Testing the Quantity-Quality Model of Fertility: Linearity, Marginal Effects, and Total Effects

Magne Mogstad
Statistics Norway
Research Department
magne.mogstad@ssb.no

Matthew Wiswall
New York University
Department of Economics
mwiswall@nyu.edu[1]

Version: 05/30/2010

Abstract

We re-examine the recent empirical evidence suggesting no tradeoff between child quantity and quality. Motivated by the theoretical ambiguity about the magnitude and sign of the marginal effects on child quality of additional siblings, we depart from previous empirical studies in allowing an unrestricted relationship between family size and child outcome. We find that the conclusion of no family size effect is an artifact of a linear specification in family size, masking substantial marginal family size effects. This is true when we perform OLS estimation with controls for confounding characteristics like birth order, or instrument family size with twin births.

Keywords: quantity-quality model of fertility, family size, birth order, non-linearity, instrumental variables

JEL-codes: C31, C14, C13

[1]Thanks to R. Aaberge, J. Angrist, C. Brinch, D. Del Boca, P. Devereux, T. Havnes, T. Hægeland, K. Salvanes, T. Skjerpen, S. Wang and a number of seminar and conference participants for useful comments and suggestions on a previous version of this paper. The Norwegian Research Council has provided financial support for this project.

1 Introduction

Motivated by the seminal quantity-quality (QQ) model of fertility by Becker and Lewis (1973), a large and growing body of empirical research has examined the relationship between family size and children's outcomes. Much of the early literature that tested the QQ model found that larger families reduced 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 model and argued that family size has no causal effect on children's outcomes. The pioneering study by Black, Devereux, and Salvanes (2005, hereafter BDS) concludes 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" (p. 697). They therefore argue that we need to "revisit models of fertility and reconsider what should be included in the 'production function' of children" (p. 698). Using data from the US, Caceres-Delpiano (2006) comes to a similar conclusion. Other recent studies reporting no effect of family size include Angrist et al. (2006) using data from Israel and Aaslund and Grønqvist (2007) using data from Sweden.

Although these studies represent a significant step forward, a concern is that their evidence for no causal family size effect 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 magnitude and sign of the marginal effects on child quality of additional siblings (Rosenzweig and Wolpin, 1980), we depart from previous empirical studies in allowing an unrestricted relationship between family size and child outcomes. Specifically, we estimate non-parametric models in family size using both empirical strategies used to test the QQ model: OLS estimation with controls for confounding characteristics like birth order, and IV estimation using instruments for family size. To rule out differences simply because of different data sources, we use the same data as BDS, administrative registers for the entire population of Norway. In addition, we follow BDS in using twin births as the instrument for family size and the same control variables.

We provide an analysis of the major empirical issues in testing the QQ model, focusing on the interpretation of linear and non-linear models of family size and child outcome, and the construction of non-parametric IV estimators using twin birth instruments. The insights from our analysis may be summarized by the following six points:

1) Although the QQ model implies that family size and child quality is jointly determined, we show that there is nothing in the theory that suggests 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 non-linearities in child quality from changes in family size. This is because parental preferences mediate exogenous changes in family size such that an increase in the number of children can have large or small, and negative or positive, effects on existing children. Hence, the relationship between family size and child outcome may not only be non-linear but even non-monotonic. Imposing a linear specification in family size when testing the QQ model is therefore worrisome. In particular, the linear OLS estimator can indicate a zero effect when in fact all marginal effects are non-zero. On top of this, the linear IV estimator can, even with homogeneous marginal effects and strong instruments, misrepresent the sign of the average causal effect of family size.

2) As in previous research, our OLS estimate of the linear model indicates an almost zero effect of family size on children's education, after controlling for birth order and other demographic variables. However, the OLS estimates of a non-parametric model in family size reveal a non-monotonic inverse U-shaped pattern, with statistically significant and sizable marginal family size effects. The reason for the almost zero effect of family size in the linear OLS estimation is that the negative and positive marginal effects at different parities cancel each other out.

BDS also perform OLS estimation of a non-parametric model in family size controlling for birth order and the same demographic variables. However, their non-parametric model in family size is constructed in such a way that their parameter estimates provide the *total effect* of family size relative to 1 child families. From these results, BDS 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" (p. 680).

However, it is the marginal effects – not the total effects – that provide the relevant comparison to the linear estimate and the appropriate test of the QQ model. In particular, the linear model provides a linearly restricted estimate of the marginal effect. Given the non-monotonic relationship between family size and children's education, the total effect specification of BDS misrepresents the sign, magnitude, and significance level of many of the marginal effects estimates. Whereas some of the total effect estimates are quite small, most of the marginal family size effects are of similar magnitude or bigger than the linear family size estimate obtained by BDS when omitting birth order controls, which they emphasize as large. In terms of the QQ model, the estimated marginal effects can be interpreted as suggesting a tradeoff between quantity and quality in large families, and (strong) complementarities between quantity and quality in small families.

3) An important and much cited finding of BDS is the large birth order effects on children's education. Like their non-parametric family size specification, BDS report birth order estimates as total effects, rather than marginal effects. Because the marginal birth order effects are monotonically negative, their total effect specification exaggerates the marginal effect of being 1 birth parity later in the birth order compared to increasing family size by 1 child. In fact, 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 actually exceeds every marginal birth order effect, except for the effect of being 2nd instead of 1st born. Moreover, the effect of having 4 instead of 3 children in the family is only slightly lower than the effect of being born 4th rather than 3rd, while the effect of having 5 instead of 4 children in the family is actually more than twice as large as the effect of being 5th instead of 4th born.

We further demonstrate that the conclusion in BDS that ”birth order effects appear to drive the observed negative relationship between family size and child education” (p. 671) does not hold once we relax the linear in family size restriction. 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) Like previous research, our IV results from the linear model using twin births as instruments show a small and fairly imprecisely estimated effect of family size. BDS argue that the lower bound of the 95 percent confidence interval for their linear estimates rule out large negative effect of family size. However, when we relax the linearity restriction and estimate a non-parametric model in family size using the twin birth instruments, we can no longer rule out large effects of family size at conventional significance levels. In fact, our IV estimates are not significantly different from the sizable OLS estimates of marginal family size effects or marginal birth order effects. One may want to be cautious in accepting the conclusion in BDS of no effect of family size, as it is *not* robust to relaxing the linear specification in the IV estimation

5) Instead of settling for the inconclusive evidence from the imprecise IV estimates, we try to gain precision by exploiting the binary nature of the family size dummy variables in the non-parametric model, as well as the unequivocally effect a twin birth has on adding another child to the family. Although previous studies of family size and children's outcome have not imposed such structure in the IV estimation, our approach produces consistent estimates under the the standard IV assumptions, and is potentially more efficient.[2] When applying this alternative IV strategy to the non-parametric model in family size, we find large and statistically significant family size effects. For first born children, the causal relationship between their education and number of siblings is clearly non-monotonic. While 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 that BDS emphasize as large. In comparison, this alternative IV strategy produces a linear IV estimate close to zero. By comparing the results from the linear and non-parametric models in family size using the alternative IV strategy, we see the important role of the linearity restriction in masking the family size effects.

6) 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. The linear OLS estimator reflects all marginal family size effects and weights them according to the sample distribution of family size, assigning the most weight to marginal effects close to the sample median family size. In comparison, the linear IV estimator only captures the marginal effects at the part of the support shifted by the specific instrument chosen. For example, using twins at second birth as the instrument weights the marginal effect of moving from 2 to 3 children most heavily, assigning far less weight to marginal effects at higher parities. The reasons for the almost zero effect of family size in both OLS and IV estimation of the linear model are that (i) negative and positive marginal effects at different parities offset each other, and that (ii) the relatively small marginal effects are weighted heavily. Importantly, OLS and IV estimates of the linear family size model assign substantially different weights to the underlying marginal family size effects. Drawing

[2] Wooldridge (2002) and Carneiro et al. (2003) provide examples of IV estimation using instrument constructed like we have here. In both applications, they find substantial improvement in the precision of the IV estimates when using constructed instruments over the IV estimates using the instruments directly.

conclusions about the endogeneity of family size by comparing linear OLS and IV estimates, as in previous studies, may therefore be unwarranted.

The paper unfolds as follows. Section 2 outlines the QQ model, focusing on the distinction between total vs. marginal effects, and 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 non-parametric models in family size. Section 5 describes the IV methods and presents IV results from the linear and non-parametric 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 quality are jointly determined. For a given number of children N, the optimal expenditure per child on child specific goods can be defined as

$$e^*(N) = \arg\max_e U(N, Q, C)$$

$$\text{s.t. } I = Ne + C \text{ and } Q = q(e).$$

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. Following their study, 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 exogenously increase in family size from $N-1$ to N children is given by

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

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 non-parametric specification of family size in the OLS estimation in BDS, the *total* effect of family size relative to 1 child families is given by

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

While $\Omega(2,1) = \Delta(2,1)$, 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 which 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 CES structure for preferences. We emphasize that this parametrization is merely to illustrate the *possibility* that the QQ model allows for non-linear and positive marginal effects of additional children. Importantly, we do not impose this parametrization in the empirical estimation.

Assume preferences and technology take the following form:

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

where

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

and child quality production technology

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

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 $U_1(N, C)$ 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 $U_1(N, C)$ 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 0.1 to a high value 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 non-linearities in the effects on child quality from changes in family size. This is because parental preferences mediate exogenous changes in family size such that 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 tradeoff. 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 non-zero. 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 non-zero. This is not merely a theoretical peculiarity, but is in fact 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 tradeoff between quantity and quality in large families, and strong complementarities between quantity and quality in small families.

It should finally be noted that non-linearities 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), 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 BDS, our data is based on administrative registers from Statistics Norway covering the entire resident population of Norway who were between 16 and 74 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 contains 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. We refer to BDS for a more detailed description of the data as well as of relevant institutional details for Norway.

To the best of our knowledge, we use the same sample selection as BDS. We restrict the sample to children who were aged at least 25 in 2000 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 BDS is that we exclude a small number of families with more than 6 children.[3] The final sample includes 1,429,126 children from 625,068 families (98 % 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 BDS. 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 age 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. As expected, fathers are, on

[3] Our main reason for excluding large families is that the estimates of the marginal birth order effects and the marginal family size effects are unstable and imprecise for families with more than 6 children. We discuss these findings below. In BDS, the relative lack of precision in these marginal effects is hidden by their total effect specification.

average, slightly older and more educated than mothers.

As in BDS, our 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 consists of children born to families with 4, 5, or 6 children.

4 Empirical Models and OLS Estimates

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

4.1 Linear vs. Non-Parametric 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 other covariates X_i. For child i, we denote her number of siblings using $s_i \in \{0, 1, \ldots, \bar{s}\}$. 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 family size model is specified as

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

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 specify a non-parametric model in family size model by including dummy variables for each number of siblings:

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

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.

All of the IV specifications in previous studies impose this linearity restriction. However, BDS relax the linearity restriction in their OLS estimation. Specifically, they estimate the following model

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

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. 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 $\gamma_1 = \psi_1$.

Although both dummy variable specifications (4) and (5) are non-parametric in family size as they fully saturate the support of the family size variable, the difference in construction is important for interpretation and has apparently created considerable confusion in the empirical literature. As discussed below, BDS interpret the usually small, and sometimes insignificant, total effects estimates from (5) as supporting the conclusion of little, if any, effect of family size drawn from the small linear estimate from (3). However, the linear estimate is directly comparable to the marginal effects estimates from (4), and not the total effects estimates from (5). Comparing marginal effects estimated from a linear model to total effects is an "apples to oranges" comparison. Even if the linear model restriction is correctly imposed, and the marginal effects are constant, the total effects are not directly comparable to the linear estimate. [4] And if the relationship between family size and child quality is non-monotonic, the total effect estimates may misrepresent the sign, magnitude, and significance level of the marginal effects. Below, we demonstrate that this is exactly what is happening.

4.2 OLS Estimates with Controls for Birth Order

Table 3 reports the OLS estimates from the linear and non-parametric models in family size. This table replicates Table IV in BDS (p. 679), except that we exclude children from families with more than 5 siblings (more than 6 children in total). We also depart from their estimation method in using specification (4), where the dummy variables for family size are constructed as marginal effects, to focus attention on the constant marginal effects restriction imposed by linearity.

[4]The total effects from a correctly imposed marginal effect model are given by $\psi_s = \beta * s$ for all s, where β is the constant marginal effect. The total effects are therefore not constant across margins even if the marginal effect is.

To provide a direct comparison to the estimation results presented in BDS, the last column of Table 3 reports results using the total effects specification (5).

The first column of Table 3 reports the OLS estimate from model (3), showing that the linear family size effect is -0.20. This suggests that each additional sibling reduces average education of all 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 non-monotonic 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 remaining columns of Table 3 add control variables (the same as BDS) 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 non-parametric estimates of the effect of family size on children's education, suggesting that OLS estimation could be biased because child quality and quality is jointly determined.

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 5 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 0 otherwise), the second dummy variable is equal to 1 if the child was born third or higher in the birth order (and 0 otherwise), and so on. In this specification of model (4), the support 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 being born 1 birth parity later in the birth order (e.g from 1st to 2nd born).

Like BDS, 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-shape pattern in family size. In particular, the inclusion of birth order controls boosts the only child penalty, as the marginal effect of moving from a 0 to 1 siblings 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, which is almost of the same magnitude as the linear estimate without controls for birth order that BDS emphasize as large.[5]

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 8 reports the weights for the linear IV 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 non-monotonic distribution of marginal family size effects and these particular OLS weights yield the near zero linear OLS estimate.

Finally, the last column of Table 3 reports results from model (5), replacing the marginal effects specification with the total effects specification for birth order and family size. In doing so, we replicate the OLS results of BDS (p. 679, Table IV, Column 6) when they use their non-parametric model in family size with controls for birth order and other confounding characteristics. Like BDS, 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, BDS 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. While the linear estimate suggests no effect of family size on child outcome, the marginal family size effects estimates can be interpreted as suggesting a tradeoff between quantity and quality in large families, and complementarities between quantity and quality

[5]We have also estimated the model in Column 6 of Table 3 for the sample of children from families with 1-10 children, including a full set of family size and birth order dummy variables. For the families with 7-10 children, the estimated marginal family size effects at these parities are negative but imprecise. At these higher parities, the estimated marginal birth order effects are more precise but unstable, alternating between positive and negative marginal effects. Estimated marginal family size effects (standard errors in parentheses): 6th sibling, -0.041 (0.032); 7th sibling -0.054 (0.051); 8th sibling -0.023 (0.077); 9th sibling -0.084 (0.11). Estimated marginal birth order effects (standard errors in parentheses): born 7th, -0.077 (0.040); born 8th, 0.18 (0.064); born 9th, -0.29 (0.10); born 10th, 0.097 (0.167).

in small families

4.3 Relative Importance of Birth Order vs. Family Size

An important and much cited finding of BDS is the large birth order effects on children's education. Like their non-parametric family size specification, BDS report birth order estimates as total effects, rather than marginal effects. Because the marginal birth order effects are monotonically negative, their total effect specification overstates the effect of being 1 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 2nd instead of 1st 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 4th rather than 3rd (-0.100), while the effect of having 5 instead of 4 children in the family (-0.089) is actually more than twice as big as the effect of being 5th instead of 4th born (-0.040).

Table 3 also shows that the conclusion in BDS that birth order effects appear to drive the observed negative relationship between family size and child education does not hold once relaxing the linear specification in family size. 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.

4.4 Results by Birth Order

Table 4 reports results from the linear family size model (3) and the non-parametric 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 4 estimates the linear family size model, whereas the bottom panel estimates the non-parametric model in family size. Contrasting the estimates from the two types of models for each birth order, 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 4. The figures present educational attainment relative to only children, whose average educational attainment is normalized to 0.

For each of the birth order sub-samples, the coefficients on the main diagonal of Table 4 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.[6] 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 non-parametric specifications is particularly stark for the sub-sample of first born children. While the linear OLS specification predicts that additional children have a zero impact on first born children, the non-parametric specification predicts significant negative effects of having more than 1 sibling. Adding a 3rd sibling is estimated to reduce educational attainment of first born children by 0.086 years, adding a 4th sibling reduces education an additional 0.16 years, and a 5th sibling child an additional 0.11 years.

5 IV Estimates

This section focuses on the second of the two empirical strategies employed by BDS and others to estimate the effects of family size on child outcome: IV estimation using twin births as instruments for family size.

5.1 Linear IV Models

Following Rosenzweig and Wolpin (1980), twin births have been commonly used as an instrument for family size. The rationale for using twins as instruments is that for some families, twin births increase the number of siblings beyond the desired family size.[7] We follow BDS in restricting the sample to children born before the twin birth, to avoid including the endogenously selected outcomes of

[6] 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 makes them to choose to have an additional child.

[7] See for example BDS for results supporting the internal validity of twin birth as an instrument for family size. Angrist et al. (2006) also use sex composition of the children as an instrument for family size. However, recent evidence suggests that sex composition may have a direct effect on children's outcomes, implying that it may not be a valid instrument for family size (see e.g. Dahl and Moretti, 2008).

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 ($c_i \geq 2$):

$$y_i = \beta s_i + X_i' \delta + \epsilon_i \quad \text{(Second Stage)}$$

$$s_i = \lambda twin_{2i} + X_i' \rho + \eta_i \quad \text{(First Stage I)}$$

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 ($c_i \geq 3$):

$$y_i = \beta s_i + X_i' \delta + \epsilon_i \quad \text{(Second Stage)}$$

$$s_i = \lambda twin_{3i} + X_i' \rho + \eta_i \quad \text{(First Stage I)}$$

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

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

$$y_i = \beta s_i + X_i' \delta + \epsilon_i \quad \text{(Second Stage)}$$

$$s_i = \lambda twin_{4i} + X_i' \rho + \eta_i \quad \text{(First Stage I)}$$

where $twin_{4i}$ is a dummy variable for whether the fourth birth was a twin birth (implying that third and fourth 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:

$$E[\epsilon_i|X_i, c_i \geq \tilde{c}] = E[\epsilon_i|c_i \geq \tilde{c}] = 0 \tag{6}$$

$$E[\epsilon_i|X_i, c_i \geq c, twin_{ci}] = E[\epsilon_i|X_i, c_i \geq c], \quad for\ all\ c \geq \tilde{c} \tag{7}$$

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

Since we follow BDS 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 non-uniform 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.

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 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}). However, this is not the case. Panel B, C and D in Table 8 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.

5.2 Non-Parametric IV Estimation

The sensitivity of the OLS results to the choice between a linear and a non-parametric 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. Following the OLS estimation above, we specify non-parametric 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 ($c_i \geq 2$)

$$y_i = \gamma_2 d_{2i} + \gamma_3 d_{3i} + \gamma_4 d_{4i} + \gamma_5 d_{5i} + X_i'\delta + \epsilon_i \quad \text{(Second Stage)}$$

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

$$y_i = \gamma_3 d_{3i} + \gamma_4 d_{4i} + \gamma_5 d_{5i} + X_i'\delta + \epsilon_i \quad \text{(Second Stage)}$$

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

$$y_i = \gamma_4 d_{4i} + \gamma_5 d_{5i} + X_i'\delta + \epsilon_i \quad \text{(Second Stage)}$$

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

First, identification of the non-parametric 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 $\gamma_2, \ldots, \gamma_5$ using the full set of twin birth

[8]It should be noted that the standard issues with non-parametric IV are avoided in our family size application. A considerable literature discusses how to non-parametrically 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 non-parametric $f(\cdot)$ function without any loss of generality.

instruments, $twin_{2i}, \ldots, twin_{5i}$. However, because of the nature of the twin birth instrument, the full set of instruments are 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, since 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 adapting the method proposed by Angrist et al (2006) and further discussed Mogstad and Wiswall (2010). 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, BDS 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 negative effects of family size at conventional significance levels. Instead of settling for the 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 unequivocally 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 assumption as the linear IV estimation of Models 1-3.

5.3 Using the Full Set of Instruments

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

$$d_{si} = \lambda_{s2} twin_{2i} + \lambda_{s3} twin_{3i} + \lambda_{s4} twin_{4i} + \lambda_{s5} twin_{5i} + X_i' \rho_s + \eta_{si}. \quad (8)$$

However, (8) is not feasible because the twin birth instruments $twin_{3i}$, $twin_{4i}$, $twin_{5i}$ 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

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

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

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

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

In order to use the full set of instruments necessary to identify the non-parametric model in family size, we instead follow the strategy proposed by Angrist et al (2006) and further discussed in Mogstad and Wiswall (2010). Specifically, we construct instruments defined for the full sample as

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

where $\hat{E}[twin_{ci} | X_i, c_i \geq c]$ is a initial stage non-parametric estimator for the conditional mean of the instrument (probability of twin birth) in the sub-sample where it is non-missing. 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 the $twin^*_{ci}$ 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 non-missing, whereas $twin_{3i}$, $twin_{4i}$ and $twin_{5i}$ are missing

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

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

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

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

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

In general, $E[twin_{ci}|X_i, c_i \geq c]$ is an unknown non-linear function that needs to be estimated. We estimate the conditional mean using a polynomial function of X_i. The non-parametric 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, including nearly 200 covariates, provides a reasonable approximation of the conditional mean function. Given that the main predictor of twinning probabilities is the mother's age at birth (BDS), this approximation is particularly well suited to our application since 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 moves the estimate of $E[twin_{ci}|X_i, c_i \geq c]$. We also provide a simulation exercise in Appendix B, which shows that instruments constructing 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.

5.4 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) implies 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 non-parametric estimators for optimal instruments. As an alternative, we impose a particular functional form when constructing our "efficient instruments," to address the concern that a higher level of small sample bias may be introduced if we use non-parametric methods and implicitly impose more over-identifying 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 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 et al. (2003) 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), 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" (p. 625).

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, given by

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

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 one. 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 non-linear 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 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, given by

$$\hat{p}_{3i} = f_3(X_i, twin^*_{3i}, \hat{\theta}_3)$$

$$\hat{p}_{4i} = f_4(X_i, twin^*_{4i}, \hat{\theta}_4)$$

$$\hat{p}_{5i} = f_5(X_i, twin_{2i}, twin^*_{5i}, \hat{\theta}_5)$$

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.

Next, we replace the instruments $twin_{2i}$, $twin^*_{3i}$, $twin^*_{4i}$, $twin^*_{5i}$ with the efficient instruments $\hat{p}_{2i}, \hat{p}_{3i}, \hat{p}_{4i}, \hat{p}_{5i}$ in the first stage specifications of Model 4, before applying standard 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 ($c_i \geq 2$)

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

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

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

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

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

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

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

where

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

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 sub-sample 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 non-linear model to predict the endogenous family size variables. This has the advantages of appropriately restricted the range to the unit interval, in addition to taking into account that twin births unequivocally increase family size by one 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 non-parametric 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 mis-specification 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 et al (2006) who interact the twin birth instruments with covariates in their study of family size effects, the efficient instruments generate an over-identified IV estimator, which may exacerbate the small sample bias in 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 showing that the constructed instruments are very strongly correlated with family size, the literature on small sample bias of the IV estimator suggests that this number of over-identifying restrictions should be of little concern (e.g. Staiger and Stock, 1997). Our simulation exercise reported in Appendix B supports this conjecture. The simulation

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

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 over-identifying restrictions is suggestive that this procedure does not increase the small sample bias of the IV estimator.

5.5 IV Estimates

Tables 5-7 present IV results for the linear models in family size in Panel I (Models 1-3) and the non-parametric 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").

Evidence for efficiency gains in using the efficient IV is found by examining the first stage R-square values. In all cases, the efficient IV have higher R-square in the first stage then the standard 2SLS. This demonstrates that the efficient instruments are more strongly correlated with the endogenous family size variables. The gains in R-square 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-square 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-Square 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 estimates using the efficient IV to have smaller standard errors than those based on the standard 2SLS.

Table 5 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. Like in BDS, 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.[10] On this basis, BDS conclude that they can rule out

[10]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 re-compute the instru-

large negative effects of family size. However, large positive effects of family size cannot be ruled out as the the upper bound of the 95 percent confidence interval is as large as 0.15. In terms of the QQ model, this results suggest that there could be complementarities between child quantity and quality of changes in family size induced by the second birth being a twin birth.

Relaxing the linearity restriction in family size, the results reported in the first column of Panel II in Table 5 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. In light of the theoretical ambiguity in the functional form of the relationship between family size and child outcome and the inverse U-shaped pattern identified in the OLS estimates with birth order control, we may therefore want to be cautious in accepting the conclusion in BDS of no large negative effects of family size as it is not robust to relaxing the linear specification in the IV estimation.

Turning to the efficient IV results of the non-parametric model in family size, reported in the second column of Panel II in Table 5, the main finding is that there are significant and large marginal family size effects on children's education. Furthermore, the results indicate a non-monotonic causal relationship between family size and children's education. For 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 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 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 birth order effects that BDS emphasize as large.[11] In terms of the QQ model, the efficient IV estimates of the non-parametric model in family size indicate a tradeoff between quantity and quality in large families, and complementarities between quantity and quality in

ment defined for the full sample, the constructed efficient instruments, and the 2SLS estimators which use these instruments.

[11] See also Conley and Glauber (2006) and Price (2008).

small families.

Comparing the efficient IV results from the linear and non-parametric models in family size reported in the first column of Table 5, 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 non-parametric 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.

The fact that we obtain much larger point estimates of marginal effects in the non-parametric model using the efficient instruments compared to those using the twin birth instruments directly, speaks to the recent discussion concerning the interpretation of IV estimation under heterogeneous treatment effects and variable treatment intensity (see e.g. Imbens and Angrist, 1994; Angrist and Imbens, 1995; Angrist et al., 2000; Heckman et al., 2006; Moffitt, 2008). As emphasized in this literature, different instruments will in general identify different local average treatment effect (LATE), even if they are valid under the same assumption. Our point estimates using the efficient instruments are estimating a LATE for a subgroup of first born children that has substantially positive marginal family size effects at low parities and large negative marginal family size effects at higher parities. Other subgroups comprising other LATEs may have marginal family size effects of different magnitude and sign at these parities. Pinning down the heterogeneity in family size effects is an important area for future research.

For later born children, the IV estimates in Table 6 and Table 7 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 non-parametric 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 non-parametric models in family size. For example, the linear estimate of the effect of family size on second born children suggest than 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).

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 non-linear 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 is *not* robust to relaxing 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 non-parametric in family size, we find a non-monotonic relationship with statistically significant and sizable marginal effects. In terms of the QQ model, this inverse U-shaped pattern suggest a tradeoff 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, including publicly provided or subsidized child care as well as welfare and tax policies, such as maternity leave laws, family allowances, single parent benefits, and family tax credits. In fact, families with children receive special treatment under the tax and transfer provisions in twenty-eight of the thirty OECD countries (OECD, 2002). Many of these policies are designed such that they reduce the cost of having one child more than the cost of having additional children, in effect promoting smaller families.[12] If a pol-

[12]For example, welfare benefits are, in many cases, reduced or even cut off after reaching a certain number of children.In the United States, a recipient of the Earned Income Tax Credit program could in 2007 receive a maximum credit of USD $2,900$ if he or she had one qualifying child; for two or more qualifying children, the maximum credit was only USD $4,700$. In addition, a number of US states have implemented family cap policies, providing little or no increase in cash benefits when a child is born to a mother who is on welfare. Another example is from Norway, offering generous benefits to single parents. However, the benefit amount received is independent of the number of dependent children. Some developing countries have implemented far more radical policies to promote smaller families, such as China's One Child Policy and an aggressive public promotion of family planning in Mexico and Indonesia. See Feyrer et al. (2008) and Del

icy 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 that there is no need to be concerned with the effects on the human capital development of existing children when designing policies promoting larger families. Our findings runs counter to this conclusion. In fact, our OLS and IV estimates indicating an inverse U-shaped pattern suggest that an efficient policy might be to target incentives for higher fertility to small families, and discourage larger families from having addition children.

Boca and Wetzels (2008) for recent reviews of the literature for developed countries.

References

AASLUND, O., AND H. GRØNQUIST (2007): "Family Size and Child Outcomes: Is There Really No Trade-Off," Working Paper 15, IFAU.

ANGRIST, J. D., K. GRADDY, AND G. W. IMBENS (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67(3), 499–527.

ANGRIST, J. D., AND G. IMBENS (1995): "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of American Statistical Association*, 90(430), 431–442.

ANGRIST, J. D., AND A. B. KRUEGER (1999): "Empirical strategies in labor economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3, pp. 1277–1366. Elsevier.

ANGRIST, J. D., V. LAVY, AND A. SCHLOSSER (2006): "Multiple Experiments for the Causal Link between the Quantity and Quality of Children," *MIT Working Paper, 06-26*.

BANDURA, A. (1977): *Social Learning Theory*. Prentice Hall, Englewood Cliffs, NJ.

BECKER, G. S. (1998): *A Treatise on the Family. Enlarged Version*. Harvard University Press, Cambridge, MA.

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.

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.

CACERES-DELPIANO, J. (2006): "The Impacts of Family Size On Investment in Child Quality," *Journal of Human Resources*, 41(4), 738–754.

CARNEIRO, P., J. HECKMAN, AND E. VYTLACIL (2003): "Estimating Marginal and Average Returns to Education," *Working paper*.

CONLEY, D., AND R. GLAUBER (2006): "Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility," *Journal of Human Resources*, 41(4), 722–737.

DAHL, G., AND E. MORETTI (2008): "The Demand for Sons," *Review of Economic Studies*, 75(4), 1085–1120.

DEL BOCA, D., AND C. WETZELS (2008): *Social Policies, Labour Markets and Motherhood: A Comparative Analysis of European Countries*. Cambridge University Press.

DONALD, S. G., AND W. K. NEWEY (2001): "Choosing the Number of Instruments," *Econometrica*, 69(5), 1161–91.

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.

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

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.

HOROWITZ, J. L. (2009): "Applied Nonparametric Instrumental Variables Estimation," *Working paper*.

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–75.

MOFFITT, R. (2008): "Estimating Marginal Treatment Effects in Heterogeneous Populations," *Working paper*.

MOGSTAD, M., AND M. WISWALL (2010): "Instrumental Variables Estimation with Partially Missing Instruments," working paper.

NEWEY, W. K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica*, 58(4), 809–37.

——— (1993): "Efficient Estimation of Models with Conditional Moment Restrictions," in *Handbook of Statistics, Vol. 11*, ed. by G. Maddala, C. Rao, and H. Vinod. Elsevier.

NEWEY, W. K., AND J. L. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71(5), 1565–78.

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

PRICE, J. (2008): "Parent-Child Quality Time: Does Birth Order Matter?," *Journal of Human Resources*, 43(1), 240–265.

QIAN, N. (2008): "Quantity-Quality and the One Child Policy: The Positive Effect of Family Size on School Enrollment in China," *working paper*.

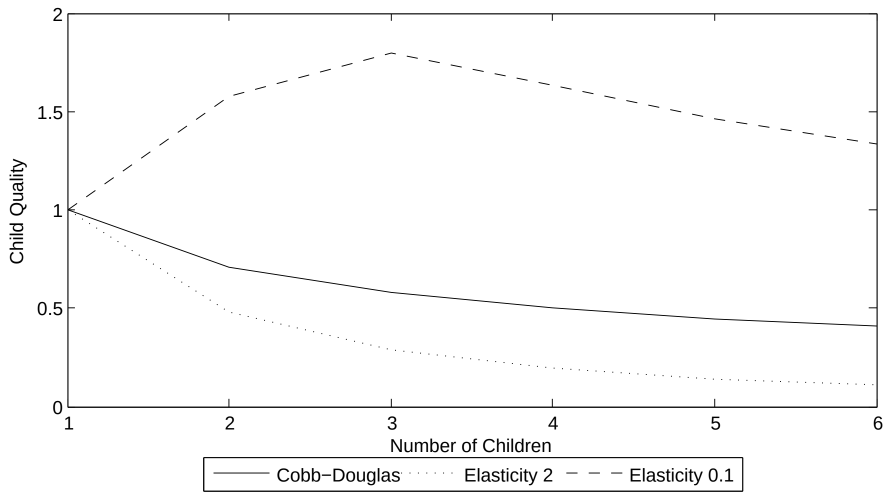
ROSENZWEIG, M. R., AND K. I. WOLPIN (1980): "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48(1), 227–240.

RUHM, C. J. (2008): "Maternal Employment and Adolescent Development," *Labour Economics*, 15(5), 958–983.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variable Regression with Weak Instruments," *Econometrica*, 65(3), 557–86.

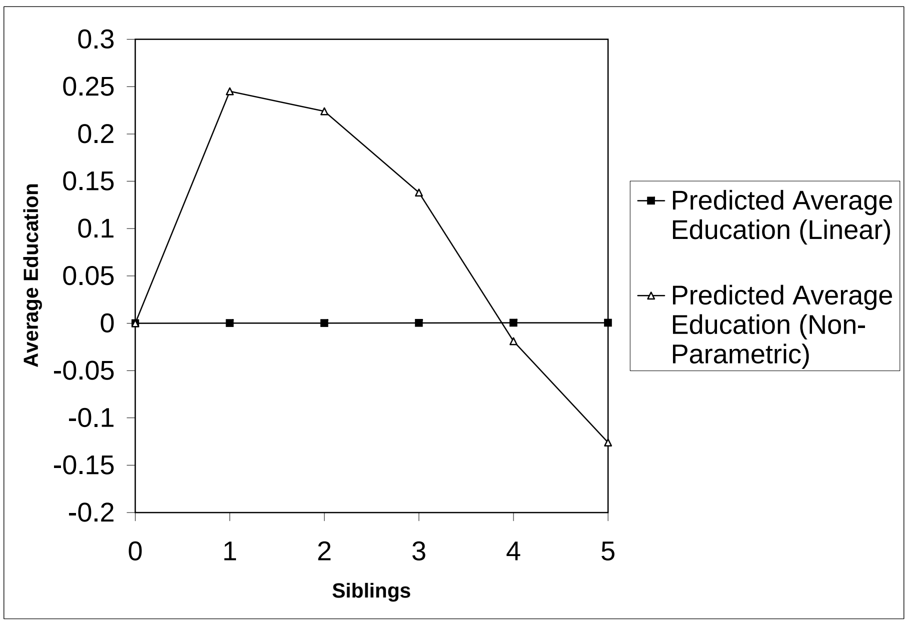
WOOLDRIDGE, J. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Figure 1: Family Size Effects in the Quantity Quality Model with Different Substitution Elasticities between the Quantity and Quality of Children



Notes: The other model parameters are set at $\alpha = 0.5$, $\gamma = 0.5$, $\theta = 0.5$.

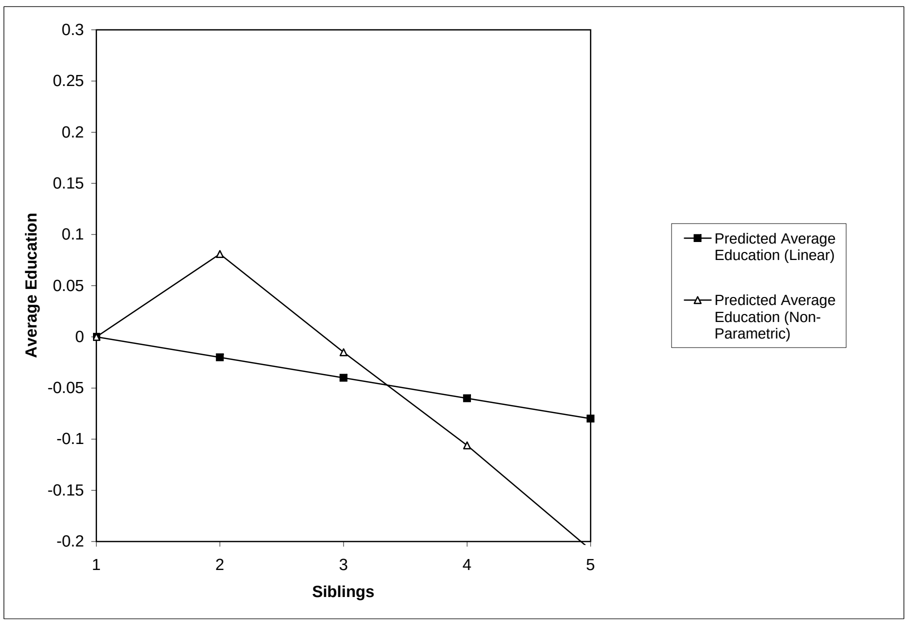
Figure 2: Average Educational Attainment for First Born Children by Number of Siblings (Relative to Only Children)



Notes: This figure graphs the linear and non-parametric in family size predictions from OLS regressions (from Table 4). These values are graphed relative to only children (0 siblings), i.e. 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 non-parametric model allows non-constant slopes. The marginal effect estimate from the linear model is close to zero (represented by a flat line), while the non-parametric estimate of the marginal effects indicates that they are non-monotonic. The linear prediction of total effect is $\hat{y} = \hat{\beta} * s$ for $s = 0, 1, \ldots, 5$, where s is number of siblings and $\hat{\beta}$ is the OLS estimate from the first panel of Table 4. Non-parametric 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 4.

Source: Administrative registers from Statistics Norway.

Figure 3: Average Educational Attainment for Second Born Children by Number of Siblings (Relative to Children with 1 Sibling)



Notes: This figure graphs the linear and non-parametric in family size predictions from OLS regressions (from Table 4). These values are graphed relative to second born children in 2 child families (1 sibling), i.e. 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 non-parametric model allows non-constant slopes. The linear prediction is $\hat{y} = \hat{\beta} * (s - 1)$ for $s = 1, 2, \ldots, 5$, where s is number of siblings and $\hat{\beta}$ is the OLS estimate from the first panel of Table 4. Non-parametric 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 4.

Source: Administrative registers from Statistics Norway.

7 Tables

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

Notes: 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.

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

Notes: 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.

Table 3: OLS Estimates of Marginal Effects from Linear and Non-Parametric Models

	Marginal Effects						
	(1)	(2)	(3)	(4)	(5)	(6)	Total Effects from (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	

Notes: 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 (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.

Table 4: OLS Estimates of Marginal Effects by Birth Order for Linear and Non-Parametric Models

	Birth Order				
	1	2	3	4	5
Panel I: Linear Estimates of Marginal Effects					
Numb. of Child.	0.0001	-0.020	-0.037	-0.037	-0.006
	(0.003)	(0.004)	(0.007)	(0.013)	(0.033)
Panel II: Non-Parametric 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)

Notes: 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.

Table 5: IV Estimates of Marginal Effects in Linear and Non-Parametric Models for First Born Children in Families with 2 or more Children

Panel I: Linear Estimates of Marginal Effects		
Instrument:	Standard 2SLS	Efficient IV
Numb. of Children	0.053	-0.0036
	(0.0495)	(0.0460)

Panel II: Non-Parametric Estimates of Marginal Effects		
Instrument(s):	Standard 2SLS	Efficient IV
Siblings	0.079	0.153
≥ 2	(0.067)	(0.063)
Siblings	-0.044	-0.474
≥ 3	(0.073)	(0.079)
Siblings	0.023	-0.800
≥ 4	(0.100)	(0.129)
Siblings	-0.051	-0.787
≥ 5	(0.181)	(0.247)

Notes: Each column is 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 re-compute the instruments defined for the full sample, the constructed efficient instruments, and the 2SLS estimators which use these instruments.

Table 6: IV Estimates of Marginal Effects in Linear and Non-Parametric Models for Second Born Children in Families with 3 or more Children

Panel I: Linear Estimates of Marginal Effects		
Instrument:	Standard 2SLS	Efficient IV
Numb. of Children	-0.051	-0.171
	(0.053)	(0.051)

Panel II: Non-Parametric Estimates of Marginal Effects		
Instrument(s):	Standard 2SLS	Efficient IV
Siblings	-0.058	-0.090
≥ 3	(0.068)	(0.068)
Siblings	-0.054	-0.586
≥ 4	(0.093)	(0.110)
Siblings	0.138	-0.504
≥ 5	(0.224)	(0.205)

Notes: Each column is 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 re-compute the instruments defined for the full sample, the constructed efficient instruments, and the 2SLS estimators which use these instruments.

Table 7: IV Estimates of Marginal Effects in Linear and Non-Parametric Models for Third Born Children in Families with 4 or more Children

Panel I: Linear Estimates of Marginal Effects		
Instrument:	Standard 2SLS	Efficient IV
Numb. of Children	-0.107	-0.191
	(0.075)	(0.074)

Panel II: Non-Parametric Estimates of Marginal Effects		
Instrument(s):	Standard 2SLS	Efficient IV
Siblings	-0.096	-0.145
≥ 4	(0.090)	(0.093)
Siblings	-0.178	-0.520
≥ 5	(0.163)	(0.176)

Notes: Each column is 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 re-compute the instruments defined for the full sample, the constructed efficient instruments, and the 2SLS estimators which use these instruments.

Table 8: 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 order					
OLS Weight	0.110	0.336	0.313	0.175	0.066
Panel B, Sample: Fam. Size ≥ 2, 1st 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: Fam. 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: Fam. Size ≥ 4, 3rd Birth					
OLS Weight	–	–	–	0.678	0.322
IV (Instr: Twin4) Weight	–	–	–	0.882	0.117

Notes: 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, Panel B, C, and D computes the weights of the linear OLS estimates reported in Columns 2,3 and 4 of Panel I in Table (4), and the weights of the linear IV estimates reported in Column 1 of Panel I in Tables 5,6, and 7.

Source: Administrative registers from Statistics Norway.

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 (2010).

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

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

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, i.e. $E[\epsilon_i | X_i, c_i \geq 2] = 0$, 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, since $twin_{3i}$ is defined only for the sub-sample with at least 3 children ($c_i \geq 3$), and is missing for the sub-sample of children from 2 child families $c_i = 2$. Assume that (7) holds, so that $E[\epsilon_i | X_i, c_i \geq 3, twin_{3i}] = E[\epsilon_i | X_i, c_i \geq 3]$.

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

$$E[\epsilon_i z_i | X_i, c_i \geq 2] = E[\epsilon_i twin_{3i} | X_i, c_i \geq 3] pr(c_i \geq 3 | X_i)$$

$$= E[\epsilon_i | X_i, c_i \geq 3, twin_{3i} = 1] pr(twin_{3i} = 1 | X_i, c_i \geq 3) pr(c_i \geq 3 | X_i).$$

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

$$E[\epsilon_i | X_i, c_i \geq 3, twin_{3i} = 1] \neq E[\epsilon_i | X_i, c_i \geq 2] = 0.$$

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

Our missing IV robust strategy first "de-means" $twin_{3i}$ by subtracting its conditional mean. Define the new instrument $twin_{3i}^*$ as

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

This instrument is valid since

$$E[\epsilon_i twin_{3i}^*|X_i, c_i \geq 2] = E[\epsilon_i(twin_{3i} - E[twin_{3i}|X_i, c_i \geq 3])|X_i, c_i \geq 3]$$

$$= E[\epsilon_i twin_{3i}|X_i, c_i \geq 3] - E[\epsilon_i E[twin_{3i}|X_i, c_i \geq 3]|X_i, c_i \geq 3].$$

Given $E[\epsilon_i|X_i, c_i \geq 3, twin_{3i}] = E[\epsilon_i|X_i, c_i \geq 3]$, we have

$$E[\epsilon_i twin_{3i}|X_i, c_i \geq 3] = E[\epsilon_i|X_i, c_i \geq 3]E[twin_{3i}|X_i, c_i \geq 3].$$

Substituting,

$$E[\epsilon_i twin_{3i}^*|X_i] = E[\epsilon_i|X_i, c_i \geq 3]E[twin_{3i}|c_i \geq 3] - E[\epsilon_i E[twin_{3i}|c_i \geq 3, X_i]|X_i, c_i \geq 3].$$

$$= E[\epsilon_i|X_i, c_i \geq 3]\{E[twin_{3i}|X_i, c_i \geq 3] - E[twin_{3i}|X_i, c_i \geq 3]\}$$

$$= 0.$$

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

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 consists 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.

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 $u_i = \alpha x_i + \epsilon_i$. 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 $twin_{2i} = 1$ and $s_i = 3$ if $twin_{3i} = 1$. The observed outcome is then $y_i = \gamma_2 d_{2i} + \gamma_3 d_{3i} + \rho x_i + \epsilon_i$, 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 $twin_{ci} = 1$ 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 non-constant 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: $\epsilon_i \sim N(0, 1)$. The second simulation assumes ϵ_i is distributed according to the Gamma distribution with shape parameter of 2 and scale parameter of 1: $\epsilon_i \sim G(1, 2)$. This parametrization 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 non-parametric model in family size), and iii) IV using the efficient instruments (that is, First Stage II in the non-parametric 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

Table B-1: Simulation Results

Distributional Assumption:	$\epsilon_i \sim N(0,1)$		$\epsilon_i \sim G(1,2)$	
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

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

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 non-parametric IV estimators based on the efficient instruments have lower average absolute value of bias and lower variance across simulations, relative to the non-parametric IV estimator using twin birth instruments directly.

C First Stage Results

Table C-1: First Stage for Table 5, Panel I

Instrument:	$twin_{2i}$	R-Sq
Numb. of Children	0.684	0.1213
	(0.012)	
Instrument:	$\hat{p}_{2i}$	R-Sq
Numb. of Children	1.30	0.1215
	(0.016)	

Notes: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 5. R-Sq is the first stage total R-square. 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 5, Panel II

Instrument:	$twin_{2i}$	$twin^*_{3i}$	$twin^*_{4i}$	$twin^*_{5i}$	R-Sq
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-Sq
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)	

Notes: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 5. R-Sq is the first stage total R-square. 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 6, Panel I

Instrument:	$twin_{3i}$	R-Sq
Numb. of Children	0.763	0.1191
	(0.014)	
Instrument:	$\hat{p}_{3i}$	R-Sq
Numb. of Children	1.23	0.1206
	(0.014)	

Notes: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 6. R-Sq is the first stage total R-square. 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 6, Panel II

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

Notes: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 6. R-Sq is the first stage total R-square. 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 7, Panel I

Instrument:	$twin_{4i}$	R-Sq
Numb. of Children	0.786	0.1066
	(0.019)	
Instrument:	$\hat{p}_{4i}$	R-Sq
Numb. of Children	1.16	0.1080
	(0.016)	

Notes: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 7. R-Sq is the first stage total R-square. 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 7, Panel II

Instrument:	$twin_{4i}$	$twin^*_{5i}$	R-Sq
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-Sq
Siblings ≥ 4	1.016	0.313	0.1088
	(0.013)	(0.019)	
Siblings ≥ 5	0.140	1.06	0.0790
	(0.009)	(0.013)	

Notes: Each row reports the first stage estimate of number of children on twin birth instrument for the indicated column from Table 7. R-Sq is the first stage total R-square. 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.