The Quantity-Quality Trade-off and the Formation of Cognitive and Non-cognitive Skills[*]

Chinhui Juhn[†]
University of Houston

Yona Rubinstein[‡]
London School of Economics

C. Andrew Zuppann[§]
University of Houston

October 5, 2016

Abstract

We estimate the impact of increases in family size on childhood and adult outcomes using matched mother-child data from the National Longitudinal Survey of Youth 1979. We find that families face a substantial quantity-quality trade-off: increases in family size decrease parental investment, decrease childhood cognitive abilities, and increase behavioral problems. The negative effects on cognitive abilities are much larger for girls while the detrimental effects on behavior are larger for boys. We also find evidence of heterogeneous effects by mother's AFQT score, with the negative effects on cognitive scores being much larger for children of mothers with low AFQT scores.

JEL Classification: J13, J24

1 Introduction

What is the impact of the quantity of children in a family on the quality of those children? Following the pioneering work of Becker and co-authors (Becker (1960), Willis (1973), Becker and Lewis (1973), and Becker and Tomes (1976)), a large body of research documents a negative relation between family size and children's educational and labor market outcomes(see Blake (1989) and Hanushek (1992) among others). While the empirical association is well documented, establishing a causal relationship between family size and human capital and earnings is challenging since parental choices are influenced by latent factors that influence children's development via other channels than family size. Rosenzweig and Wolpin (1980) first addressed this problem by using twin births as a natural source of variation in family size.

Other researchers have followed their instrumental variables approach. These studies, while using similar methods, have found conflicting results. On the one hand, Rosenzweig and Wolpin (1980) originally found that increases in family size do indeed lower average schooling of children in Indian families. Similar results have also been found in China in Rosenzweig and Zhang (2009) and in Liu (2014). In the U.S. context, Caceres-Delpiano (2006) and Conley and Glauber (2006) have found, using twin births and sibling-gender composition to instrument for family size, that children in larger families are less likely to attend private school. In contrast to these studies, several prominent studies have found little evidence of a quantity-quality trade-off. Black et al. (2005) use large samples from Norway and find that the negative relationship between family size and education disappears once birth order controls are included or twin births are used to instrument for family size. Similarly, Angrist et al. (2010) use the Israeli Census and alternative instruments - twin births and sex composition of children - and find no relationship between family size and education.

Understanding why the quantity-quality trade-off seemingly holds in some countries but not in others is thus an important question for further study.[1]

Another important aspect of the quantity-quality trade-off that has yet to be fully explored in the literature is the exact mechanisms that underlie the trade-off. Most of the literature to date has focused on adult outcomes such as education and earnings but have not examined directly measures of parental investment which lies at the heart of the original theory by Becker. How is the home environment different in larger families? Are there fewer material resources to go around per child, such as books and toys, or is parental time the critical factor?

In this paper we take a different approach. We take advantage of the mothers of the National Longitudinal Survey of Youth 1979 who are surveyed with their children in the corresponding Children and Young Adult Survey (jointly hereafter NLSY) to evaluate the impact of family size on the formation of cognitive abilities and personality traits. A key feature of the NLSY is that the data contains detailed information about childhood cognitive abilities and non-cognitive traits (measured for ages 5-14 and ages 4-14) biannually from 1986 to 2012, as well as measures of parental investment and home environments. We exploit these rich panel data and make use of the timing of expansions in family size to estimate the dynamics of the quality-quantity trade-off among young children and shed light on parental

---

[1]More recently, researchers have proposed alternative explanations for why conventional instrumental variable methods may erroneously lead to finding no evidence of a quantity-quality trade-off. Mogstad and Wiswall (forthcoming) argue that a linear specification for family size may mask significant heterogeneity in how the trade-off varies for the second birth versus the third birth, etc. Bagger et al. (2013) argue that both conventional OLS and IV approaches for testing the quantity-quality trade-off may have a theoretical limitation because it is impossible to simultaneously vary family size while keeping the distribution of birth orders constant within the household. Thus if an additional child changes the optimal trade-off along the entire birth order distribution OLS regressions at the individual level cannot correctly estimate the quantity-quality trade-off alongside birth order effects. They propose an alternative estimation strategy: estimate birth-order effects using family fixed effects, residualize to net out birth-order effects and then estimate the relationship between average family size and average outcomes, net of these birth-order effects.

investment.

Using child fixed effects estimation strategies we find evidence that families face a substantial quantity-quality trade-off: increases in family size decrease parental investment, decrease childhood performance on cognitive tests and measures of social behavior. Our fixed effect estimates indicate that the arrival of a younger sibling reduces measures of parental investment as well as cognitive and non-cognitive outcomes of older children by approximately one-tenth of a standard deviation. Importantly, we find that these negative effects are not merely temporary disruptions following a birth but in fact persist throughout childhood. When we decompose parental investments into time, resources, affection, and house safety, we find time to be the most critical input that decreases with additional children.

While fixed effect estimates condition out family and child time-invariant factors it does not address the endogeneity of the timing of births. We address this concern in three ways. First, we use an event study approach and check for pre-trends. Next, we utilize the natural gap between conception and birth. Finally, we exploit the plausibly random timing of the NLSY survey with respect to a sibling's birth that allows us to compare first-borns in households that just expanded to their counterparts in households that are about to expand.

We find negligible evidence of pre-trends in parental investment measures, test scores and behavioral problems leading up to births. We also find no evidence of a conception effect. The decline in parental investments and first-born's outcomes are driven by the actual arrival of the younger sibling. Moreover, we find that families appear to time fertility to minimize the negative impact of additional siblings, suggesting that endogenous timing of fertility may in fact be biasing us against finding a quantity-quality trade-off.

The impact of family size may vary by type of parents or households. It might also vary

4

by gender of the children. Indeed we find that the negative effects on cognitive abilities are much larger for girls, while the detrimental effects on behavior are larger for boys. The negative effects on cognitive scores are also much larger among children of mothers with low Armed Force Qualification Test (AFQT) scores within race groups. Last but not least we look at how mothers' employment and hours worked change after childbirth. We find that there are disparate work responses between these same groups; mothers with high AFQT scores are more likely to leave employment and reduce hours worked following the birth of younger siblings. This disparity we find within the U.S. across families with different levels of resources provides a possible explanation for the disparate results documented in the literature across developed and developing countries.

By examining early childhood outcomes, as well as direct measures of parental investment, our paper also makes a natural and important connection between the quantity-quantity literature and recent studies on the importance of early childhood development pioneered by Heckman and co-authors (Cunha and Heckman (2007) and Cunha et al. (2010), see also Almond and Currie (2011) for a survey). Previous studies have found that a child's early environment, both at home and in school, plays a crucial role in the development of cognitive and non-cognitive abilities and that this development has lifelong impacts on later adult outcomes. Our study contributes to this literature by examining how the arrival of additional siblings impact parental decision making and children's early environment influence his or her development of human capital.

The paper proceeds as follows. Section 2 describes our data. Section 3 describes the quantity-quality trade-off in a dynamic setting. Section 4 lays out our empirical strategy. Section 5 presents our main results from fixed effects models, including heterogeneous effects by gender and mother's characteristics. Section 6 addresses the issue of endogenous timing

of births. Section 7 concludes.

2   Data

The National Longitudinal Survey of Youth - 1979 (NLSY) provides a unique opportunity to evaluate the quantity-quality trade-off as it contains information on childhood development as well as adult education and labor market outcomes. We match mothers from the 1979 survey with all their children from the Children and Young Adult survey, resulting in a sample containing 4,925 mothers and 11,464 children. Children were surveyed biannually from 1986 to 2012. By matching children to their mothers we can identify siblings, as well as the precise timing of when family size expands.

The matched NLSY mother-child data contains detailed information about childhood cognitive and non-cognitive abilities as well as longer term outcomes. Children aged 5 to 14 are given Peabody Individual Achievement Tests (PIATs) that measure cognitive skills in mathematics, reading recognition, and reading comprehension.[2] We report results of the combined math and reading recognition scores which were constructed by taking a simple average of the two.[3] To measure non-cognitive abilities, the survey calculates a Behavioral Problem Index (BPI) for children aged 4 to 14 from subindices which measure particular problems including antisocial behaviors, anxiety, dependence, headstrongness, hyperactivity, and social problems. To measure parental investment, the NLSY asks questions to construct a HOME (Home Observation Measurement of the Environment-Short Form) score, "a unique

---

[2]For 1986 to 1992 children aged 5 and older were given PIAT tests. This was changed to children aged 5 to 14 in 1994 and subsequent surveys. We include all respondents in our main analysis; results are robust to excluding children aged 15 and older. A similar change in age eligibility is true for respondents of the BPI as well. Details available upon request.

[3]We have also examined math and reading scores separately. We find somewhat stronger results for math and weaker results for reading when we disaggregate. Results are available upon request.

observational measure of the quality of the cognitive stimulation and emotional support provided by a child's family." Examples of these questions include how many books a child has, how often parents read to the child, and whether parents assist with homework. HOME scores have been shown to be a significant determinant in a child's development(Mott (2004), Todd and Wolpin (2007), Cunha and Heckman (2007), Cunha et al. (2010), and others). Table A.1 presents summary statistics of the children in our matched mother-child sample.

2.1 Sibling-pair data

For our fixed effects estimates, we construct a "sibling-pair" data set where we match test scores, HOME and behavioral index measures of older siblings with the birthdate of each subsequent younger sibling. We use the whole sample of children to estimate the effects of age and other demographic variables. However, since our fixed effects estimation relies on before-after comparisons of outcomes for the same child, our main coefficient of interest is identified from a sub-sample of children with valid observations before and after a sibling's birth. While this is not an issue for HOME scores which are measured since birth, the behavioral problem index is measured for children aged 4-14 and NLSY PIAT tests are only administered to children aged 5-14. For test scores, older children who identify the younger sibling effect are significantly older than the average child at the birth of a younger sibling.

Table A.2 shows summary statistics of demographic characteristics of mother and children in both the full sample of older siblings and our most restricted sample of older siblings who have valid PIAT test scores before and after the arrival of a younger sibling. Not surprisingly, children in the restricted sample have substantially greater spacing between births (26 months). However on other dimensions these children appear similar to the full sample. Mothers have similar cognitive and non-cognitive ability measures, have a similar

7

distribution of race, and start their fertility at similar ages. Children in the restricted sample are of similar birth weight and slightly less likely to be living in a blended family.

3 The quantity-quality tradeoff in a dynamic setting

A simple cross-sectional analysis of family size on child outcomes is confounded by differential selection as larger families are likely to differ from smaller families in other important ways. Additionally, there may be concerns of reverse causality. Child outcomes themselves may affect whether parents choose to have an additional child. For example, parents may follow an optimal stopping rule where if the just-born child is of high quality, they decide to have an additional child. The opposite may also occur if parents decide to have additional children to compensate for a just-born child who is of low quality.

While these issues may be addressed by instrumenting family size with twin births, there are a number of problems with using twins in our data. First, the birth of twins is a rare event and our small sample sizes of twin births limit our ability to estimate the family size effect with precision, particularly when we estimate by parity, our preferred specification. An additional concern is that twins may have a direct effect on other siblings through the close spacing of births or by being of poorer health and drawing resources away from other siblings (Angrist et al. 2010). These concerns suggest that the twins instrument may fail the exclusion restriction and therefore may be an invalid instrument for family size.[4]

In light of these concerns, we take a different approach.[5] The panel structure of the

---

[4]Another possible instrument for family size - sibling sex composition of the first two children as an additional instrument, a method used by Angrist et al. (2010) - also should not be used with our data. In the NLSY matched mother-children data this instrument is weak - the F-test in the first stage is only 5.83.

[5]A curious reader may still wonder what estimates we would obtain using the presence of twins in the household as an IV despite the aforementioned weaknesses. In the NLSY matched mother-children data a twins IV estimation strategy generally finds the presence of a quantity-quality tradeoff during childhood as

8

NLSY gives us multiple observations on the same child and we can examine how the arrival of siblings impact cognitive and non-cognitive scores as well as the measure of the HOME environment of the older child. Issues of selection or reverse causality are addressed by controlling for individual child fixed effects.

Including child fixed effects provides a way to test how the quantity-quality tradeoff actually manifests in children over their childhood. The bulk of previous research, starting with Becker's original model of the quantity-quality tradeoff, has been essentially static. Parents simultaneously choose both the number of children and the average quality of their children. We extend the original model introduced in Becker and Lewis (1973) and Becker and Tomes (1976)) to a dynamic setting in which parents invest in the "quality", i.e. human capital, of children in each period. We adopt the key feature of the original model which is that the marginal cost of investing in average quality of children increases with the number of children. Below we lay out the key implications of this model on the time path of parental investments and human capital accumulation of older siblings, implications which we can then take to the panel data available in the NLSY.

At each time t parents choose investment (I_i) in their children, $i = 1, ..., N$ as a function of their total wealth/resources (M_t) and the stock of current child human capital/quality (q_i).

Human capital for each child follows an *accumulating* production process

$$q_{t+1} = f(I, q_t)$$

By accumulating we mean that childhood human capital is fundamentally a stock and de-

---

measured by test scores and parental investments and then into adulthood as measured by years of education and wages. The estimated effects are imprecisely measured, likely due to the small number of twins in our dataset. Details of these estimates are provided in Appendix 1.

pends on the past history of parental investments.[6] This implies that parental investments today lead to higher long-run levels of child quality. We also assume that children with greater human capital stock have higher marginal returns from investment and that there are diminishing marginal returns of investments. That is, we assume $f_{qI} \geq 0$ and $f_{II} \leq 0$.

The marginal cost of investing in each child depends on the number of children in the household: $c(n_t)$. Having additional children increases the shadow price of investing in any particular child, so $c' > 0$. This reflects the fundamental quantity-quality logic that larger families are, *ceteris paribus*, more constrained in spending their time and resources in all their children so the opportunity cost of investing in any one particular child is increased.

Following the expansion of family size this lowered investment in older siblings results in permanently lower levels of human capital. This is true for two reasons: First, the higher cost of investment persists throughout the rest of childhood leading to lower per-period investment, which leads to a lower long-run human capital levels due to the accumulative nature of human capital. Secondly, possible dynamic complementarities between investment across childhood means that a lowered human capital growth path may further lower optimal parental investment in child quality.

As an illustration of how the quantity-quality tradeoff manifests via lowered parental investment consider an example where human capital is produced according to

$$f(I, q) = \delta \cdot q + ln(I)$$

and the marginal cost of investing in any child is multiplicative in the number of children in

---

[6]This is a fairly weak assumption, as it only rules out that human capital does not depend at all on parental investments or that human capital depends *only* contemporaneously on parental investment.

the household:

$$\theta_i(n) = \gamma \cdot n$$

In figures 1 and 2 we plot optimal parental investment and total human capital for the first child in the household in three scenarios: one with no siblings, one with a sibling born at age 4, and one with a sibling born at age 8.[7]

We see that the arrival of a younger sibling results in lowered investment in the older sibling as well as places the firstborn on a significantly lower path of human capital accumulation. Further, this simple framework shows that shorter spacing between the older and younger siblings has a negative impact on long-run outcomes of the older child, even though spacing does not effect investment levels *per se*. This is consistent with the literature that finds short spacing between births is disadvantageous for older siblings (Buckles and Munnich 2012).

This simple model provides testable predictions about how the quantity-quality tradeoff manifests in a dynamic setting and how parental investment and measures of child development should evolve over time. First, the birth of another sibling should lower parental investment in older children when compared to levels prior to that birth. As parental time and resources are limited, a family that increases in size now faces a tradeoff between time spent with the younger sibling and the older sibling that was not present before. Of course, parents who are precisely planning the timing of all their children's birth may recognize that they will be more resource-constrained in the future so, if parents are forward-looking they may increase investment in their older child directly prior to a younger sibling birth.

Second, the tradeoff should not just appear as a temporary shock to parental invest-

---

[7]See Appendix for mathematical details in solving for optimal investment decisions.

11

ment in the first months and years following the arrival of a younger sibling but instead in a permanently level of investment and correspondingly worse cognitive and non-cognitive outcomes. The quantity-quality tradeoff is not just a statement that the first year of a younger child's life is a disruptive event in the household. Instead, the time and resource allocation of the parents has fundamentally changed as the family grew in size. Further, as parental investment is an important input into the production of childhood human capital, we expect that children will see their outcomes worsen even more over a longer time horizon as the accumulation of more and more years with reduced investment translate into worse and worse outcomes. Even without taking an explicit stand on what the exact production technology is that transforms parental investment into human capital, so long as there is any accumulation of human capital over time we would expect to find increasingly worse outcomes over time following the birth of a sibling.[8]

4   Empirical strategy

Empirically then, we can test a dynamic version of the quantity-quality tradeoff by using multiple observations of the same child and measure how older siblings' parental investment and childhood outcomes change before and after the birth of a younger sibling. Or more simply, we can include child fixed effects. The panel structure of the NLSY gives us multiple observations on the same child and this allows us to examine how the arrival of siblings impact cognitive and non-cognitive scores as well as the measure of the home environment.

We first investigate how the arrival of a younger sibling affects older children via an event-study framework where we estimate coefficients on year dummies prior to and after

---

[8]Previous work by Todd and Wolpin (2006) has found that the NLSY HOME measures at young ages do lead to lowered cognitive test scores at older ages.

the arrival of the additional sibling while controlling for child fixed effects. Our estimating equation is the following:

$$Y_{ijt} = \beta_0 + \sum_{r=-R}^{-1} D_{ijt}^r \delta_r + \sum_{r=1}^{R} D_{ijt}^r \delta_r + X_{it}\beta_2 + \lambda_i + \varepsilon_{ijt} \tag{1}$$

where D_{ijt}^r is a dummy variable equal to 1 if the survey date, t, is r months relative to sibling birth month at parity j, with the birth year $(r = 0)$ being the omitted category.

To improve our precision, we also group the many after years together and re-estimate a simple before vs after fixed effects model. Our baseline specification to measure the extent to which parental investment and child outcomes change following the arrival of a younger sibling in the household is:

$$Y_{ijt} = \beta_0 + \beta_1 1\{after\}_{ijt} + X_{it}\beta_3 + \lambda_i + \varepsilon_{ijt} \tag{2}$$

where we compare an older child before and after the arrival of a younger sibling.

In a dynamic setting, the decline in parental investment following a family size expansion will manifest in lower outcomes over time - controlling for age effects - as children will have faced more and more years of lowered investment. This is in contrast to an alternative hypothesis where an additional sibling is temporarily disruptive in the first few months or year after her birth but the family environment returns to the pre-birth state after a while. To test for differential effects between the short and long run we estimate two specifications.

First, we look directly at patterns of long-run accumulation of investment and human capital by allowing for a level difference in outcomes directly after birth and then including

a linear term in the number of years following the arrival of a sibling.

$$Y_{ijt} = \beta_0 + \beta_1 1\{1yearafter\}_{ijt} + \beta_2\{\text{Years after}\}_{ijt} + X_{it}\beta_3 + \lambda_i + \varepsilon_{ijt} \tag{3}$$

In this specification, β_1 is the immediate change in outcomes and β_2 measures how that changes over time.

Second, we estimate both a short-run and long-run effect of an additional sibling by estimating the change in the first 3 years after the birth and the change in all the years more than 3 after the birth:

$$Y_{ijt} = \beta_0 + \beta_1 1\{0-3 \text{ years after}\}_{ijt} + \beta_2 1\{3+\text{years after}\}_{ijt} + X_{it}\beta_3 + \lambda_i + \varepsilon_{ijt} \tag{4}$$

where β_1 measures the change in investments and outcomes that occur in the 3 years following a sibling birth and β_2 captures the longer run changes that occur in all later years.

5 Results

Figure 3 graphs the event-study estimates, including both pre and post trends, in our outcome measures. While our confidence intervals are large, we see that following the arrival of a younger sibling, older children receive less parental investment, have worsened cognitive outcomes, and have increased behavioral problems. Further, it looks like these negative effects on the older child are not temporary. If anything, the effects worsen the further away from the younger child's birth we measure, providing suggestive evidence of the deceleration of human capital growth following an additional child. We also find little evidence of pretrends which may be biasing our results. There is little evidence of an upsurge in parental

14

investments or a change in the older child's achievement preceding a sibling birth. One possible explanation is that parents are not forward-looking when it comes to investment in their firstborns.

The estimates presented in Figure 3 are somewhat imprecise. Table 1 presents estimates from fixed effects models of the impact of an additional child on older children's outcomes at second birth parity.[9] Column 1 estimates the average decline in test scores and parental investments after the second child is born. We see that parental investment in the older child, as measured by the HOME Inventory index, falls by 3 percentile points after the birth of a younger child. Older siblings are generally worse off in both cognitive and non-cognitive outcomes following the birth of a younger child. Cognitive scores falls by 2.8 percentile points which is slightly larger than one-tenth of a standard deviation. Similarly, behavioral problems also increase, with the index increasing by more than one-tenth of a standard deviation.

Column 2 shows how the presence of an additional child in the household leads to a long-run lowering of child quality. Parental investment immediately falls by 2.9 percentile points following the arrival of a new sibling. This lowered investment persists at the same level in the long-run; the estimated effect of every additional year after birth is small and insignificant. In cognitive and non-cognitive outcomes, however, a different pattern emerges. Although there is an immediate negative effect in both outcomes directly after a sibling birth, we find that these effects not only persist as time passes but actually worsen! Column 3 of Table 1 summarizes the results by contrasting the short run (the first 3 years following a sibling birth) and the long run (the remaining years after 3 years). Not only do we fail to

---

[9]Results at third parity are reported in Table A.3. We find that, at third parity, the results for HOME score and behavioral problems remain significant and while cognitive test results are still negative they are not significant.

find evidence that the impact is transitory, effects appear to substantially worsen over the longer run.

Our findings are in line with our model predictions for how the quantity-quality tradeoff manifests in practice. Upon the arrival of an additional sibling, parents are faced with increased resource constraints in the household and permanently lower investment in the older child. This leads to a deceleration in human capital growth and apparent worsening of outcomes relative to before the younger sibling was born.

5.1 Understanding the decline in parental investment: HOME subcomponents

The matched mother-child data of the NLSY provide rich detail on the home environment. We utilize this data to shed further light on what aspect of the home environment changes as an additional sibling is introduced. We decompose the HOME index into four broad sub-categories corresponding to time, resources, affection, and home safety. We chose a set of questions for each of these aspects of parenting that are asked to children in the 3-14 age range in the NLSY. We chose questions that both (a) are available and comparable across all child ages from 3 to 14 and (b) can reasonably be categorized as representing time, resources, affection, or safety. For each of these categories we then form a single measure by performing factor analysis on the questions and constructing the first factor that explains the most variance between the questions. The details of each question are listed in Appendix Table A.4.[10]

Table 2 reports our estimates of the impact of an additional child on these different

---

[10]Exploratory factor analysis on a larger set of questions in the HOME Index suggests that grouping these questions together into these separate clusters is a good fit in explaining the patterns of variation in the overall HOME Index. Details of this factor analysis are available upon request.

16

measures of parental investment. The table shows that parental time is the most critical input that falls after the arrival of an additional child. The coefficient -0.107 indicates that our index measure of parental time falls by approximately one tenth of a standard deviation. We find little evidence of a decline in other measures with the possible exception of the home safety measure in the longer run.

5.2 Heterogenous effects across gender, race, and mother's AFQT

One advantage of the NLSY sample is that we have detailed characteristics on the mothers of these children and can investigate whether there is heterogeneity in the impact of an additional child across different types of children and families. To do this we interact the child's and mother's characteristics with the after dummy in our fixed effects regressions.

Table 3 reports the effects by gender. We report estimates of a pooled specification where age effects are restricted to be equal across gender. The table shows different family size effects by gender. The effects on cognitive scores are small and insignificant for boys. They are large and significant for girls. In contrast, the family size effect is larger for boys for "Behavioral Problems." These latter results are consistent with Bertrand and Pan (2013) who find that disruptive behavior of boys are particularly sensitive to adverse shocks to the family environment.

Table 4 reports the effects separately by mother's AFQT group. There are strikingly different results by mother's AFQT score. For children with mothers with below median AFQT scores, the arrival of younger siblings have large and significantly negative effects on cognitive skills while for children with mothers with above median score, the effects are much smaller and not significant. With regards to "Behavioral Problems" index, however, the results are opposite. It is the children of high ability mothers who are more likely to

act up when a younger sibling arrives. Mother's AFQT effect could be confounded by race. Column 2 examines white mothers only, where median AFQT is constructed conditional on the race of the mother. Even conditioning on race, we find a strong AFQT gradient, with the HOME index and cognitive scores falling more for children with mothers with below median AFQT scores, and behavioral problems increasing more for children with mothers with above median AFQT scores. Among black mothers shown in Column 3 we continue to observe a strong AFQT gradient for cognitive scores although no differential effects by AFQT for Home index and behavioral outcomes.

The differential effects by mother's AFQT score provide insight into why empirical papers have found a quantity-quality trade-off in some, but not in all, cases. In countries with a comprehensive welfare system and a strong public education system, parental resource constraints may not be binding for educational outcomes. In a country such as the U.S. with limited income support programs and limited quality public education, lower ability mothers with limited monetary resources may face a real trade-off between quantity of children and the resources she can devote to invest in their skills.

To investigate this hypothesis that mothers of different cognitive ability face different resource constraints and that this effect leads to the heterogeneity in the quantity-quality trade-off we look at how maternal work behavior adjusts in response to increases in family size. Table 5 presents estimates of the effect of an additional sibling on whether the mother worked last week in the top panel and total annual hours worked in the bottom panel. We see that there are differential responses by mother's AFQT. As shown in the top panel, mothers who have below median AFQT scores do not respond to an additional child by reducing work while mothers above the median do. When we examine hours worked last year in the bottom panel, we see declines for both mothers above and below AFQT scores.

18

However, examining what happens over the long run in Column 3, we see that hours worked diverge with above median AFQT mothers reducing hours even further in the longer run. This evidence is consistent with maternal resources being an important factor in whether there is a quantity-quality trade-off among children.

6   Endogenous birth timing

Our fixed effects estimates control for all unobservable characteristics of children that are constant over time but do not address the issue of endogenous timing of fertility. Time-varying shocks may both drive parents' decision of whether and when to have a child as well as affect older sibling's test scores. It may be the case that parents choose to have additional children when the negative impact on older children is minimal. For example, they may choose to have additional children at the time of an expected promotion or when a grandparent retires. Alternatively, parents may choose to have additional children when times are good in terms of either household income or older children's outcomes. If these shocks are transitory, however, then the household environment will worsen due to regression to the mean. In this case we would be overstating the negative impact of additional siblings. A related concern is that parents make compensatory investments in older children prior to the birth of an additional sibling. In this case, the amount of parental investments in the "before" period may reflect a temporary upsurge, thereby exaggerating the drop in parental investment in the "after" period.

We conduct three exercises to assess and account for time-varying unobservable shocks that are correlated with the decision to have an additional child. First, we check for pre-trends in our outcome variables prior to the birth of the sibling. Second, we estimate our

model in first-differences to distinguish between the timing of the decision to get pregnant and the actual arrival of an additional child. Third, we utilize the randomness of the timing of NLSY interviews with respect to pregnancies and births to compare families observed just prior to the birth of a child to those observed just after the birth of a child.

6.1 Pre-trends

To examine whether endogenous timing is a concern we first check for pre-trends using the event-study framework we specified in Equation 1. As discussed previously, Figure 3 graphs the pre and post trends in our outcome measures. While our confidence intervals are large, we find little evidence of pre-trends which may be biasing our results. There is little evidence of an upsurge in parental investments or a change in the older child's achievement preceding a sibling birth.

6.2 First-differences with controls for timing of pregnancy decision

The decision of when to have an additional child may be correlated with unobservable time-varying shocks. However, there is a built-in time lag between when parents decide to have an additional child and when the child actually arrives. Our second method exploits this time lag to distinguish between the impact of actual sibling arrival from the timing of the pregnancy decision. To more precisely identify these separate channels, we switch to a first-differences specification. While we do not directly observe when parents decide to start a pregnancy, we use mother's pregnancy to proxy for the timing of the pregnancy decision.

Let P_{ijt} be an indicator of a household's decision to have a child of parity j (to older

20

child i), which is equal to 1 when a household decides to have an additional child and is then 1 forever. S_{ijt} is an indicator variable equal to 1 if the younger sibling has arrived. Taking first differences, we have

$$\Delta Y_{ijt} = \beta_1 \Delta S_{ijt} + \beta_2 \Delta P_{ijt} + \nu_{ijt} \tag{5}$$

We do not directly observe P_{ijt} as it is when the parents decide to start a pregnancy (and not when the mother actually becomes pregnant). In what follows, we proxy for P_{ijt} with an indicator variable which equals 1 once the mother becomes pregnant with the child of parity j, approximately 10-months before the arrival of the child.[11] Note that our primary interest is not in the pregnancy state itself but in the timing of the decision to have an additional child. As such, our indicator variable is turned on to equal 1 when the mother becomes pregnant and is then 1 forever.

Estimates of this first-differences specification with and without the control for "Pregnancy occurred" are presented in Table 6. Column 1 presents baseline estimates of our first-differences; we see similar estimates as from the fixed effects estimates. Column 2 presents estimates when we regress an older child's change in outcomes on the change in "Pregnancy occurred" variable. According to Column 2, in the 10th month before birth, it appears that older children have worse cognitive test scores and actually improved behavioral outcomes even though neither effect is statistically significant. Column 3 reports estimates including both variables. The important takeaway is that the actual arrival of the sibling remains negative and significant even when we include a control for the timing of the pregnancy decision.

---

[11]We have also investigated 9 and 12-months leads to approximate pregnancy and found similar results.

6.3 Comparing just after with just before

In general our identification strategy can be thought of as exploiting the timing of the arrival of the sibling to identify the effect of family size for those families who have a child at parity j. Using notation from the event study framework, we have:

$$Y_{ijt} = \beta_0 + \sum_{r=-R}^{-1} D_{ijt}^r \delta_r + \sum_{r=1}^{R} D_{ijt}^r \delta_r + X_{it}\beta_2 + \varepsilon_{ijt} \tag{6}$$

where again $r = 0$ refers to the birth month. The "treatment" effect is the average of the δ_r coefficients in the after period relative to the average in the before period. The identifying assumption is that there are no omitted factors in the error term that are correlated with Y_{ijt} and the timing of the sibling arrival. One possible violation of this assumption is if there is an individual specific component of the error term that is correlated with the timing of sibling birth. This problem can be addressed in our fixed effects framework. As already discussed, however, another violation of the identifying assumption which cannot be addressed with fixed effects is time-varying shocks that are correlated with both the timing of the sibling birth and Y_{ijt}. In this sub-section we assess the extent that time-varying latent factors contaminate our fixed effects and first differences estimates.

Our basic approach is to compare families who make "similar" timing decisions. In other words we compare outcomes of the first born children whose sibling was "just born" to first born children whose sibling is "about to be born." The idea is to compare families with as similar as possible spacing between first and the second born. The smaller the gap in spacing between the treatment and the control group the smaller should be the bias associated with the timing of the second birth.

Formally, we estimate the following equation:

$$Y_{ijtr} = \beta_0 + \beta_{1,r}1\{after\}_{ijt,r} + X_{it}\beta_2 + \varepsilon_{ijt} \tag{7}$$

where Y_{ijtr} is again the outcome of interest for an older child i at younger sibling birth parity j measured sometime between r months before and 12 months after the birth of the 2nd child and t refers to the interview date. We run our regression on a sample where the child is interviewed between r months prior to the sibling's birth and 12 months after the sibling's birth at parity j. $1\{after\}_{ijt,r}$ is a dummy variable equal to 1 if the child is interviewed between 0 to 12 months after the sibling's birth.[12]

Random assignment across "before" and "after" categories is more plausible if we restrict the time difference between the interview and the sibling's birth to a narrow window.[13] In our main results, we set r equal to 12 months. We also conduct exercises where we increase r, the size of the window in the "before" period. In effect, in these exercises we compare older children who "just had" a sibling to all older children who will have a sibling sometime within the next 2, 3, or 5 years. As our window of comparison, r, increases, the

---

[12]More specifically, we estimate the effects of child age, birth spacing, and other characteristics of the mother using the entire sample of mothers and children. We calculate the residuals and estimate $\beta_{1,r}$ on the sub-sample of children who are interviewed between r months prior to the sibling's birth and 12 months after the sibling's birth.

[13]For such a narrow window, the precise timing of births varies due to two plausibly random events: the timing of when the NLSY interviews a woman within a calendar year and the inability of parents to precisely time when their pregnancy will occur. The NLSY conducts interviews in several months throughout the calendar year of a particular survey wave. Although the NLSY attempts to keep the month of interview constant for a single respondent this is often not possible for many reasons. Most notably the NLSY often changes the months of surveying from wave to wave (e.g., from June-December in 1994 to April-October in 1996). The month of interview can also change from wave to wave because of problems contacting the respondent or due to organizational/budgeting delays. Another source of plausible randomness is that parents are unable to precisely time the birth of a child to a particular month. Even if we assume that all pregnancies are planned there is still a great deal of uncertainty about when the pregnancy and birth will actually occur. Of course, parents are capable of moving the date of birth by a few days or a week through induced labor or voluntary Caesarean sections but this will change, at most, one month.

23

assumption of random assignment into before and after categories becomes worse and more selection bias resulting from endogenous timing is introduced. The change in estimates of the quantity-quality trade-off as we increase r therefore tells us about the direction of bias due to endogenous timing.

We report our findings in Table 6 and Figure 4. Two main facts emerge. First, estimates are similar in magnitude to fixed effects and first differences, but only when we restrict the sample to a narrow window. Second the gap between treated and control groups shrinks, in absolute terms, as we expand the experiment window. Column 1 of Table 7 shows estimates with $r = 12$ where we restrict the sample to those observations where the interview date is within 12 months before or 12 months after a sibling birth. Estimates of the sibling effect are negative for all three measures. Given the small samples, however, we lose precision and the effect on cognitive scores is no longer significant. The effects on HOME index and behavioral problems, however, remain. In Column 2 and Column 3 we broaden the window in the "before" period to 24 and 36 months while keeping the "after" period to 12 months. When we extend the "before" period, the sibling effect becomes less negative which suggests that families time fertility decisions as to minimize the negative impact on older children.

To further illustrate this pattern Figure 4 shows how the estimated effect of an additional sibling varies as we expand the period of time before birth (i.e. as we increase r). The highlighted numbers in the figures correspond to those already shown in Table 7 and the dotted lines refer to the 95 percent confidence intervals around our estimates. While the estimates are somewhat imprecise, the figure shows that the effect of an additional sibling becomes smaller as we expand the window in the "before" period, thereby allowing for more endogenous timing of fertility. The evidence here suggests that endogenous timing of fertility biases our coefficients to be too small, and therefore, biases us against finding a

quantity-quality trade-off.

As a check on whether the before and after assignments for our narrowest window, $r = 12$, are random, Table A.5 shows average mother and child characteristics of the before and after groups in the years before they entered the sample. If previous shocks to child test scores influenced the timing of births within our narrow window then we would expect to find differences across these groups. Looking at Table A.5 it does not appear there are any significant differences in either maternal characteristics or child test scores across these groups.

7   Conclusion

Using a variety of approaches we have documented a significant trade-off between quantity and quality of children for NLSY mothers and their children. On average, children in larger families have lowered parental investment and worse cognitive and non-cognitive outcomes. We found this trade-off in three separate specifications that control for possible confounds. First, we examined older children in households that experienced an exogenous shock to family size due to twin births. Second, we compared older children to themselves before and after the arrival of a younger sibling. Finally, we examined older children who were surveyed within a narrow window of a sibling birth with the "before" and "after" status being plausibly random.

These results differ substantially from previous research so it is worthwhile to consider why our estimates are so different. One potential explanation for our results in contrast to (Black et al. 2005) may be institutional differences between Norway and the U.S. In particular, at the margin, parental investments may matter more in the U.S. where a substantial

fraction of young men and women, particularly from lower income backgrounds, are at risk of not finishing high school. A recent paper by (Black et al. 2010) offers some support for the idea that the particular country and the particular cohort examined matters. In contrast to their earlier paper which examined an older cohort, their recent paper finds a negative impact of family size on IQ among younger birth cohorts in Norway. Likewise, Li et al. (2008) also find that the negative family size effect on schooling is particularly strong in rural areas of China where the public education system is less well developed.

Although we consistently find, on average, the presence of a quantity-quality trade-off in the NLSY sample, the strong differences across mother's AFQT score provide suggestive evidence of the importance of resource availability or institutions in determining the extent of the quantity-quality trade-off. High AFQT mothers of all races appear to face less of a trade-off than mothers with low AFQT scores. Differences in mother's AFQT scores are correlated with a wide set of lifestyle differences that could explain these differences. For instance, having worse child care coverage, maternity policies, or flexibility in household labor supply could all make the presence of an additional child more detrimental to other children in the household. Understanding the mechanisms underlying these differentials in the quantity-quality trade-off remains an important question for future research.

References

Douglas Almond and Janet Currie. *Human Capital Development before Age Five*, volume 4 of *Handbook of Labor Economics*, chapter 15, pages 1315–1486. 2011.

Josh D. Angrist, Victor Lavy, and Analia Schlosser. Multiple experiments for the causal link

between the quantity and quality of children. *Journal of Labor Economics*, 28:773–824, 2010.

Jesper Bagger, Javier A. Birchenall, Hani Mansour, and Sergio Urzua. Education, birth order, and family size. Nber working papers, National Bureau of Economic Research, Inc, June 2013.

Gary S. Becker. An economic analysis of fertility. *Demographic and Economic Change in Developed Countries*, 1960.

Gary S. Becker and H. Gregg Lewis. On the interaction between the quantity and quality of children. *Journal of Political Economy*, 81:S279–288, 1973.

Gary S. Becker and Nigel Tomes. Child endowments and the quantity and quality of children. *Journal of Political Economy*, 84:S143–162, 1976.

Marianne Bertrand and Jessica Pan. The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*, 5: 32–64, 2013.

Sandra E. Black, Paul J. Devereux, and Kjell G. Salvanes. The more the merrier? the effects of family size and birth order on children's education. *The Quarterly Journal of Economics*, 120:669–700, 2005.

Sandra E. Black, Paul J. Devereux, and Kjell G. Salvanes. Small family, smart family? family size and the iq scores of young men. *Journal of Human Resources*, 35:33–58, 2010.

Judith Blake. *Family Size and Achievement*. University of California Press, 1989.

Kasey Buckles and Elizabeth Munnich. Birth spacing and sibling outcomes. *Journal of Human Resources*, 47:613–642, 2012.

Julio Caceres-Delpiano. The impacts of family size on investment in child quality. *Journal of Human Resources*, 41:738–754, 2006.

Dalton Conley and Rebecca Glauber. 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*, 41:722–737, 2006.

Flavio Cunha and James J Heckman. The technology of skill formation. *American Economic Review*, 97(2):31–47, May 2007.

Flavio Cunha, James J Heckman, and Susanne M. Schennach. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931, May 2010.

Eric Hanushek. The trade-off between child quantity and quality. *Journal of Political Economy*, 100:84–117, 1992.

Hongbin Li, Junsen Zhang, and Yi Zhu. Cheap children and the persistence of poverty. *Demography*, 45:223–243, 2008.

Haoming Liu. The quality-quantity trade-off: evidence from the relaxation of china's one-child policy. *Journal of Population Economics*, 27:565–602, 2014.

Magne Mogstad and Matthew Wiswall. Testing the quantity-quality model of fertility: Estimation using unrestricted family size models. *Quantitative Econoomics*, forthcoming.

Frank L. Mott. The utility of the home-sf scale for child development research in a large

national longitudinal survey: The national longitudinal survey of youth 1979 cohort. "Parenting: Science and Practice", 4(2-3):259–270, 2004.

Mark R. Rosenzweig and Kenneth I. Wolpin. Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica*, 48:227–240, 1980.

Mark R. Rosenzweig and Junsen Zhang. Do population control policies induce more human capital investment? twins, birth weight and china's one-child policy. *Review of Economic Studies*, 76:1149–1174, 2009.

Petra E. Todd and Kenneth I. Wolpin. The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human Capital*, 1(1):91–136, 2007.

Robert J. Willis. A new aproach to the economic theory of fertility behavior. *Journal of Political Economy*, 81:S14–64, 1973.

Table 1: Individual fixed effects: 2nd parity

Panel A: Parental Investment

		(1) Child FEs	(2) Child FEs	(3) Child FEs
After sibling birth	HOME Inventory	-3.030** (0.986)	-2.967** (0.989)	
Every additional year			-0.514 (0.724)	
0-3 years after				-3.066** (0.986)
3+ years after				-4.925*** (1.306)
N		14218	14218	14218

Panel B: Ability Measures

		(1) Child FEs	(2) Child FEs	(3) Child FEs
After sibling birth	Cognitive Ability	-2.812** (1.044)	-2.526** (1.045)	
Every additional year			-2.812*** (0.581)	
0-3 years after				-2.519** (1.056)
3+ years after				-4.011*** (1.183)
N		10645	10645	10645
After sibling birth	Behavioral Problems	3.875** (1.194)	4.053*** (1.213)	
Every additional year			0.167** (0.0789)	
0-3 years after				3.960*** (1.196)
3+ years after				2.870** (1.424)
N		11897	11897	11897

Each row presents estimates for a different outcome of interest. Controls are child's age at test and child fixed effects. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.

∗: significant at 10% level. ∗∗: significant at 5% level. ∗ ∗ ∗: significant at 1% level.

Table 2: Effects for HOME subcomponents: 2nd parity

		(1) Individual FEs	(2) Individual FEs
After sibling birth	HOME, time with child	-0.107***	
		(0.0217)	
0-3 years after			-0.0740***
			(0.0219)
3+ years after			-0.172***
			(0.0232)
N		7095	7095
After sibling birth	HOME, resources spent	-0.0269	
		(0.0300)	
0-3 years after			-0.00678
			(0.0308)
3+ years after			-0.0541*
			(0.0315)
N		10361	10361
After sibling birth	HOME, parental quality/affection	0.0151	
		(0.0308)	
0-3 years after			0.0178
			(0.0313)
3+ years after			0.0113
			(0.0331)
N		8096	8096
After sibling birth	HOME, house environment	-0.0475	
		(0.0334)	
0-3 years after			-0.00785
			(0.0350)
3+ years after			-0.0755**
			(0.0344)
N		12020	12020

Each row presents estimates for a different outcome of interest. Controls are child's age at test and child fixed effects. The indices are constructed as the fitted (first) factor following a factor analysis on a set of questions for each subcomponent. See Table A.4 for list of questions used to construct each index. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.

∗: significant at 10% level. ∗∗: significant at 5% level. ∗ ∗ ∗: significant at 1% level.

Table 3: Fixed effects estimates by gender - 2nd parity

Panel A: Parental Investment			
		Individual FEs	Difference (Std. Err.)
Boys	HOME Inventory	-3.113**	
		(1.184)	
Girls		-2.943**	0.170
		(1.230)	(1.390)
N		14218	14218
Panel B: Ability Measures			
		Individual FEs	Difference (Std. Err.)
Boys	Cognitive Ability	-1.025	
		(1.290)	
Girls		-4.947**	-3.922*
		(1.618)	(2.019)
N		10645	10645
Boys	Behavioral Problems	4.430**	
		(1.653)	
Girls		3.277**	-1.153
		(1.642)	(2.273)
N		11897	11897

Each row presents estimates for a different outcome of interest. Regressions include controls for child age and child fixed effects. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.

∗: significant at 10% level. ∗∗: significant at 5% level. ∗ ∗ ∗: significant at 1% level.

Table 4: Fixed effects estimates by mother's AFQT - 2nd parity

Panel A: Parental Investment		Individual FEs	White mothers	Black mothers
Below median	HOME Inventory	-3.260**	-4.884**	-0.419
		(1.507)	(1.717)	(2.621)
Above median		-2.943**	-3.056**	0.0169
		(1.103)	(1.333)	(2.187)
N		14218	7690	3813
Difference		0.317	1.828	0.436
(Std. Err.)		1.635	1.827	3.035

Panel B: Ability Measures		Individual FEs	White mothers	Black mothers
Below median	Cognitive Ability	-6.006***	-5.953**	-5.555**
		(1.822)	(2.323)	(2.464)
Above median		-1.035	-0.211	2.118
		(1.219)	(1.622)	(2.010)
N		10645	5498	3091
Difference		4.971	5.742	7.672
(Std. Err.)		2.149	2.784	3.086
Below median	Behavioral Problems	1.921	-0.294	6.183**
		(1.714)	(2.143)	(3.124)
Above median		4.926**	6.577**	6.800**
		(1.541)	(2.103)	(2.110)
N		11897	6276	3298
Difference		3.006	6.872	0.617
(Std. Err.)		2.252	2.945	3.636

Each row presents estimates for a different outcome of interest. Regressions include controls for child age and child fixed effects. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. Mother's median AFQT is constructed conditional on the race of the mother. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.

∗: significant at 10% level. ∗∗: significant at 5% level. ∗ ∗ ∗: significant at 1% level.

Table 5: Impact on maternal labor supply

	Whole sample	By AFQT	Short vs. Long Run
Worked last week (CPS)	-0.0981***		
	(0.0291)		
Below median		-0.0409	
		(0.0383)	
Above median		-0.122***	
		(0.0333)	
Below median, 0-3 years after			-0.0512
			(0.0397)
Below median, 3+ years after			-0.0274
			(0.0459)
Above median, 0-3 years after			-0.119***
			(0.0337)
Above median, 3+ years after			-0.119**
			(0.0427)
N	8451	8451	8451
Mean of dependent variable	0.383	0.383	0.383

	Whole sample	By AFQT	Short vs. Long Run
Hours last year	-357.9***		
	(32.05)		
Below median		-297.4***	
		(45.91)	
Above median		-381.0***	
		(35.69)	
Below median, 0-3 years after			-371.8***
			(48.57)
Below median, 3+ years after			-280.8***
			(55.00)
Above median, 0-3 years after			-358.6***
			(36.24)
Above median, 3+ years after			-421.5***
			(45.81)
N	14877	14877	14877
Mean of dependent variable	1145	1145	1145

Each row presents estimates for a different outcome of interest. Regressions include controls for child age and child fixed effects. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.

$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table 6: First difference estimates: 2nd parity

Panel A: Parental Investment		(1)	(2)	(3)
Sibling arrived	HOME Inventory	-1.866		-2.014
		(1.171)		(1.459)
Pregnancy occurred			-0.814	0.315
			(1.244)	(1.554)
N		10247	10247	10247
Panel B: Ability Measures		(1)	(2)	(3)
Sibling arrived	Cognitive Ability	-2.333**		-2.615*
		(1.174)		(1.434)
Pregnancy occurred			-1.144	0.562
			(1.259)	(1.560)
N		7339	7339	7339
Sibling arrived	Behavioral Problems	4.706***		8.017***
		(1.402)		(1.849)
Pregnancy occurred			-1.590	-6.734**
			(1.606)	(2.109)
N		8274	8274	8274

Each row presents estimates for a different outcome of interest. Controls are child's age at test and child fixed effects. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.

*: significant at 10% level. **: significant at 5% level. ***: significant at 1% level.

Table 7: Controlling for endogeneity of timing through smaller sampling windows around births: 2nd parity

Panel A: Parental Investment			
	0-12 months	0-24 months	0-36 months
HOME Index	-2.928**	-2.227*	-1.896
	(1.481)	(1.199)	(1.171)
N	1426	1973	2355
Panel B: Ability Measures			
	0-12 months	0-24 months	0-36 months
Cognitive Ability	-1.887	-0.537	-0.989
	(2.296)	(1.872)	(1.789)
N	438	567	675
Behavioral Problems	8.964***	5.688**	5.413**
	(2.433)	(2.058)	(2.018)
N	639	813	963

Each row presents estimates for a different outcome of interest. Column (1) restricts sample to interviews taken 0-12 months before and 0-12 months after the birth of a younger sibling. Column (2) restricts sample to interviews taken 0-24 months before and 0-12 months after the birth of a younger sibling. Column (3) restricts sample to interviews taken 0-36 months before and 0-12 months after the birth of a younger sibling. Regressions include controls for child age, gender, mother's age at first birth, AFQT, race, Rosenberg and Rotter scores. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010. $*$: significant at 10% level. $**$: significant at 5% level. $* * *$: significant at 1% level.
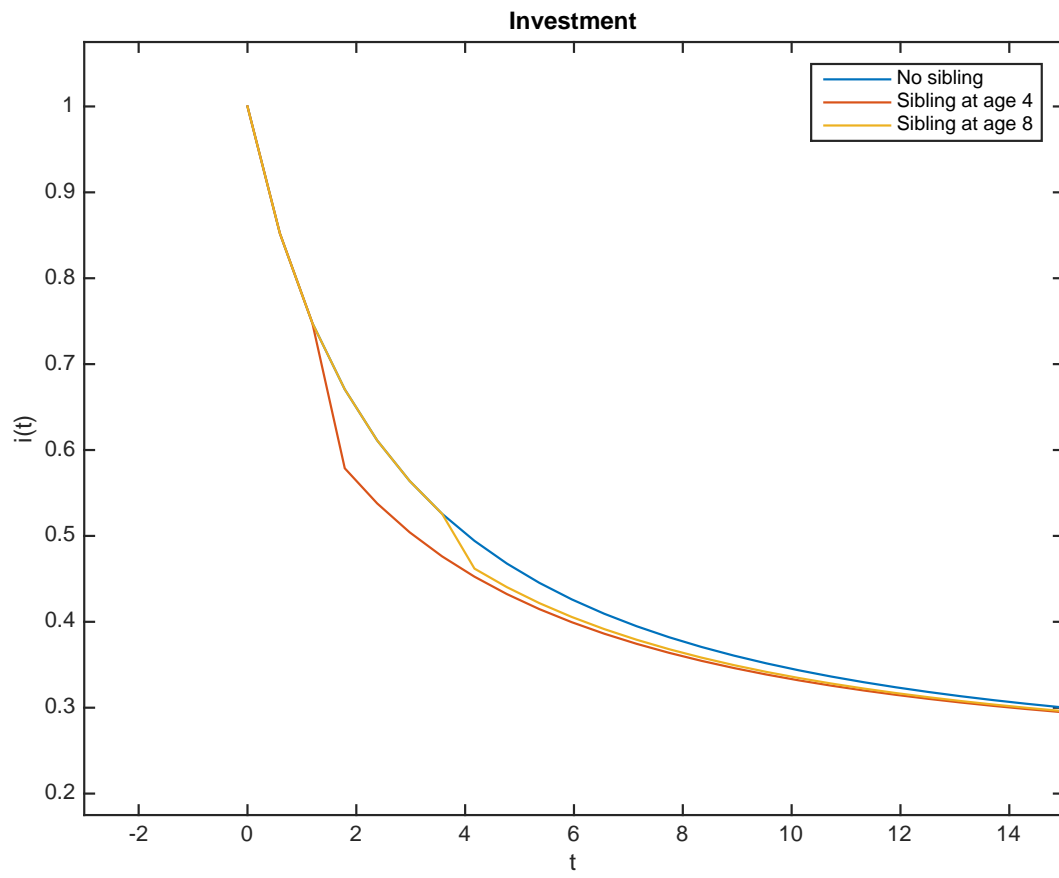
Figure 1: Time path of investment in older child

Figure 2: Time path of quality of older child
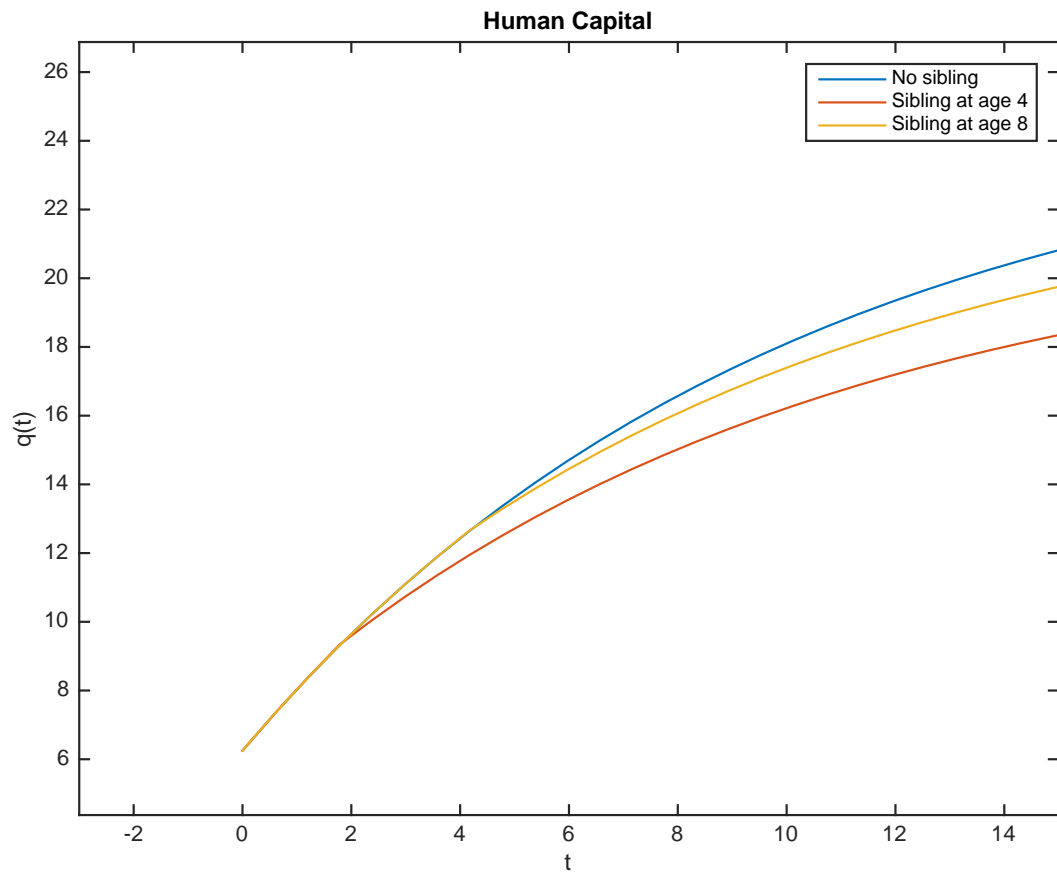
Figure 3: Effect of an additional sibling on firstborns before and after 2nd sibling birth

(a) HOME Inventory

(b) Cognitive Ability

(c) Behavioral Problems
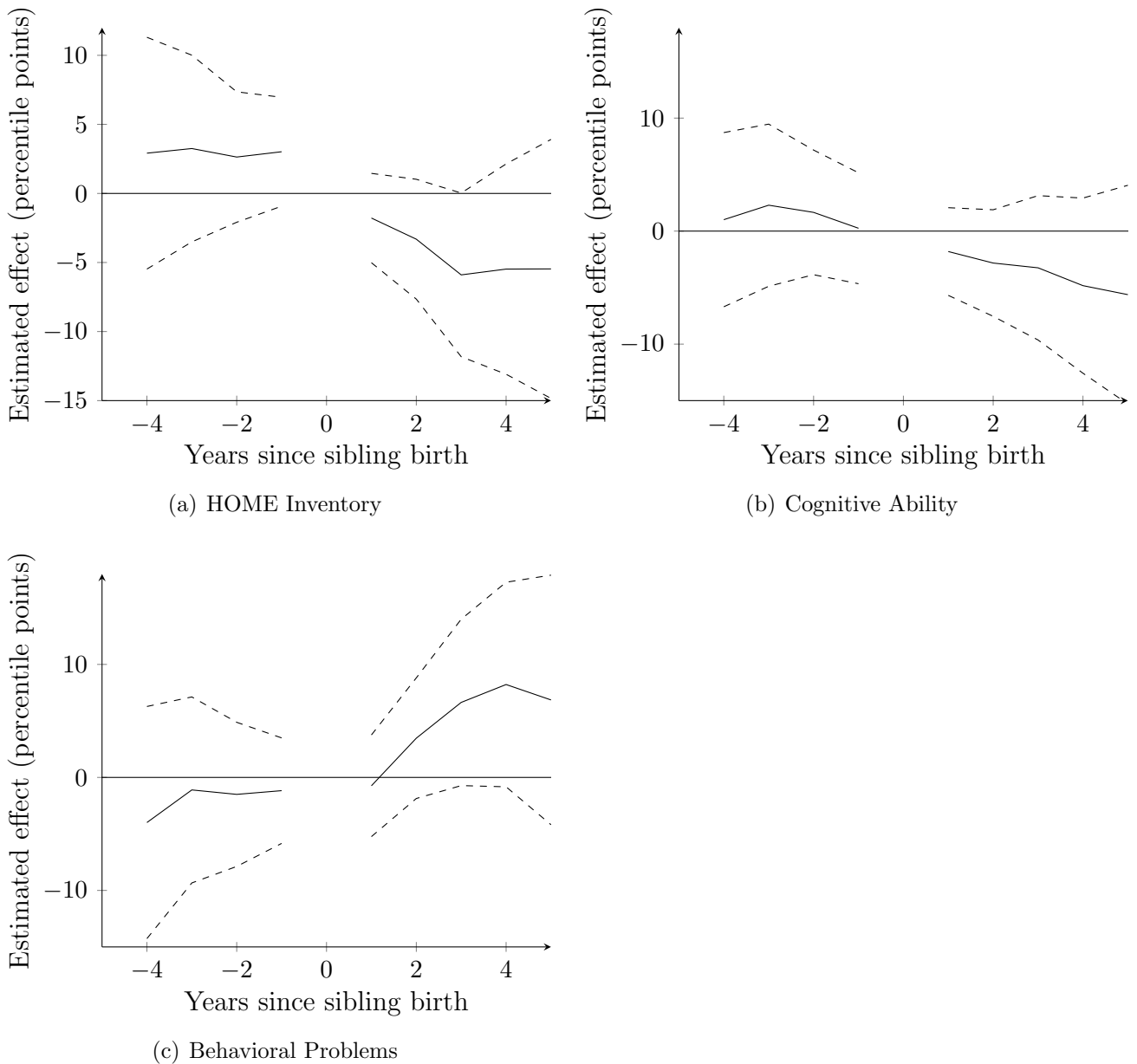
Each panel presents estimates of the effect of an additional child on an older sibling's outcome for the years preceding and following the year of birth. The omitted category is the year of birth. Regressions include controls for child age and child fixed effects. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. The dashed lines are 95% confidence intervals for robust standard errors clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.

Figure 4: Estimated impact of an additional sibling for different window lengths

(a) HOME Inventory

(b) Cognitive Ability

(c) Behavioral Problems
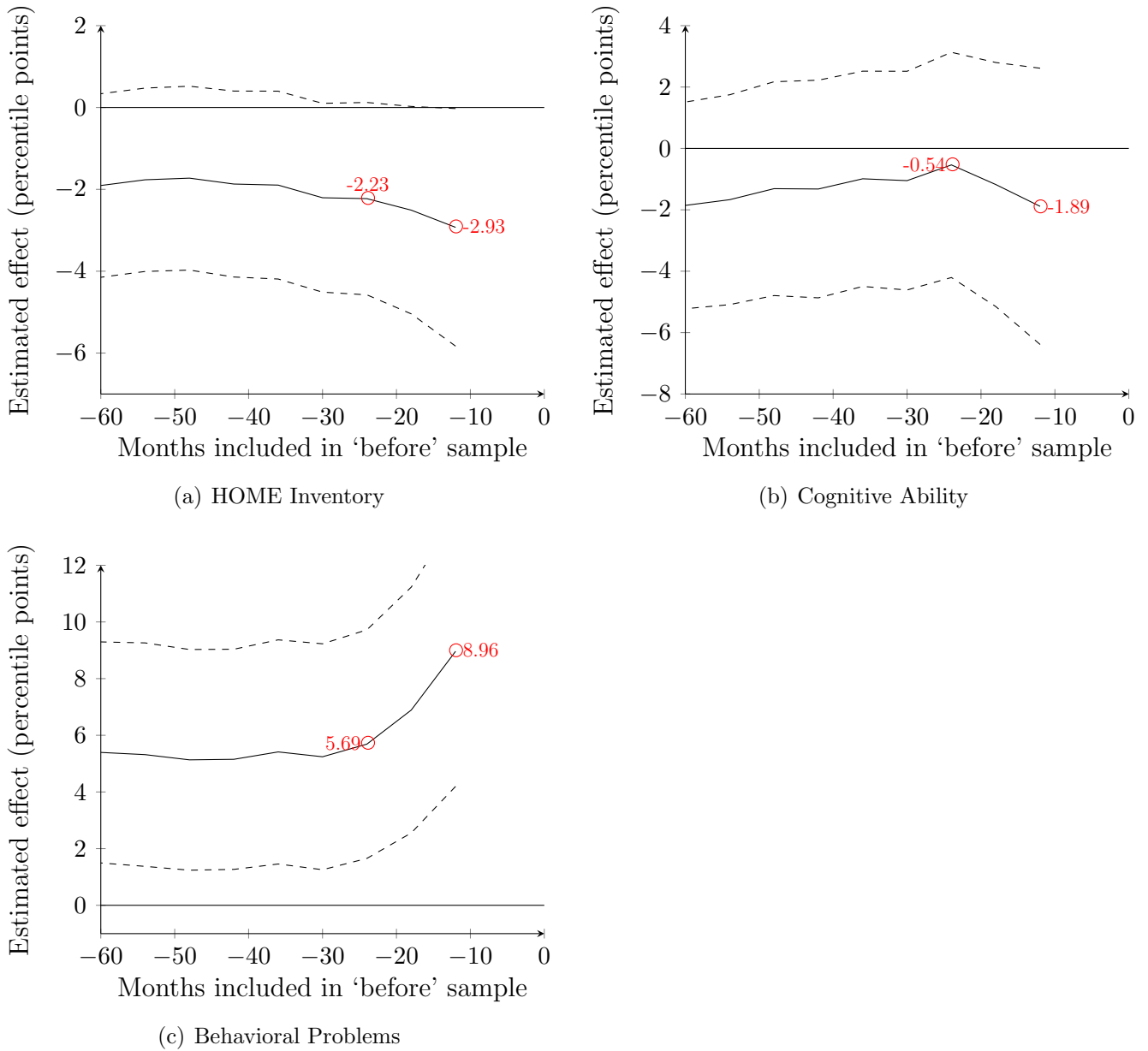
Each panel shows how the estimated effect of an additional sibling changes as we increase the sample to include more months prior to a sibling birth. The circled estimates correspond to the point estimates presented in Table 7. Regressions include controls for child age, gender, mother's age at first birth, AFQT, race, Rosenberg and Rotter scores. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.

A Tables

Table A.1: Children of NLSY79: Summary Statistics

	Mean	Std. Dev.	# of Obs.
Age in 2010	23.5	4.89	6084
At least 25 years old in 2010	0.43	0.50	6084
Female	0.49	0.50	6084
Black	0.15	0.36	6084
Hispanic	0.076	0.26	6084
Birth weight of child (oz)	118.7	21.0	5619
Family size	2.85	1.27	6084
Twins in family	0.037	0.19	6084
Yrs of education — At least 25 years old	13.1	2.43	2738
HS grad — At least 25 years old	0.80	0.40	2738
2009 Wage + salary income	25128.8	23304.8	2477
Log 2009 wage + salary income	9.97	0.98	2002
Worked 35+ hours at 1st job, 2010	0.77	0.42	2396
Child was convicted of a crime	0.26	0.44	2736
Child had a teenage pregnancy indicator	0.098	0.30	2740
Cognitive ability	59.8	24.4	33426
HOME Inventory score (percentile)	52.0	28.5	49680
Behavioral Problems Index (percentile)	57.9	28.1	38334

Data from Children of the NLSY79, 1986-2010. The data is based on 4,925 mothers and 11,464 children. For the first panel, each child is a unit of observation. For the second panel, each child is an observation, conditional on being at least 25 years old during at least one survey wave. For the third panel, a child-year is the unit of observation. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. Statistics calculated using NLSY sampling weights.

Table A.2: Summary statistics: Sibling-pair data

	Full Sample	Restricted Sample	Difference
Mother's AFQT pctile	37.12	37.81	0.69
	(27.02)	(27.01)	
Mother's self-esteem index	21.65	21.68	0.03
	(4.104)	(4.132)	
Mother's locus-of-control index	8.935	8.949	0.014
	(2.508)	(2.519)	
Age of mother at birth of child	20.30	20.75	0.45*
	(2.969)	(2.803)	
Black	0.204	0.194	0.10
	(0.403)	(0.396)	
Hispanic	0.104	0.108	0.04
	(0.305)	(0.311)	
Birth order	1.405	1.424	0.019
	(0.722)	(0.725)	
Birth weight of child (oz)	116.4	116.5	0.12
	(20.44)	(20.16)	
Spacing to next birth (months)	4.129	6.163	2.034***
	(3.253)	(3.425)	
Current number of siblings	2.576	2.227	-0.349
	(1.274)	(1.120)	
Living in blended family	0.0961	0.0776	-0.0185*
	(0.295)	(0.268)	
Living with single mother	0.315	0.349	0.034
	(0.465)	(0.477)	
Observations	5742	2469	

Data from Children of the NLSY79, 1986-2010. "Restricted sample" includes older children-younger siblings pairs where the older child has a valid PIAT test score. PIAT tests were administered to children aged 5 to 14. Standard deviations in parentheses. To test for significant differences between the restricted and full sample for each variable we regress the variable on a dummy indicating restricted sample status.
∗: significant at 10% level. ∗∗: significant at 5% level. ∗ ∗ ∗: significant at 1% level.

Table A.3: Individual fixed effects: 3rd parity

Panel A: Parental Investment

		(1) OLS	(2) Individual FEs	(3) Individual FEs
After sibling birth	HOME Inventory	-2.604** (0.958)	-1.850** (0.846)	
0-3 years after				-1.835** (0.845)
3+ years after				-2.985** (1.153)
N		13966	13966	13966

Panel B: Ability Measures

		(1) OLS	(2) Individual FEs	(3) Individual FEs
After sibling birth	Cognitive Ability	-0.660 (1.004)	-0.634 (0.776)	
0-3 years after				-0.658 (0.775)
3+ years after				-0.245 (0.998)
N		10374	10374	10374
After sibling birth	Behavioral Problems	0.229 (1.194)	2.351** (0.907)	
0-3 years after				2.334** (0.906)
3+ years after				2.868** (1.197)
N		11508	11508	11508

Each row presents estimates for a different outcome of interest. Controls are child's age at test and child fixed effects. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.

$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.

Table A.4: HOME questions used to form subcomponent indices

Subcomponent Index	Question Text
HOME, time with child	"How often do you read stories to child?" "How often has a family member taken or arranged to take child to any type of museum?" "How often does child eat a meal with both mother and father?"
HOME, resources spent	"About how many books does child have?" "About how many magazines does your family get regularly?" (Ages 3-5) "Does your family get a daily newspaper?" (Ages 6-14) "Does child have the use of a CD player, tape deck, or tape recorder, or record player at home and at least 5 children's records or tapes?" (Ages 3-5) "Is there a musical instrument that child can use here at home?" (Ages 6-14)
HOME, parental affection	"How many times in the past week have you shown child physical affection?" "Mother's voice conveyed positive feeling about child?" "Mother conversed with child excluding scolding or suspicious comments?"
HOME, home safety	"Child's play environment is safe?" "Interior of the home is dark or perceptually monotonous?" "All visible rooms of house/apartment are reasonably clean?" "All visible rooms of house/apartment are minimally cluttered?"

Table A.5: Comparing the pre-trend characteristics of the 12 month sampling window, 2nd parity births

	Before group	After sample	Difference
Mothers AFQT Score (percentile)	44.8	44.0	0.779
SELF ESTEEM SCORE 80	22.2	21.9	0.232
Mother's Rotter (locus-of-control) score (percentile)	8.84	8.88	-0.0341
Age of mother at birth of child	24.4	24.7	-0.276
Black	0.27	0.25	0.0150
Hispanic	0.19	0.20	-0.0139
HGC BY RS MOTHER 79	11.0	10.9	0.0170
Worked last week	0.30	0.32	-0.0166
Birth weight of child (oz)	114.9	116.1	-1.145
	Before group	After sample	Difference
HOMEP, previous wave	-2.08	0.018	-2.094
COG, previous wave	2.40	2.11	0.292
BPIP, previous wave	2.34	5.65	-3.314
	Before group	After sample	Difference
Survey Year	1989.8	1989.8	0.0126
Age	3.88	3.48	0.400*
Difference in age between respondent and 2nd birth	8.15	6.76	1.392***
Female	0.48	0.51	-0.0286

Panel (a) shows child characteristics for children who will be interviewed either one year before or one year after a younger sibling's birth and Panel (b) shows similar characteristics for two years. Data from Children of the NLSY79 sample, 1986-2010.
∗: significant at 10% level. ∗∗: significant at 5% level. ∗ ∗ ∗: significant at 1% level.

B Twins as an Instrument for Family Size

B.1 Twins in the NLSY mother-child data

Table B.1 shows the distribution of households of various family sizes as well as the incidence and birth of twins in the mother-child data. Using twins as an IV for family size is identified off individuals who have twins as younger siblings. For example, there are 35 families where twins are the second and third children born in the family, implying that there are 35 older children in these families who are in our sample. Likewise, there are 19 families where twins are third and fourth children born implying there are 38 older children added to our sample from these families. Altogether, the table indicates that there are 142 of these individuals in our data. The small number of older children residing in twin households is a concern. It is worth investigating these twin households in more detail to ensure that there are not substantial observable differences between households with and without twins.

Table B.2 shows demographic information for mothers in our sample split by the presence of twins in the household. There is not a significant difference in mothers that have twins and those that do not. If anything, mothers without twins have slightly lower cognitive and non-cognitive measures.

The two bottom panels of Table B.2 provides suggestive evidence for the quantity-quality tradeoff. Although maternal characteristics do not differ across households with and without twins there do appear to be substantial differences across children in these households. Children born in households with twins have worse test scores at all ages, lower parental investments, and more behavioral problems.

B.2 OLS and IV estimates of the effect of family size

Our OLS and instrumental variable estimates directly follows Black et al. (2010). We estimate the effects of family size on outcomes as described by the following equations:

$$Y = \beta_0 + \beta_1 FAMSIZE + X\beta_2 + \varepsilon \tag{8}$$

$$FAMSIZE = \alpha_0 + \alpha_1 TWIN + X\alpha_2 + \nu \tag{9}$$

where Y is an outcome of interest, eg. years of education, $FAMSIZE$ is the number of children in the family, and X is vector of individual characteristics. $TWIN$ is a dummy variable indicating whether the family has any twins. For households with twins present, we restrict the sample to children who were born prior to the twin birth. Families with twin births have had a plausibly exogenous increase in family size. Differences in outcomes for older children in households with twins versus households without twins can thus be interpreted as a causal estimate of how family size affects outcomes.[14]

Table B.3 presents both OLS and IV estimates of the impact of family size on years of education.[15] As reported in column 1, the OLS estimate is large and highly significant. The coefficient implies that raising final family size by an additional child reduces average schooling of children by -.13 years. Since we add additional maternal controls, the coefficient is slightly smaller than the -0.18 reported in Black et al. (2005) and -.20 reported in Blake (1989) using U.S. data. Similar to Black et al. (2005), we also find that the addition of

---

[14]Readers may be curious why we do not use the sibling-sex composition of the first two children as an additional instrument, a method used by Angrist et al. (2010). Unfortunately, in the NLSY matched mother-children data this instrument is weak - the F-test in the first stage is only 5.83.

[15]In order to ensure the highest grade completed in our sample represents completed education we restrict our sample to children who were surveyed at least once after age 24. We include demographic controls for the child such as age, gender, and race and maternal controls such as mother's race, mother's AFQT, mother's age at first birth, and mother's non-cognitive measures such as Rosenberg score and Rotter scale score.

birth order controls in Column 2 reduces the size and significance of the family size effect considerably.

Column 4 reports estimates of family size using younger twin birth as an instrument. In contrast to Black et al. (2005), our IV estimates are negative and large. Older children in families with twins have education levels that are four-tenths of a year lower compared to children in families without twins. A similar pattern is found for other adult outcomes. As shown in Table B.4, IV estimates indicate that larger families not only have children with lower educational attainment, but also lower earnings, and increased likelihood of criminal behavior and teenage pregnancies.

Column 4 shows that the first stage coefficient for the IV is 1.81, which suggests that having a twin birth increases the final family size by more than one child. One potential concern is that families with higher desired fertility are more likely to experience a twin birth. Black et al. (2005) address this potential selection problem by estimating the family size effect by parity. Column 5 reports the results for second parity only. While the coefficient is still large and negative, we lose considerable precision given the small number of twin births in our data.[16]

Education has been the focus of most previous studies on the quantity-quality trade-off. While education is the most common measure of child quality, it is likely to be a function of institutions as well as parental investments. For example, the lack of family size effects in Black et al. (2005) may reflect the existence of a generous public welfare system and a

---

[16]We have also experimented with the parity-pooled approach suggested by Angrist et al. (2010). Using this method, we combined twin births at different parities by regressing the presence of a twin birth for older children at a given birth parity on other covariates, calculating the residual from this regression, interacting this residual with a dummy indicating whether a child was an older sibling at that birth parity, and using this interacted residual as the instrumental variable for an increase in family size. See section IV.C in Angrist et al. (2010) for technical details on this estimation strategy. Using this approach our coefficient on family size is even larger at -0.833 and significant at the 5 percent level.

strong public education system in Norway where at the margin, parental investments may not result in variations in educational outcomes. An important advantage of the NLSY is that it provides a variety of child cognitive and non-cognitive skill measures as well as direct measures of parental investments that affect the home environment. Thus, we can more directly test the original quantity-quality theory which is about parental choices and child skill formation.[17]

Table B.5 presents estimates of how family size impacts outcomes during childhood. The estimating equations are modified versions of equations 1 and 2 with t subscripts:

$$Y_{it} = \beta_0 + \beta_1 FAMSIZE_{it} + X_{it}\beta_2 + \varepsilon_{it} \tag{10}$$

$$FAMSIZE_{it} = \alpha_0 + \alpha_1 TWIN_{it} + X_{it}\alpha_2 + \nu_{it} \tag{11}$$

where $FAMSIZE_{it}$ refers to current family size and $TWIN_{it}$ refers to an indicator variable whether there are twins currently in the household. We report estimates clustering observations at the child level.[18]

According to the IV estimates reported in Column 2, an additional sibling reduces the HOME score - the NLSY's measure of parental investment in a child - by 1.7 percentile points. This finding speaks directly to the trade-offs in a quantity-quality model of child-rearing. Parents in larger families reduce their per-child investment. Similarly, an extra sibling reduces an older sibling's cognitive score by 2.6 percentile points. Since the standard

---

[17]Several previous papers have moved in this direction by using alternative child quality measures. (Hanushek 1992) finds a strong effect of family size on reading comprehension and vocabulary test scores using data from the Gary Income Maintenance Experiment merged with school level data. Interestingly, Black et al. (2010) find a negative family size effect on child IQ in the Norwegian data even though they did not find any effect on education. Caceres-Delpiano (2006) and Conley and Glauber (2006) find that children in larger families are less likely to attend private school and more likely to repeat a grade.

[18]We have also clustered at the mother level with similar results.

deviation of cognitive scores is 24 percentile points, this translates into a slightly larger than a tenth of a standard deviation reduction. We do not find significant effects of family size on the Behavioral Problem Index (BPI), however.

Table B.1: Distribution of family size and twin births

| | Family size | Birth order of twins within family |
	# of HHs	# of HHs
1	1179	44
2	1966	35
3	1126	19
4	430	10
5	135	6
6	49	3
7	22	
8	7	
9	4	
10	2	
11	2	
Total	4922	117

Data from Children of the NLSY79, 1986-2010.

Table B.2: Maternal characteristics by twin and non-twin households

	HH w/o twins	HH w/ twins	Difference
Mothers AFQT Score (percentile)	40.2	44.0	-3.779
Mother's self-esteem index	22.1	22.6	-0.554
Mother's locus-of-control index	8.82	8.74	0.0745
Age of mother at birth of child	22.9	23.3	-0.436
Black	0.26	0.26	0.00234
Hispanic	0.17	0.17	0.0000758
	HH w/o twins	HH w/ twins	Difference
Cognitive ability	55.0	48.0	7.007***
HOME Inventory score (percentile)	46.8	42.8	3.937**
Behavioral Problems Index (percentile)	58.9	61.3	-2.347
Birth weight of child (oz)	116.4	116.7	-0.286
	HH w/o twins	HH w/ twins	Difference
Yrs of education — older than 22	12.7	11.9	0.810**
HS grad — older than 22	0.75	0.57	0.175***
Log 2009 wage + salary income	9.92	9.61	0.308
Teen pregnancy	0.076	0.22	-0.147***

Data from Children of the NLSY79 sample, 1986-2010. For the first panel, the mother is the unit of observation. For the second panel, a child-year is the unit of observation. For the third panel, a child (surveyed after age 25) is the unit of observation. For households with twins present we restrict the sample to children born prior to the birth of the twins. The last column reports p-values of an unpaired t-test for difference in means between households with and without twins.

Table B.3: Effect of an additional child on years of education

	(1) Maternal controls	(2) + Birth order controls	(3) OLS, — IV sample	(4) IV, full sample	(5) IV, second parity births
Family Size	-0.131**	-0.0287	-0.0613	-0.395***	-0.531
	(0.0519)	(0.0531)	(0.0595)	(0.143)	(0.586)
2nd Child		-0.600***	-0.620***	-0.506***	
		(0.123)	(0.126)	(0.141)	
3rd Child		-0.664**	-0.622**	-0.203	
		(0.265)	(0.273)	(0.337)	
4th Child		-1.441***	-1.332***	-0.601	
		(0.299)	(0.311)	(0.445)	
5th Child		-1.911**	-1.978**	-0.958	
		(0.752)	(0.799)	(0.963)	
6th Child		-2.038	0.312	1.661	
		(1.481)	(1.264)	(1.387)	
7th+ Child		-2.574***			
		(0.332)			
N	7931	7931	7316	7316	4320
r2	0.139	0.155	0.155	0.143	0.106
				First Stage Coeff.	First Stage Coeff.
				1.805***	0.809***
				(0.148)	(0.177)

All results include controls for child age and gender. Maternal controls are mother's age at first birth, race, AFQT, Rosenberg and Rotter scores. Robust standard errors are clustered at the child level. Data from Children of the NLSY79 sample, 1986-2010. The sample is restricted to children who are surveyed after age 25. The sample size changes between Columns (2) and (3) as children born after the birth of twins are dropped from the overall sample. The sample size changes between Columns (4) and (5) as we restrict the sample to only second-parity births.
∗: significant at 10% level. ∗∗: significant at 5% level. ∗∗∗: significant at 1% level.

53

Table B.4: Long-term outcomes

	(1) OLS	(2) IV, full sample	(3) IV, second parity births
Years of education	-0.0287	-0.395**	-0.531
	(0.0531)	(0.143)	(0.586)
N	7931	7316	4320
Graduated HS	-0.0204**	-0.126***	-0.00929
	(0.00967)	(0.0376)	(0.106)
N	7931	7316	4320
Log earnings	-0.0289	-0.184**	-0.104
	(0.0310)	(0.0598)	(0.136)
N	4098	3732	2091
Criminal conviction	0.0169*	0.0746*	0.268*
	(0.00969)	(0.0399)	(0.157)
N	7931	7316	4320
Teen pregnancy	0.0129*	0.0936**	-0.0259
	(0.00731)	(0.0311)	(0.0549)
Birth order controls	Yes	Yes	No
N	7931	7316	4320

Each row presents estimates for a different outcome of interest. All results include controls for child age, gender, mother's age at first birth, AFQT, race, Rosenberg and Rotter scores. For years of education, high school graduation, and log earnings the sample is restricted to children who are surveyed after age 25. For teen pregnancy and criminal conviction the sample is restricted to children who are surveyed after age 20. Robust standard errors are clustered at the individual level. Data from Children of the NLSY79 sample, 1986-2010.
∗: significant at 10% level. ∗∗: significant at 5% level. ∗ ∗ ∗: significant at 1% level.

Table B.5: Childhood outcomes

Panel A: Parental Investment

	(1) OLS, — IV sample	(2) IV, full sample	(3) IV, second parity births
HOME Index	-1.127***	-1.764	-1.453
	(0.328)	(1.119)	(2.033)
N	40529	40529	14241

Panel B: Ability Measures

	(1) OLS, — IV sample	(2) IV, full sample	(3) IV, second parity births
Cognitive Ability	-0.909**	-2.647**	-1.464
	(0.309)	(1.047)	(1.963)
N	28108	28108	10664
Behavioral Problems	-0.912**	0.0405	-1.975
	(0.379)	(1.584)	(3.050)
N	31984	31984	11916
Birth order controls	Yes	Yes	No

Each row presents estimates for a different outcome of interest. Cognitive ability is constructed by averaging a child's PIAT math and reading recognition test scores. All results include controls for child age, gender, mother's age at first birth, AFQT, race, Rosenberg and Rotter scores. Robust standard errors are clustered at the child level. Data from Children of the NLSY79 sample, 1986-2010. The sample size changes between Columns (2) and (3) as we restrict the sample to only second-parity births.

$*$: significant at 10% level. $**$: significant at 5% level. $***$: significant at 1% level.