw a sample of 10,000 observations from the data generating process. We conduct two simulations. The first simulation assumes ε_i is distributed standard Normal: $\varepsilon_i \sim N(0,1)$. The second simulation assumes ε_i is distributed according to the Gamma distribution with a shape parameter of 2 and a scale parameter of 1: $\varepsilon_i \sim G(1,2)$. This parameterization implies that distribution of ε_i has skewness of $2/\sqrt{2}$ and excess kurtosis of 3. By contrast, the Normal distribution has skewness and excess kurtosis of 0.

For each simulated sample, we compute three estimators of the γ_2 and γ_3 parameters: (i) OLS, (ii) IV using the twin birth instruments directly (that is, first stage I in the unrestricted model in family size), and (iii) IV using the efficient instruments (that is, first stage II in the unrestricted model in family size). In both IV estimations, we deal with the missing instrument problem for twin_{3i} as discussed in Section 5.3. The efficient instruments are constructed as described in Section 5.4.

The results in Table B.1 display several finite sample characteristics for each estimator. Across the $R = 500$ replications of the data generating process, we calculate the mean of the absolute bias for each parameter: $\frac{1}{R}\sum_{r=1}^{R} |\hat{\gamma}_{sr} - \gamma_s|$ for $s = 1, 2$, where γ_s is the true parameter and $\hat{\gamma}_{sr}$ is the rth simulation estimate. We also calculate the standard deviation of the estimates across the simulations: $\sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{\gamma}_{sr}-\bar{\hat{\gamma}}_s)^2}$, where $\bar{\hat{\gamma}}_s$ is the mean of the estimates across the simulations. Finally, we calculate mean squared error as the variance in the estimators across the replications plus the mean squared bias.

For each parameter and error distribution assumption, the OLS estimator is severely biased with the mean absolute value of bias around 1 or higher. All the IV estimators have substantially lower levels of bias than the OLS estimators. However, the IV estimators using the twin births directly have higher levels of bias, higher variance, and higher mean squared error than the IV estimators using the efficient instruments. This is true across parameters and assumptions about the distribution of the error. When the ε_i follows a Gamma distribution that is highly non-Normal, the finite sample bias is larger than when the ε_i distribution is Normal. However, the finite sample bias increases for the IV estimator using the twin birth instruments directly as well, and the finite sample bias is still smaller for the IV using the efficient instruments compared to the IV using the twin birth instruments directly.

This simulation indicates that using a misspecified probit model to generate the instruments does not introduce any larger degree of finite sample bias relative to the more standard IV estimation using linear functions of the instruments. Both when the simulation assumes a Normal distribution for the error terms and when using a Gamma distribution with a high degree of skewness, the unrestricted IV estimators based on the efficient instruments have lower average absolute value of bias and lower variance across simulations, relative to the unrestricted IV estimator using twin birth instruments directly.

Appendix C: First stage results

Table C.1. First stage for Table 6, panel I.

Instrument	twin_{2i}	R^2
No. of children	0.684 (0.012)	0.1213
Instrument	$\hat{p}_{2i}$	R^2
No. of children	1.30 (0.016)	0.1215

Note: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 6. The R^2 is the first stage total R-squared. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors in parentheses are robust to heteroskedasticity, but clustering is not necessary given that each regression includes only 1 child from each family.

Table C.2. First stage for Table 6, panel II.

Instrument	twin_{2i}	twin^*_{3i}	twin^*_{4i}	twin^*_{5i}	R^2
Siblings ≥ 2	0.518	0.010	−0.013	−0.005	0.0969
	(0.007)	(0.009)	(0.015)	(0.028)	
Siblings ≥ 3	0.127	0.661	0.018	−0.017	0.1003
	(0.005)	(0.018)	(0.011)	(0.021)	
Siblings ≥ 4	0.033	0.092	0.708	0.024	0.0656
	(0.003)	(0.004)	(0.007)	(0.013)	
Siblings ≥ 5	0.010	0.017	0.090	0.756	0.0423
	(0.002)	(0.002)	(0.003)	(0.007)	
Instrument	$\hat{p}_{2i}$	$\hat{p}_{3i}$	$\hat{p}_{4i}$	$\hat{p}_{5i}$	R^2
Siblings ≥ 2	1.012	0.134	0.173	0.191	0.0983
	(0.013)	(0.013)	(0.021)	(0.040)	
Siblings ≥ 3	0.209	0.960	0.459	0.550	0.1053
	(0.010)	(0.010)	(0.016)	(0.031)	
Siblings ≥ 4	0.048	0.079	1.08	0.874	0.0774
	(0.006)	(0.006)	(0.009)	(0.018)	
Siblings ≥ 5	0.014	−0.001	0.112	1.30	0.0544
	(0.003)	(0.003)	(0.005)	(0.009)	

Note: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 6. The R^2 is the first stage total R-squared. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors in parentheses are robust to heteroskedasticity, but clustering is not necessary given that each regression includes only 1 child from each family.

Table C.3. First stage for Table 7, panel I.

Instrument	twin_{3i}	R^2
No. of children	0.763	0.1191
	(0.014)	
Instrument	$\hat{p}_{3i}$	R^2
No. of children	1.23	0.1206
	(0.014)	

Note: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 7. The R^2 is the first stage total R-squared. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors in parentheses are robust to heteroskedasticity, but clustering is not necessary given that each regression includes only 1 child from each family.

TABLE C.4. First stage for Table 7, panel II.

Instrument	$twin_{3i}$	$twin^*_{4i}$	$twin^*_{5i}$	R^2
Siblings ≥ 3	0.648 (0.009)	0.020 (0.014)	−0.016 (0.027)	0.1110
Siblings ≥ 4	0.103 (0.006)	0.703 (0.009)	0.023 (0.178)	0.0878
Siblings ≥ 5	0.020 (0.003)	0.100 (0.005)	0.756 (0.010)	0.0534
Instrument	$\hat{p}_{3i}$	$\hat{p}_{4i}$	$\hat{p}_{5i}$	R^2
Siblings ≥ 3	1.01 (0.0133)	0.289 (0.019)	0.361 (0.036)	0.1141
Siblings ≥ 4	0.131 (0.009)	1.01 (0.013)	0.681 (0.024)	0.0964
Siblings ≥ 5	0.018 (0.005)	0.114 (0.007)	1.20 (0.013)	0.0642

Note: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 7. The R^2 is the first stage total R-squared. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors in parentheses are robust to heteroskedasticity, but clustering is not necessary given that each regression includes only 1 child from each family.

TABLE C.5. First stage for Table 8, panel I.

Instrument	$twin_{4i}$	R^2
No. of children	0.786 (0.019)	0.1066
Instrument	$\hat{p}_{4i}$	R^2
No. of children	1.16 (0.016)	0.1080

Note: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 8. The R^2 is the first stage total R-squared. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors in parentheses are robust to heteroskedasticity, but clustering is not necessary given that each regression includes only 1 child from each family.

Table C.6. First stage for Table 8, panel II.

Instrument	twin_{4i}	twin^*_{5i}	R^2
Siblings ≥ 4	0.693	0.026	0.1059
	(0.014)	(0.026)	
Siblings ≥ 5	0.101	0.757	0.0731
	(0.008)	(0.015)	
Instrument	$\hat{p}_{4i}$	$\hat{p}_{5i}$	R^2
Siblings ≥ 4	1.016	0.313	0.1088
	(0.013)	(0.019)	
Siblings ≥ 5	0.140	1.06	0.0790
	(0.009)	(0.013)	

Note: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 8. The R^2 is the first stage total R-squared. All models include covariates for gender, child's age (in 2000), mother's age (in 2000), father's age (in 2000), mother's education, and father's education. Standard errors in parentheses are robust to heteroskedasticity, but clustering is not necessary given that each regression includes only 1 child from each family.

References

Aaslund, O. and H. Grønquist (2010), "Family size and child outcomes: Is there really no trade-off." *Labour Economics*, 17 (1), 130–139. [158, 160]

Angrist, J. D. and G. Imbens (1995), "Two-stage least squares estimation of average causal effects in models with variable treatment intensity." *Journal of the American Statistical Association*, 90 (430), 431–442. [167, 172, 173]

Angrist, J. D. and A. B. Krueger (1999), "Empirical strategies in labor economics." In *Handbook of Labor Economics*, Vol. 3 (O. Ashenfelter and D. Card, eds.), 1277–1366, Elsevier, Amsterdam. [166, 167]

Angrist, J. D., V. Lavy, and A. Schlosser (2010), "Multiple experiments for the causal link between the quantity and quality of children." *Journal of Labor Economics*, 28, 773–824. [158, 160, 170, 174, 175, 179]

Bandura, A. (1977), *Social Learning Theory*. Prentice Hall, Englewood Cliffs, NJ. [162]

Becker, G. S. (1998), *A Treatise on the Family. Enlarged Version*. Harvard University Press, Cambridge, MA. [162]

Becker, G. S. and H. G. Lewis (1973), "On the interaction between the quantity and quality of children." *Journal of Political Economy*, 81 (2), 279–288. [158, 159, 183]

Black, S. E., P. J. Devereux, and K. G. Salvanes (2005), "The more the merrier? The effects of family size and birth order on children's education." *Quarterly Journal of Economics*, 120, 669–700. [157, 158, 160, 162, 163, 165, 167, 168, 174]

Brinch, C., M. Mogstad, and M. Wiswall (2014), "Beyond LATE with a discrete instrument." Working paper. [183, 184]

Caceres-Delpiano, J. (2006), "The impacts of family size on investment in child quality." *Journal of Human Resources*, 41 (4), 738–754. [158, 160]

Carneiro, P., J. Heckman, and E. Vytlacil (2011), "Estimating marginal and average returns to education." *American Economic Review*, 101 (6), 2754–2781. [177]

Dahl, G. and E. Moretti (2008), "The demand for sons." *Review of Economic Studies*, 75 (4), 1085–1120. [170]

Del Boca, D. and C. Wetzels (2008), *Social Policies, Labour Markets and Motherhood: A Comparative Analysis of European Countries*. Cambridge University Press, Cambridge. [158]

Donald, S. G. and W. K. Newey (2001), "Choosing the number of instruments." *Econometrica*, 69 (5), 1161–1191. [179]

Feyrer, J., B. Sacerdote, and A. Stern (2008), "Will the stork return to Europe and Japan? Understanding fertility within developed nations." *Journal of Economic Perspectives*, 22 (3), 3–22. [157]

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. [158]

Hanushek, E. A. (1992), "The trade-off between child quantity and quality." *Journal of Political Economy*, 100 (1), 84–117. [158]

Heckman, J., S. Urzua, and E. Vytlacil (2006), "Understanding instrumental variables in models with essential heterogeneity." *Review of Economics and Statistics*, 88 (3), 389–432. [183]

Horowitz, J. L. (2009), "Applied nonparametric instrumental variables estimation." Working paper. [174]

Moav, O. (2005), "Cheap children and the persistence of poverty." *The Economic Journal*, 115, 88–110. [158]

Mogstad, M. and M. Wiswall (2012), "Instrumental variables estimation with partially missing instruments." *Economics Letters*, 114 (2), 186–189. [174, 175, 184]

Newey, W. K. (1990), "Efficient instrumental variables estimation of nonlinear models." *Econometrica*, 58 (4), 809–837. [176, 179]

Newey, W. K. (1993), "Efficient estimation of models with conditional moment restrictions." In *Handbook of Statistics*, Vol. 11 (G. S. Maddala, C. R. Rao, and H. D. Vinod, eds.), Elsevier, Amsterdam. [176, 179]

Newey, W. K. and J. L. Powell (2003), "Instrumental variable estimation of nonparametric models." *Econometrica*, 71 (5), 1565–1578. [174]

OECD (2002), *Taxing Wages: 2001 Edition*. Organization for Economic Cooperation and Development, Paris. [157]

Qian, N. (2008), "Quantity–quality and the one child policy: The positive effect of family size on school enrollment in China." Working paper. [172]

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. [158, 160]

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, 1149–1174. [170]

Ruhm, C. J. (2008), "Maternal employment and adolescent development." *Labour Economics*, 15 (5), 958–983. [162]

Staiger, D. and J. H. Stock (1997), "Instrumental variable regression with weak instruments." *Econometrica*, 65 (3), 557–586. [179]

Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA. [177]

Co-editor Petra Todd handled this manuscript.

Submitted November, 2012. Final version accepted March, 2015.