The Effect of Maternal Labor Supply on Children: Evidence from Bunching*

Carolina Caetano[1], Gregorio Caetano[1], Eric Nielsen[2], and Viviane Sanfelice[3]

[1]University of Georgia
[2]Federal Reserve Board
[3]Temple University

December 2021

Abstract

We study the effect of maternal labor supply in the first three years of life on early childhood cognitive skills. We pay particular attention to heterogeneous effects by the skill of the mother, by the intensity of her labor supply, and by her pre-birth wages. We correct for selection using a control function approach which uses the fact that many mothers are bunched at zero working hours – skill variation in the children of these bunched mothers is informative about the effect of unobservables on skills. We find that maternal labor supply typically has a significant, negative effect on children's early cognitive skills, with more negative effects for higher-skill mothers. By contrast, we do not find significant heterogeneity depending on the pre-birth wage rate of the mother. These findings suggest that there may be more scope to avoid short-term, unintended consequences of maternal labor supply through policies that promote more flexible work arrangements rather than through policies that increase the financial rewards to working.
JEL Codes: D13, I21, I2, J01, J22, C24. Keywords: cognitive skills, bunching, maternal labor supply, early childhood, skill development.

1 Introduction

This paper estimates the effect of mothers working longer hours during the first three years of a child's life on that child's cognitive skills around age 6. We use data on maternal work histories in the National Longitudinal Surveys of Youth 1979 (NLSY79) linked to childhood skill measures from the Children of the National Longitudinal Surveys (CNLSY), focusing on mothers whose children were born between 1979 and 2008. We aim to understand an important aspect of the trade-off mothers may face when deciding whether and how much to work. On the one hand, maternal labor supply may be detrimental to children's skills because time spent at work is time not spent with children. Indeed, there is a wealth of evidence suggesting that an enriching environment with high-quality parent/child interactions in early childhood is important for subsequent skill development

*We would like to thank Joseph Altonji, Joseph Hotz, Josh Kinsler, Rodrigo Pinto, David Slichter, Christopher Taber, Hao Teng, Christopher Ruhm, and seminar participants at various institutions. The analysis and conclusions set forth here are those of the authors and do not indicate concurrence by other members of the research staff, the Board of Governors, or the Federal Reserve System.

(e.g., Todd and Wolpin 2007, Del Boca, Flinn, and Wiswall 2014, Hsin and Felfe 2014, Bono, Francesconi, Kelly, and Sacker 2016). On the other hand, additional work hours will bring in additional income, which may itself have a direct, positive impact on skills (Blau 1999, Milligan and Stabile 2011, Dahl and Lochner 2012, Løken, Mogstad, and Wiswall 2012).

The trade-off between time working and time at home has become increasingly salient as maternal labor supply has increased in recent decades (Eckstein and Lifshitz 2011, Fogli and Veldkamp 2011). Understanding the sign and magnitude of this effect is critical both for understanding the sources of childhood skill differences and as an input into many policy-relevant analyses. For instance, many public policies – including child allowances and tax credits, subsidized child care, or even the progressivity of the tax code – will alter mothers' labor supply choices.[1] Such policies may have unintended consequences on skill development, with important implications for intergenerational mobility and inequality in general (Blau and Currie 2006, Currie and Almond 2011).

This paper takes a distinct approach in focusing on heterogeneous effects by the skills of mothers and by the quantity of their labor supply. Both of these dimensions should alter the intensity of the trade-off between maternal labor supply and time at home. More skilled mothers tend to earn higher wages, but their time not working may also be more valuable (in terms of skill production) to their children. It is unclear whether the additional resources (financial or otherwise) earned by skilled working mothers can better offset any detrimental effect of working.[2] Moreover, on the margin, this trade-off is likely to change depending on whether the mother works longer hours (Ettinger, Riley, and Price 2018). These two dimensions may interact, as higher-skilled mothers tend to work longer hours (Cortes and Tessada 2011, Adda, Dustmann, and Stevens 2017, Chen, Grove, and Hussey 2017).

Estimating these effects is challenging because maternal labor supply may be correlated with unobservables that are themselves inputs in childhood skill production. Prior research has addressed this endogeneity using standard approaches including family fixed effects and instrumental variables (IVs). We discuss this prior related work in greater detail in Section 2. Methodologically, we add to the literature by using a novel control function approach that does not require IVs, leveraging instead the fact that maternal labor supply is bunched at zero (Caetano, Caetano, and Nielsen 2021). We argue that mothers bunch at zero because they are at a corner solution: the constraint that hours worked cannot be negative is binding for them. Bunched mothers have different levels of the unobservable confounder, but they all choose the same amount of working hours (namely,

[1] The literature on tax credits effects on female labor supply is large; for some references, see work by Eissa and Liebman 1996, Averett, Peters, and Waldman 1997, Meyer and Rosenbaum 2001, Grogger 2003, Bosch and Van der Klaauw 2012, Blundell, Costa Dias, Meghir, and Shaw 2016, Bick and Fuchs-Schündeln 2017. For work on child care effects on maternal labor supply at early ages (less than 3 years old) see Baker, Gruber, and Milligan 2008, Goux and Maurin 2010, Givord and Marbot 2015, Carta and Rizzica 2018, Yamaguchi, Asai, and Kambayashi 2018, Gathmann and Sass 2018, and Andresen and Havnes 2019.

[2] On the one hand, their spouse or others in her network might be more available to the child (Kalenkoski, Ribar, and Stratton, 2009; Sayer and Gornick, 2012), they may be able to afford higher-quality childcare (Blau and Hagy, 1998), or they may be better able to substitute market-purchased goods for their own time (Anderson and Levine, 1999). On the other hand, the time of a higher-skilled mother may be less substitutable (from the perspective of the child) with any of these options (Ruhm, 2009; Carneiro, Meghir, and Parey, 2013; Polachek, Das, and Thamma-Apiroam, 2015).

zero). This allows us to isolate the effect of confounders on skills (i.e. separately from the effect of working hours on skills), and thus to build a control function to control for confounders. Instead of exclusion restrictions, our approach relies on a testable distributional assumption.

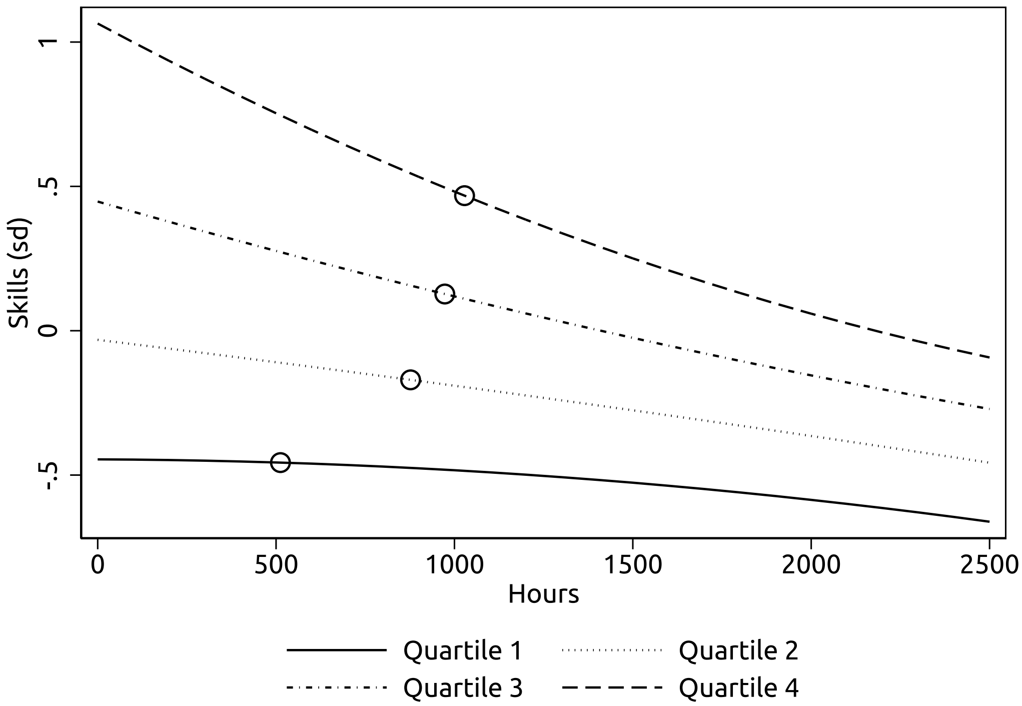
We believe this control function approach adds value to the understanding of the effects of maternal labor supply because it offers a sensible identification strategy that allows us to use the full NLSY79/CNLSY sample – we do not need to restrict our analysis to families with siblings (as in fixed effects models) or to families for whom a particular instrumental variable is available.[3] This enables us to uncover effects broken down along these important dimensions of heterogeneity, which enrich our understanding of the relationship between maternal labor supply and child development. We provide many robustness checks that speak to the validity of the identification strategy in our empirical context.

We find that maternal labor supply has, on average, negative effects on children's cognitive skills in the short-run. Our estimates imply that an additional 10 hours of maternal labor per week during the first three years of a child's life lowers the child's cognitive skills at age six by about 10% of a standard deviation (s.d.). We also find substantial heterogeneity in these effects by the mother's skill, measured by the Armed Forces Qualifying Test (AFQT) score, and some heterogeneity by the total number of hours worked. These heterogeneous results are shown in Figure 1. The hollow circles show the average observed childhood cognitive skills and post-birth maternal work hours for each quartile of the maternal AFQT distribution. The lines show the counterfactual skills of children observed at each of these hollow points if their mother worked a different number of hours above or below the average value in the data. First, we find evidence that the schedule is nonlinear, with the degree of concavity/convexity changing depending on the skill level of the mother. However, although statistically significant, this curvature is economically not very important – the linearity assumption that is typically made in this literature seems to be a good approximation for the range of hours and AFQT scores observed in the data. Second, we find substantial heterogeneity depending on the skill of the mother. Labor supplied by higher-skilled mothers tends to have more negative effects, while for lower-skilled mothers the effects are closer to zero.

Of course, maternal labor supply has many other positive effects which may justify the implementation of work-promoting policies. Specifically, the additional income the mother earns may be beneficial to the child in the long run via several channels – better schools and social networks, support for college admissions, reduced levels of stress, etc. Furthermore, the additional income will generally affect all family members, including the mother herself, in various positive ways. There are also important considerations with respect to career timing. Although a mother may have liked to cut back her hours during her child's early years, she might prefer to remain in the labor force full time, even if doing so is detrimental to the child in the short-run, because of potential negative long-term effects on her career. Better career growth, and the higher resources that come with

[3] Another potential concern with understanding heterogeneous effects using IVs is that the "compliers" of a given IV may disproportionately have a certain skill level, making it difficult to compare the estimates across skill levels in a meaningful way. By contrast, we demonstrate empirically that mothers of all skill levels are well-represented in the bunched group at zero work hours.

Figure 1: Children's Cognitive Skills – Quartiles of Maternal AFQT



Note: Total effects based on estimates from Table 3. The hollow circles represent the average skills and working hours of all observations in the corresponding quartile of the maternal AFQT distribution.

it, could in turn benefit all family members, including the child, in the long run. Thus, it would be valuable to investigate further whether there is room to conduct work promoting-policies while mitigating the scope for this negative unintended consequence for higher-skilled mothers.

With these considerations in mind, we investigate why working longer hours seems to have particularly detrimental short-run effects on the cognitive skills of the children of higher-skilled mothers. A natural candidate explanation is that the last hour worked may be particularly costly for the children of mothers working longer hours, who are also disproportionately higher skilled. This potential explanation is ruled out by our finding that the effect is approximately linear, as shown in Figure 1. In fact, to the extent that we find some evidence of nonlinearity, Figure 1 shows that the last hour might be particularly costly for children of mothers working longer hours only for mothers who are sufficiently low-skilled.

Another potential explanation is that the additional money earned by these higher-skilled mothers may not be enough to offset their opportunity cost of working, at least in the short-run. It is therefore valuable to consider heterogeneity of the effects by another dimension beyond skills: the mother's wage rate. If the negative effects accrue mostly to the children of high-skilled, *low-wage* mothers, then policies that provide financial support to low-wage mothers could be effective by allowing their families to pay for goods and services to offset their absence. However, if the negative effects accrue also to the families of high-skilled, *high-wage* mothers, suggesting that close substitutes for high-skilled maternal time are unavailable even for this high-earning group, then providing financial support may be ineffective. In this case, it would make more sense to focus on policies aimed at promoting flexible location and work schedules for mothers, thus allowing mothers to maintain (or increase) their labor supply while not reducing their interactions with their children.

In order to investigate these issues, we study the heterogeneity of the effects of maternal work hours (beyond skill and number of hours) by the hourly pre-birth wage rate of the mother. Allowing for this third dimension of heterogeneity leads to more noisy estimates, as expected. However, this exercise is still informative. We find some evidence that mothers with higher pre-birth wages may be able to mitigate some of the detrimental effect of their absence, although the degree of mitigation appears to be modest at best. It appears that current institutions and norms do not provide sufficient scope for families with higher-skill working mothers to mitigate any detrimental effects of the mother working on the child's cognitive skills in the short-run, even if the mother is also a high-earner.

These findings provide important insights about the effects of maternal work during the first years of the children. First, using a different identification strategy, we confirm what has largely been found in prior literature (see Section 2): maternal labor supply on average has a negative short-run effect on children's cognitive skills. Second, we provide new evidence that this short-term unintended consequence tends to be small for low-skill mothers, even those who work long hours. Third, our analysis gives us a clue about how work-promoting policies could potentially avoid short-term unintended consequences for the children of higher-skill mothers as well. It may be safer (from the perspective of mitigating any unintended consequences for the child) to focus on policies that encourage greater flexibility in work arrangements for mothers, rather than focusing on policies that provide financial support for mothers to mitigate the consequences of their absence. Indeed, policies that increase flexibility in the schedule and location of jobs could allow mothers to spend more time with their children without sacrificing their work hours. Additionally, enhancing spouses' flexibility along the same lines might be complementary, further mitigating any potential cost to children.[4] Some of these policies, including remote and hybrid work, have recently expanded dramatically around the Covid-19 pandemic due to the necessity of social distancing, especially for higher skilled workers (Bartik, Cullen, Glaeser, Luca, and Stanton, 2020; Bick, Blandin, and Mertens, 2020; Dingel and Neiman, 2020). It would be valuable to understand the long-run impact of these changes.

The rest of this paper is organized as follows. Section 2 relates this paper to previous work in the literature, while Section 3 presents our data. Section 4 discusses our empirical approach. Section 5 presents our main empirical findings. Finally, Section 6 provides a concluding discussion. An appendix contains a number of robustness analyses.

[4]It is worth emphasizing that while our results pertain to the effects of maternal labor supply, similar considerations and concerns would in principle apply to the labor supply choices of any parent or caregiver. Our focus on mothers is purely pragmatic – data sources such as the NLSY79/CNLSY do not allow one to connect paternal labor supply to measures of childhood skills.

2 Related literature

The vast majority of studies on this topic use the NLSY79/CNLSY data and focus on estimating the impact of maternal hours worked during the three first years of a child's life on the child's skills at an early age, as we do. To overcome the endogeneity of maternal labor supply, these studies use either (i) a considerable set of control variables (Desai, Chase-Lansdale, and Michael, 1989; Baydar and Brooks-Gunn, 1991; Vandell and Ramanan, 1992; Parcel and Menaghan, 1994; Hill and O'Neill, 1994; Waldfogel, Han, and Brooks-Gunn, 2002; Baum II, 2003; Ruhm, 2004, 2009), (ii) local labor market conditions as an instrumental variable (Blau, Grossberg, et al., 1992; James-Burdumy, 2005), (iii) family fixed effects (Waldfogel et al., 2002; James-Burdumy, 2005), or (iv) dynamic choice models that simultaneously consider a mother's choice to work and invest in the child's cognitive skill (Bernal, 2008).

The results in these studies vary widely, making it difficult to draw a clear conclusion about the magnitude of the effect of maternal employment. Nonetheless, on balance, this literature finds that maternal labor supply has either a null or a detrimental effect on children's cognitive skills. Our results are consistent with these findings. For instance, Ruhm (2004), which adopts a selection-on-observables approach, finds that each additional twenty hours worked per week during the first three years of life is associated with a 0.11 standard deviation decrease on the reading assessment and a 0.08 standard deviation decrease in the mathematics assessment. Similarly, Bernal (2008) finds that working full-time and using childcare for one year is associated with a 0.13 standard deviation reduction in test scores. Other papers finding negative effects include Desai et al. (1989), Baydar and Brooks-Gunn (1991), Hill and O'Neill (1994), and Baum II (2003). Using fixed-effects models, James-Burdumy (2005) finds null effects in some cases and negative effects in others. Parcel and Menaghan (1994) similarly find null effects, while Blau et al. (1992) and Waldfogel et al. (2002) find negative effects in the first year of the child's life and offsetting, positive effects subsequently. Finally, Vandell and Ramanan (1992) reports positive effects of early maternal employment on math achievement for children from low-income families, which is consistent with our heterogeneous results.

Our analysis matches the context of this literature: we also focus on the impact of maternal labor supply in the first three years of the child on the child's early outcomes, and we also use the NLSY79/CNLSY data. Because of the similar context, we complement the main findings in this literature in many ways: (a) We confirm the main findings of negative effects with a different approach to control for confounders; (b) We confirm that the linearity assumption made in this literature is a good approximation for the range of hours and skills in the data; (c) We provide new results about heterogeneity by skills; (d) We investigate the direct vs. income-mediated channel of the effect, providing further context to the findings of this literature while shedding light on the potential impacts of different policies.

We are not the first study to investigate the direct vs. the income-mediated channel of maternal labor supply. Two recent papers investigate such effects, but in contexts different than those in the literature discussed above. Agostinelli and Sorrenti (2021) use the NLSY79/CNLSY to estimate time and income effects of maternal labor supply when children are 4-16 years old on the children's contemporaneous outcomes, instrumenting for maternal labor supply with local labor market conditions and for family income with Earned Income Tax Credit (EITC) expansions. They find negative direct effect of maternal hours worked and positive income effect that are not fully offsetting, as we do. Using Norwegian registry data, Nicoletti, Salvanes, and Tominey (2020) estimates the direct and income-mediated effects of maternal labor supply during the first five years of the child on test scores at ages 11 and 15. To handle the endogeneity of maternal work hours and family income, the authors construct instruments for each based on the characteristics of the peers of the parental peers. They find a negative direct effect of maternal labor supply on test scores and a positive income effect that fully offsets the negative direct effect.

3 Data

We use data from two linked surveys: the NLSY79, which gives us information about mothers, and the CNLSY, which gives us information about their children. The NLSY79 follows a cohort of young adults aged 14-22 from 1980 through the present, while the CNLSY follows the children born to the women in the NLSY79 sample.[5] Linked together, these surveys provide a unique source of information on children and their parents, including detailed information on maternal labor supply, childhood cognitive development, and household characteristics. Our final sample is a cross-sectional data set of children born from 1979 to 2008 for whom information on cognitive measures, maternal labor supply, and family characteristics are available. Children who were reported not to be living with their mother in the first years of life are dropped from our sample. We also drop observations who report working exactly 40 hours per week for all 52 weeks during each of the three years, as this lack of variation across weeks suggests that these reported hours do not reflect the actual working hours of the mother. However, replicating our analysis using these observations yields nearly the same results in all instances.

Following the economic literature in child development, we measure cognitive skills using the reading recognition and math tests from the Peabody Individual Achievement Test (PIAT). The reading recognition test is designed to measure reading comprehension based on a child's ability to recognize and pronounce words. The math test assesses attainment in mathematics beginning with early skills, such as recognizing numerals, and progressing to advanced concepts in geometry and trigonometry. The PIAT was administered to all children over the age of 5 in each CNLSY wave. Because our focus is on early childhood skill development, we adopt as our outcome a unified score for childhood cognitive skills constructed by applying factor analysis to the age-standardized math and reading PIAT scores from the first time each child in the CNLSY is assessed, which happens around age 6.[6] Throughout the analysis, we measure skills in standard-deviation (s.d.) units.

[5] The NLSY79 interviews are annual from 1979-1994 and biennial thereafter. The CNLSY interviews are biennial starting in 1986.

[6] These age-specific scores are based on a nationally representative sample of children and are normalized to have a mean of 100 and a standard deviation of 15.

We measure our primary variable of interest, maternal labor supply, using the average number of hours worked annually by the mother in the first three years of the child's life. The NLSY79 collects extensive weekly information on employment status and hours worked. This allows us to construct a weekly work history for each mother after giving birth. Some mothers may report that they are working shortly after giving birth when they are actually on paid maternity leave (Baum II, 2003). We can only distinguish these two possibilities – working after birth versus paid maternity leave – in the survey waves from 1988 onward. To avoid losing a large portion of our sample and yet to avoid measurement error due to maternity leave, we begin to measure hours worked in the fourth month following the month of birth.[7] For instance, for a child born in July, we compute hours worked by the mother starting in the first week of November. For this child, maternal labor supply in the first year of life would be computed from the first week of November of the year of birth until last week of October in the following year. We continue this yearly computation for the next two years in order to measure hours worked by the mother in the second and third year of the child's life. Finally, our treatment variable is computed by taking the average of annual number of hours worked by the mother in these three years.[8]

A key explanatory variable in this study is the mother's cognitive skill, which we measure using the Armed Forces Qualifying Test (AFQT). The AFQT was administered to almost all NLSY79 respondents in the base year of the survey. The AFQT is a general measure of achievement in math and reading and is a primary eligibility criterion for service and placement in the United States Armed Forces. Because of its use in U.S. military personnel decisions, the AFQT has undergone extensive vetting and has been used in numerous prior economic studies as a proxy for cognitive skill or human capital (Neal and Johnson, 1996; Hirsch and Schumacher, 1998; Arcidiacono, Bayer, and Hizmo, 2010).[9]

In addition to maternal AFQT, we construct a number of other control variables based on the child, mother, and household characteristics. Unless otherwise specified, control variables such as the mother's education and marital status are computed at the year of birth. We opt for this approach in order to keep our control variables pre-determined.[10]

Table 1 presents summary statistics of our sample. The table first shows the mean and standard deviation of each element used to generate the children cognitive skill measure. These variables are normalized by age and follow a nationally representative sample with a mean of 100 and standard deviation 15. On average, children in our sample score above the national average on the PIAT

[7]The findings in this paper do not change if we start counting hours in the month immediately after the month the child is born.

[8]For some years, the NLSY79 reports weekly employment information over 53 weeks instead of 52 weeks. In order to avoid this type of measurement error, we discard information about hours worked in the 53rd week of a year, if any. In practice, this change turns out to be immaterial for the results.

[9]The AFQT is based on a subset of tests from the Armed Services Vocational Aptitude Battery (ASVAB). Throughout, we use the current (post-1989 renormalization) definition of AFQT math as the sum of the arithmetic reasoning and mathematics knowledge subscores of the ASVAB.

[10]For children born after 1994 in odd years, the survey was not conducted in their year of birth. Then, we measure control variables in the year before birth, except family size which is measured at the year after birth in order for the child itself to be counted as part of the family.

reading recognition, and marginally below the average on math.

Table 1: Summary Statistics

	Mean	Std.Dev.
Outcome variables		
PIAT Reading Recognition	105.33	14.04
PIAT Math	99.72	14.03
Treatment variable		
Mother's average hours worked in 3 first years	847.64	838.18
Bunching variables		
Mother worked 0 hours in 3 first years	0.25	0.44
Control variables		
Mother's AFQT score	38.20	28.21
Mother's wage year prior to the birth of the child	14.69	11.04
Mother's education less than high school	0.23	0.42
Mother's education completed high school	0.43	0.50
Mother's education some college	0.19	0.40
Mother's education completed college	0.10	0.30
Mother's education more than college	0.04	0.20
Mother's age less than 20 years old	0.11	0.32
Mother's age 20 to 24 years old	0.33	0.47
Mother's age 25 to 29 years old	0.28	0.45
Mother's age 30 to 34 years old	0.18	0.39
Mother's age 35 years old or more	0.09	0.29
Mother's spouse present	0.60	0.49
Mother's spouse highest grade	12.83	2.69
Child's age at test (in months)	75.07	14.13
Sex of child (male=1, female=0)	0.51	0.50
Birth order of child	2.06	1.18
Child is Hispanic	0.21	0.40
Child is Black	0.29	0.45
Family size	3.85	1.91
Lives in north region	0.15	0.36
Lives in north-central region	0.23	0.42
Lives in south region	0.35	0.48
Lives in west region	0.19	0.39
Observations	6924	

Note: Unless specified, control variables are measured at the child's year of birth. For children born in odd years after 1994 (years that the survey is not conducted), control variables are measured at the year before birth, except family size which is measured at the year after birth. Among the control variables we also include indicator variables for the year the child took the PIAT test. The mother's wage variable is conditional in being greater than zero and it is measured per hour in 2019 dollars.

Next, the table reports statistics about maternal employment status and hours worked in the three first years of the child's life. The average annual number of hours worked in the three years following birth is 848 hours (approximately 16 hours per week) with substantial variation across children. One quarter of children in our sample have mothers who do not work during the three first years following birth.

Turning to maternal skill, on average mothers in our sample scored 38 out of 100 in the AFQT.

Since this test is set to have mean 50 and standard deviation 10 in the overall population, the mother of the average child in the sample is about one standard deviation below the national average. We also note that the AFQT scores vary notably across mothers. For our analysis, we standardize AFQT within our sample, so that it has mean zero and standard deviation one.

The remainder of the table displays summary statistics for our control variables. Most children have mothers who had completed high school and were at least 25 years old at the time of birth. Children were about 75 months old (6 years old) when they took the PIAT. The sample of children is equally balanced on gender, and is composed of 21% of Hispanic and 29% Black children. Finally, in 60% of the cases, the mother's spouse is present in the household at the time of birth, and the average child is born to a family of about three other members.

List of Controls

Here we detail the complete list of controls used in the analysis. For the mother, we use variables meant to capture her human capital: AFQT, AFQT squared, and indicators for completed education at birth: less than high school, high school only, some college, college, and more than college. We also control for her age at birth by including indicators for whether her age was ≤ 19, $\in [20, 24]$, $\in [25, 29]$, $\in [30, 34]$, or ≥ 35. As household-level controls, we include an indicator for whether the mother's spouse is present in the household, the spouse's education at birth, and the natural log of total family size. Finally, as child-level controls, we include indicators for the child's sex, race, birth order, Census geographic region, age in months at the time of the cognitive assessment, and indicators for the year of cognitive assessment.

4 Empirical Strategy

We start by specifying the child's cognitive skill S as

$$S = f(L, X; \theta) + g(X) + \epsilon, \tag{1}$$

where g is nonparametric, L is the number of hours the mother works, X is a vector of pre-determined controls, and ϵ is the unobservable error term.

We will specify the parametric function $f(\cdot; \theta)$ in alternative ways in order to improve our understanding of heterogeneity in the effects of L on S. Regardless of the specification of the parametric function $f(\cdot)$, our goal is to identify θ. The challenge is that L is endogenous – L and ϵ are not independent from each other conditional on X. Thus, a regression of S on L and X alone will yield a biased estimate of θ.

4.1 A Model of Constrained Labor Supply

In this section, we model how work hours, L, is chosen by mothers, and then show how we can leverage bunching in this model to control for unobserved heterogeneity. Mothers face a constrained optimization problem when deciding how many hours to work because the actual number of work hours chosen is constrained to be non-negative. Specifically, we write the *desired* number of work hours, L^* as

$$L^* = h(X) + \eta, \tag{2}$$

where h is nonparametric, and

$$L = \max\{0, L^*\}, \quad \text{with } \mathbb{P}(L^* < 0) > 0. \tag{3}$$

Equation (2) is written without loss of generality, as η is simply defined as the difference between L^* and $h(X)$. Equation (3) is motivated by the fact that many mothers are bunched at zero hours, as shown in Table 1 (see also Figure 2 below). Intuitively, some mothers are at a corner solution: they would like to have spent less than zero hours working in order to do something else with that additional time, but they cannot.

Next, we add some structure. We open the error term in equation (1) so that $\epsilon = \delta(X)\eta + \varepsilon$:

$$S = f(L, X; \theta) + g(X) + \delta(X)\eta + \varepsilon, \quad \text{with } \mathbb{E}[\varepsilon | L, X, \eta] = 0, \tag{4}$$

where η is the potential confounder, as it enters both equations (2) and (4) whenever $\delta(X) \neq 0$. This equation assumes that any unobserved confounder can be written as a linear function of η, $\delta(X)\eta$, where the slope may change depending on the value of the covariates X.

We exploit bunching to identify θ under the model specified by equations (2), (3) and (4). To see how bunching allows us to do this, consider the children of mothers bunched at $L = 0$. All such children have the same treatment $L = 0$, and remaining differences in their observables X can be controlled for. Thus, any systematic differences in S between children with the same value of $L = 0$ and the same observables X must occur because of differences in η, per equation (4). This allows us to isolate the effect of η on S, which we can use to build a control function approach to identify θ. Next, we detail how we implement this idea.

4.2 Control Function Approach

The model laid out in the previous section defined by equations (2), (3) and (4) implies

$$\mathbb{E}[S|L, X] = f(L, X; \theta) + \underbrace{g(X) - \delta(X)h(X)}_{m(X)} + \delta(X)[L + \mathbb{E}[L^*|L = 0, X]\mathbf{1}(L = 0)]. \tag{5}$$

If we identify $\mathbb{E}[L^*|L = 0, X]$, then we can add the term $L + \mathbb{E}[L^*|L = 0, X]\mathbf{1}(L = 0)$ to the equation as another control, allowing us to identify $f(\cdot)$, $m(\cdot)$ and $\delta(\cdot)$. In order to identify $\mathbb{E}[L^*|L = 0, X]$, we need a second identifying assumption, this time on the distribution of η (which affects the distribution of L^* via equation (2)).

Assumption 1. *(Nonparametric Tail Symmetry) For all censored quantiles* q_0*,*

$$\eta|X \textit{ has symmetric tails below } q_0 \textit{ and above } 1 - q_0,$$

In order to provide more context about the robustness of our results to distributional assumptions, we also show results under stronger but more standard assumptions:

Assumption 1′. *(Semiparametric Normal)*

$$\eta | X \sim \mathcal{N}(l(X), \sigma^2(X)),$$

Assumption 1″. *(Semiparametric Uniform)*

$$\eta | X \sim U[\kappa(X), \mu(X)].$$

These alternative distributional assumptions are testable, at least partially (see Caetano et al. 2021 for details). While the normal distribution is a standard choice, we also consider the uniform distribution because it seems to fit the data well for most values of X. We also consider nonparametric full symmetry, which, while stronger than tail symmetry, is testable. We are never able to reject nonparametric symmetry in our application. Therefore, we take as our preferred estimates throughout the paper those estimated under nonparametric tail symmetry (Assumption 1).

Our control function approach thus makes two identifying assumptions beyond the standard rank condition:[11] (a) the selection-on-unobservables assumption in equation (4), and (b) the nonparametric tail symmetry assumption discussed above. Importantly, this control function approach makes no exclusion restriction, so there is no need for IVs, which are difficult to find in this context. Intuitively, we circumvent the need for IVs because of the constraint in the choice of L. Due to the binding restriction for some mothers, $\mathbb{E}[L^*|L = 0, X]$ is often negative, which makes the last term of equation (5) discontinuous at $L = 0$. This allows us to identify $f(\cdot)$ and $\delta(\cdot)$ separately using the discontinuity in $\mathbb{E}[Y|L, X]$ at $L = 0$. Note that controls X play no special role in this identification strategy, beyond controlling for further endogenous variation. In particular, there is no separability or exogeneity assumption in equation (4) with regard to X.

4.3 Estimation Details

Section 3 details the list of controls X, and some of the elements of this vector are continuous. As discussed in Caetano et al. (2021), in order to maintain the predominantly nonparametric structure of the model (instead of resorting to linear structures due to the high dimensionality of X) it is a good idea to "discretize" X before estimating the expectation $\mathbb{E}[L^*|L = 0, X]$. Let $\{\hat{\mathcal{C}}_1, \ldots, \hat{\mathcal{C}}_K\}$ be a finite partition of the support of X into sets, which we call clusters, and let $\hat{C}_K = (\mathbf{1}(X \in \hat{\mathcal{C}}_1), \ldots, \mathbf{1}(X \in \hat{\mathcal{C}}_K))'$ be the cluster indicators. In the estimation of the expectation, we substitute X with $\hat{C}_K$, which has finite support. The estimator $\hat{\mathbb{E}}[L^*|L = 0, X] = \hat{\mathbb{E}}[L^*|L = 0, \hat{C}_K]$ is thus constructed using a two-step procedure in which first X is discretized and then one of the distributional assumptions is applied separately for each cluster.

[11] The rank condition that allows for the linear independence between the first and last terms of equation (5) follows trivially from bunching, $\mathbb{P}(L^* < 0) > 0$, and from $f(\cdot)$ being continuous at $L = 0$. In our context, $f(\cdot)$ is continuous at $L = 0$, since working a few hours in the first three years of the child's life should have a small effect on the cognitive skill of the child around age 6.

In general, if $\mathbb{E}[L^*|L = 0, X]$ is continuous, then $\hat{\mathbb{E}}[L^*|L = 0, \hat{C}_K]$ will approximate $\mathbb{E}[L^*|L = 0, X]$ as K grows. Indeed, note that as K grows the observations within the same cluster have increasingly more similar values of X.[12]

Using a similar logic, we also use the same clusters to make sure the functions $m(X)$ and $\delta(X)$ in equation (5) approximate nonparametric functions of X. For instance, we specify $m(\cdot)$ as $m(X) = X'\tau + \sum_{k=1}^{K} \alpha_k \mathbf{1}(X \in \mathcal{C}_k)$ (the specification of $\delta(\cdot)$ is analogous), so the cluster indicators control nonparametrically for differences across clusters, while differences within cluster due to X are controlled linearly. As the number of clusters K increases, the nonparametric match improves, leaving less unexplained variation within cluster. We show that the main estimates do not change when the total number of clusters K grows, thus suggesting that our approach of conditioning on these indicators well approximates conditioning on a nonparametric function of X.

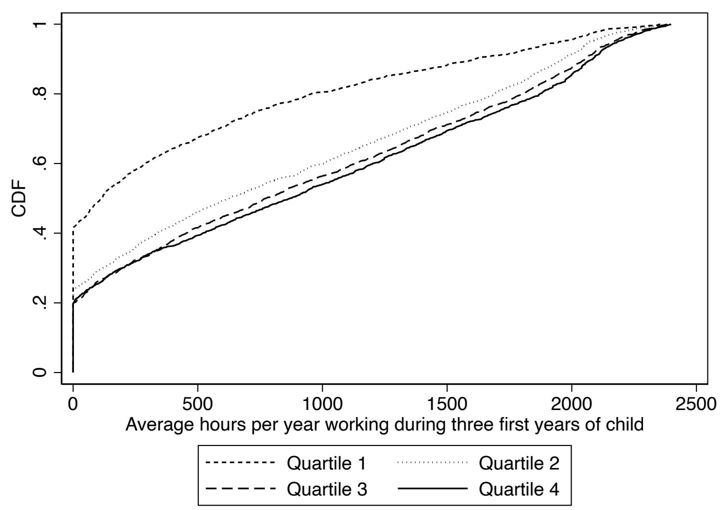
4.4 Evidence of Bunching and Selection

Before proceeding to the main results, we present here evidence in support of the model and identification strategy laid out above.

Evidence of Bunching

We begin by showing in Figure 2 that maternal labor supply has a notable bunching point at $L = 0$ for mothers of all ability levels. Specifically, the curves show the cumulative distribution of mothers' working hours, L, for each quartile of maternal AFQT. About 40% of the mothers in the lowest

Figure 2: Evidence of Bunching by Quartile of the Maternal AFQT Distribution



Note: This figure shows the cumulative density function (CDF) of the average yearly hours mothers have worked in the three years following the birth of their child for each quartile of the maternal AFQT distribution.

quartile choose to work exactly zero hours, while only a small proportion of them choose to work

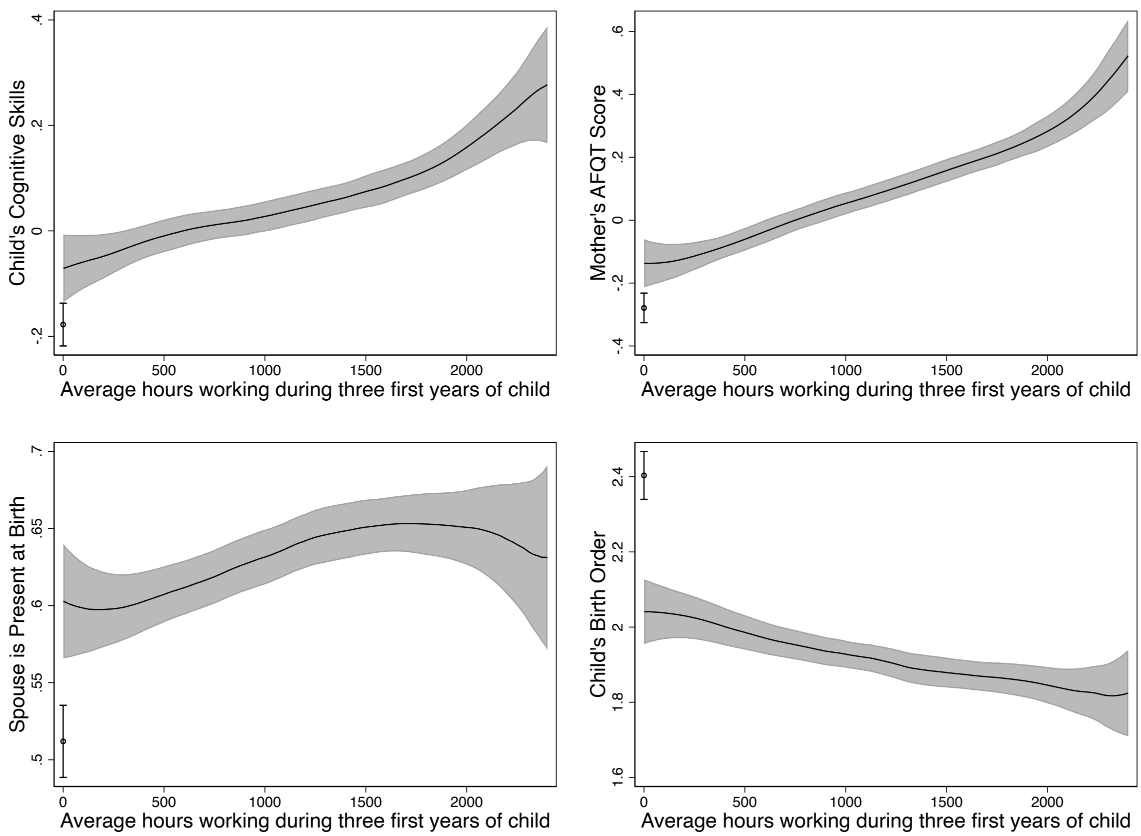
[12]We show results using hierarchical clustering (with the Gower measure of distance and Ward's linkage) for its simplicity, stability, and ease of interpretation as we vary the number of clusters.

a few hours. The degree of bunching at $L = 0$ tends to vary with the AFQT score of the mother, as expected. Nonetheless, we find a substantial amount of bunching at zero hours for all quartiles. This evidence suggests that many mothers, irrespective of their ability, find themselves at a corner solution when deciding how many hours to work. This will allow us to identify $\delta(X)$, the effect of the confounder on skills, from a well-rounded sample of mothers.

Evidence of Selection

Next, we show evidence that the effect of η on skills S tends to be positive (i.e., $\mathbb{E}[\delta(X)] > 0$). The top left panel from Figure 3 shows the local linear regression of S on L, estimated for the sample such that $L > 0$. We also show the average value of S in the subsample at $L = 0$. The figure makes clear that the children whose mothers bunch at $L = 0$ have discontinuously lower cognitive skills. The sign of the implied selection is intuitive – mothers whose unobservables make them more likely to work longer hours are also likely to have higher skill children.

Figure 3: Evidence of Positive Selection



Note: This figure shows the local linear regression of S (child's cognitive skills at age 6, top left panel) or an observed key covariate (other panels) on L (average hours working per year during the first 3 years of the child) along with the 95% confidence interval. The bandwidth is 400 hours. At $L = 0$, the average along with the 95% confidence interval is also shown.

The remaining panels from Figure 3 provide additional context by presenting the discontinuity in key covariates that are correlated with the cognitive skill of the child, S. Mothers at $L = 0$ have discontinuously lower AFQT scores and lower likelihood of the spouse being present at birth, and the children of such mothers tend to have discontinuously more older siblings. Overall, the discontinuities in Figure 3 suggest positive selection into labor supply.

Of course, it is in principle possible that observables X are enough to control for this selection. However, this hypothesis turns out not to be true. When we test the assumption $\mathbb{E}[\epsilon|L, X] = 0$ in equation (1) with Caetano (2015)'s discontinuity test based on the bunching at $L = 0$, we reject this selection-on-observables assumption even at the 1% level of significance. This motivates our selection-on-unobservables approach described in Section 4.2.

5 Results

In this section, we present our estimates of θ for different specifications of the function $f(L, X; \theta)$. We also specify $\delta(X)$ to be simple functions of the controls X, in order to facilitate the interpretation of our results, as these functions make clear how selection varies with some key observables. In the Appendix we show that our estimates of θ are robust to specifications that allow $\delta(X)$ to depend nonparametrically on X.

5.1 Homogeneous, Linear Results

We start from the most parsimonious specification of $f(L, X; \theta)$,

$$f(L, X; \theta) = \beta L \tag{6}$$

This is a helpful benchmark specification of $f(L, X; \theta)$, as it is the most comparable to the specifications in prior literature.

Table 2 presents these homogeneous, linear estimates. Column (i) presents the results of simple regressions of skills on maternal working hours with no additional controls. Cognitive skills are strongly positively associated with maternal work hours. However, column (ii), which adds observable controls, $m(X)$, to the specification in column (i), shows that pre-determined observables remove most of this positive relationship – the residual regression coefficient is close to zero and insignificant.

Columns (iii)-(v) present our estimates correcting for endogeneity using the control function approach outlined in Section 4. Each column differs only in the assumption made on the distribution of $\eta|X$. Column (iii) supposes that $\eta|X$ is normally distributed, with mean and variance that depend on the cluster to which X belongs. Column (iv) supposes that $\eta|X$ is uniformly distributed with the upper and lower limits of the support depending on the cluster of X. Finally, column (v) supposes that $\eta|X$ is symmetric in the tails, so that the unobserved portion of $\eta|X$ below 0 is the mirror image of the corresponding tail above 0, conditional on the cluster of X. As discussed in Section 4, these are our preferred estimates, as they rely on our weakest distributional assumption.

Table 2: The Effect of Maternal Hours Worked on Early Childhood Cognitive Skills

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
β	0.014**	0.000	-0.016**	-0.019**	-0.019**
	(0.001)	(0.001)	(0.005)	(0.006)	(0.005)
δ			0.014**	0.017**	0.017**
			(0.004)	(0.005)	(0.005)

Note: This table shows estimates of the effect of an additional 100 hours per year working in the three years following the child's birth on the child's early cognitive skills. N=6,924. Bootstrapped standard errors in parentheses (1,000 bootstrap samples). The corrected specifications use 50 clusters and include cluster indicators as controls. See Table 6 in Appendix B for analogous results when using a nonparametric function $\delta(X)$. ** p<0.05, * p<0.1.

Table 2 reveals that maternal labor supply has quantitatively large and statistically significant negative effects on cognitive skills. Column (v) suggests that, on the margin, an additional 100 hours of maternal labor supply annually over the first three years of a child's life tends to lower cognitive skills by around 0.019 standard deviations. This effect is economically large given the sample variance in maternal labor supply – a one standard deviation (s.d.) increase in maternal labor supply over the first three years of life would translate to reductions in cognitive skills of about 0.15 s.d. The results assuming normal or uniform distributions for η are quite similar to the symmetric estimates.

Table 2 also shows the estimates of δ, the average effect of confounder η on the outcome Y. These estimates are positive and significant, consistent with what we find in Section 4: mothers who work more hours tend to be positively selected relative to those who work fewer hours.

Robustness

It is useful to note what happens to the estimates of β as the controls X are added - i.e. as we go from columns (i) to (ii) in Table 2. The estimate of β, which was large and positive without controls (column (i)), goes down to zero when controls are added (column (ii)). Thus, our negative results in columns (iii)-(v) simply show that when we further control for unobservables, the estimate of β goes down further, indicating that selection on unobservables tends to go in the same direction as selection on observables. In Appendix A, we formalize this idea by implementing the method developed in Oster (2019) to show that our main estimate from column (v) in Table 2 is consistent with selection on unobservables going in the same direction as selection on observables. It is also consistent with a degree of selection on unobservables that is smaller than the degree of selection on observables, which is plausible (Altonji, Elder, and Taber (2005)). Importantly, this sensitivity analysis relies on completely different assumptions and does not use bunching of the treatment variable in any way, so we view these results as evidence that our findings are not an artifact of our assumptions.

All of the results we report in any of our tables use $K = 50$ clusters (see Section 4.3 for details). In Figure 4 in Appendix B, we replicate the analysis in column (v) of Table 2 for $K = 2, \ldots, 100$. The estimates of β are very similar for all values of K, and remarkably so for K ranging from $K = 50$ all the way to $K = 100$. Since as K grows we are better able to approximate a nonparametric match on controls X, this gives us confidence that $K = 50$ is enough for such approximation.

5.2 Heterogeneity by Maternal AFQT and Labor Supply

We now turn to assessing heterogeneity in the effect of maternal labor supply by the AFQT skill of the mother and the intensity of her labor supply. We specify $f(L, X; \theta)$ as a quadratic function of hours, with the slope and convexity allowed to be heterogeneous by the normalized AFQT skill of the mother, A:

$$f(L, X; \theta) = \beta L + \beta_A A L + \beta_L L^2 + \beta_{AL} A L^2. \tag{7}$$

Table 3 presents the heterogeneous results. As before, we discuss the results of our preferred estimates under nonparametric tail symmetry. The other distributional assumptions yield qualitatively and quantitatively similar estimates. As with the homogeneous, linear estimates in Table 2, we use $K = 50$ clusters, but our results turn out to not depend on the choice of K (see Figure 5 in Appendix B).

Table 3 reveals a number of interesting patterns. First, $\beta < 0$ – the estimated effect on cognitive skills of the first 100 working hours for a mother with average skills ($A = 0$) is negative. Second, $\beta_A < 0$ – the effect of work hours on cognitive skills is more negative the higher is the skill of the mother. This effect is statistically significant and quite large; each additional standard deviation of mother's AFQT lowers β by 0.026 s.d. Third, $\beta_L > 0$ – the effect of work hours for a mother with average AFQT skills become larger (or less negative) the more total hours are worked, although the degree of convexity is small and not significant.[13] Fourth, $\beta_{AL} > 0$ – the effects of working hours are more convex for higher skill mothers. These results are illustrated in Figure 1 in the Introduction for four hypothetical children who are representative of each quartile of the distribution of skill of their mothers. Each child's skill and corresponding maternal working hours is represented by a hollow circle in the figure, which together illustrate that mothers of higher skill tend to work longer hours. As their mothers works longer hours, the children's skills go down, but more intensely for the children with higher skilled mothers.[14] The figure also shows that, for the range of hours in the data, the curves are approximately linear. On average, the estimates in Table 3 imply more negative βs than the homogeneous, linear estimate in Table 2: the average estimated effect in the sample is -0.027.

[13] Note that the terms multiplying both β_L and β_{AL} have L^2 in them, a very large number for most working mothers. Thus, the coefficient estimates are much smaller than the other estimates. For readability, we present these estimates multiplied by 1,000.

[14] For some children in our sample whose mothers have sufficiently low skills, Table 3 actually implies positive effects. In particular, of the 6,924 children in our sample, 592 have positive estimated effects of maternal hours on skills. These positive effects are quite small, however, with an average of just 0.002 and a maximum of 0.005.

Table 3: The Effect of Maternal Hours Worked on Early Childhood Cognitive Skills by the AFQT Score and Labor Supply of the Mother

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
β	0.018**	0.003	-0.023**	-0.026**	-0.030**
	(0.004)	(0.003)	(0.008)	(0.010)	(0.009)
β_A	0.047**	-0.010**	-0.017**	-0.021**	-0.026**
	(0.003)	(0.004)	(0.008)	(0.009)	(0.009)
β_L ($\times 1000$)	-0.004**	-0.001	0.002	0.001	0.002
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
β_{AL} ($\times 1000$)	-0.015**	0.003**	0.004**	0.004**	0.004**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
δ			0.016**	0.022**	0.023**
			(0.005)	(0.007)	(0.006)
δ_A			0.005	0.009	0.013**
			(0.005)	(0.006)	(0.006)

Note: This table shows estimates of the effect of an additional 100 hours per year working in the three years following the child's birth on the child's early cognitive skills. N=6,924. Bootstrapped standard errors in parentheses (1,000 bootstrap samples). The corrected specifications use $K = 50$ clusters and include cluster indicators as controls. Figure 5 in Appendix B shows that each of the parameter estimates are not sensitive to the choice of K. See Table 7 in Appendix B for analogous results when using a nonparametric function $\delta(X)$. ** p<0.05, * p<0.1.

The estimates of $\delta(X)$ are also intuitive. While we continue to find evidence of positive selection (as in the homogeneous results), here we also find evidence that the positive selection is more intense for mothers with higher skills (that is, $\delta_A > 0$).

Robustness

It is worth mentioning that our key heterogeneity finding, $\beta_A < 0$, is true even when we only control for X under a standard OLS regression, as can be seen in column (ii) from Table 3.

We can provide further assurances about the true value of β_A under relatively weak assumptions. As discussed in Caetano et al. (2021), the sign of δ_A can be inferred under no distributional assumption. This means that under only the selection on unobservables assumption from equation (4), we can infer that the true value of β_A is smaller (i.e. more negative) than the estimate of β_A from column (ii), which matches our findings.

Table 7 in the appendix also shows that the findings in this section do not change when we allow $\delta(X)$ to vary nonparametrically with X. Finally, Figure 5 in Appendix B demonstrates that the results in column (v) of Table 3 are robust to using different values of K between 50 and 100.

5.3 Further Heterogeneity by Pre-Birth Wages

The results so far suggest a misalignment between the work incentives facing mothers and the short-term benefits this work yields for their children's skills. The mothers who work the most are on average those whose work has the most negative short-term consequences for their children. Moreover, it does not appear that working longer hours is the reason why maternal work is particularly detrimental to the skills of the children from higher-skilled mothers. One explanation for this pattern is that the additional money earned by these higher-skilled mothers might not be enough to offset the higher opportunity cost of working (and thus not interacting as much with their child). In order to further investigate this hypothesis, we expand equation (7) by allowing for a third dimension of heterogeneity, the pre-birth wage rate of the mother, W:[15]

$$f(L, X; \theta) = \beta L + \beta_A AL + \beta_W WL + \beta_L L^2 + \beta_{AL} AL^2 + \beta_{WL} WL^2. \tag{8}$$

We allow for the slope and convexity of the effect of hours to depend on maternal skills and wage rates in a separable way.[16] Intuitively, we want to separately identify the effect in skills due to the loss of interaction between the child and the mother because she is working and any potential way to offset this loss using the additional money received due to this work.[17] We assume that the heterogeneous effects via A (AFQT skills) holding W (wages) constant tend to incorporate mostly the loss of interaction between the mother and the child, while the heterogeneous effects via W (pre-birth wages) tend to incorporate mostly the potential for additional earnings to offset this loss, which may happen via higher-quality child care, increased goods purchases (better food, more books, etc.), a reduction in parental stress, or in many other ways that may be difficult to observe. Indeed, it is plausible that the skills that are valued in the job market (affecting wages) aside from AFQT may have only a small effect on the quality of the interaction between the mother and the child during the first three years of life of the child.

Table 4 presents the estimates of the function $f(\cdot)$ as specified in equation (8). For this table, we must restrict the sample to women who were working prior to giving birth and we must include the pre-birth wage rate as an additional control. The results from the more restricted functions $f(L, X; \theta)$ estimated in Tables 2 and 3 are similar when we impose such changes. As expected, the estimates are more noisy once this third dimension of heterogeneity is included. However, this exercise is still informative. Comparing our preferred symmetric estimates to the analogous estimates in Table 3 reveals a number of noteworthy results. First, as expected, the heterogeneity in the effect by the mother's skill (β_A) becomes more intense, since now it does not incorporate

[15] Specifically, W denotes a residualized wage measure in which the effect of age of mother-at-birth and year fixed effects have been removed. We standardize this measure so that it is in standard deviation units like our AFQT measure. The estimates of β_W and β_{WL} should thus be more comparable to the estimates of β_A and β_{AL}.

[16] We also tried to allow for further heterogeneity, by including the interaction $A \cdot W$. Including this additional interaction barely changes the main estimates, while the interaction itself is imprecisely estimated.

[17] Using other variables in the data to more directly conduct this analysis is infeasible. There are two potential variables from the NLSY that speak more directly to whether the mother needs to distance herself from the child when working: whether the job has flexible working hours and how many hours per week the person works from home. Unfortunately, these variables have too many missing observations to be useful for our purposes.

Table 4: The Effect of Maternal Hours Worked on Early Childhood Cognitive Skills by the AFQT Score, Pre-Birth Wages, and Labor Supply of the Mother

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
β	-0.004	-0.002	-0.009	-0.013	-0.017
	(0.005)	(0.004)	(0.014)	(0.026)	(0.017)
β_A	0.039**	-0.016**	-0.027**	-0.047**	-0.035**
	(0.004)	(0.004)	(0.014)	(0.023)	(0.017)
β_W	0.007*	0.001	0.014	0.029	0.018
	(0.004)	(0.005)	(0.016)	(0.026)	(0.019)
β_L ($\times$1000)	0.002	0.001	0.001	0.001	0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
β_{AL} ($\times$1000)	-0.012**	0.005**	0.006**	0.007**	0.007**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
β_{WL} ($\times$1000)	-0.002	0.001	-0.000	-0.001	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
δ			0.006	0.010	0.013
			(0.010)	(0.022)	(0.013)
δ_A			0.008	0.027	0.015
			(0.009)	(0.020)	(0.013)
δ_W			-0.009	-0.024	-0.014
			(0.010)	(0.021)	(0.014)

Note: This table shows estimates of the effect of an additional 100 hours per year working in the three years following the child's birth on the child's early cognitive skills. N=3,994. Bootstrapped standard errors in parentheses (1,000 bootstrap samples). In addition to the controls in Table 3, these models also include pre-birth wages and the interaction of pre-birth wages and maternal AFQT. The corrected specifications use 50 clusters and include cluster indicators as controls. See Table 8 in Appendix B for analogous results when using a nonparametric function $\delta(X)$. ** p<0.05, * p<0.1.

the offsetting income mechanism. Second, there is some evidence of a positive offsetting effect due to wages (β_W), although this is not significant at standard levels. Note also that there is a bit more evidence of nonlinearity by mother's skill (β_{AL}), and there is no evidence of nonlinearity by wages ($\beta_W L$). Still, the linearity assumption made in the literature continues to be a reasonable approximation for the range of hours, skills and wages in the sample.

Because of the noisy results, it is difficult to rule out some positive value for β_W. A positive value for β_W is intuitive: mothers with higher pre-birth wages are likely paid more for each hour they work post-birth, and this additional income could be beneficial to their child's skills through a variety of mechanisms already discussed.[18] Nonetheless, it seems that there is much less heterogeneity across

[18] We also estimate specifications analogous to Table 3 but with pre-birth wages taking the place of AFQT, also

different values of W than across different values of A. To see this, consider the point estimates of $\beta_A = -0.035$ and $\beta_W = 0.018$ from the table, and note that a mother who is 1 s.d. above the average in terms of A tends to be only one quarter of a s.d. above the average in terms of W. While the child of such a mother is expected to lose 0.035 s.d. in cognitive skills for each additional hour she works (relative to the average mother), only $0.018/4 = 0.0045$ s.d. would be expected to be offset by her higher earnings. In other words, a mother who is 1 s.d. above the average in skills would have to earn roughly 8 times the salary that she would be expected to earn given her skill to fully offset the short-run impact on the cognitive skills of her child.

Although more noisy, the estimates of $\delta(X)$ continue to suggest some positive selection, and disproportionately positive selection for higher-skilled mothers, as before. The table also shows some evidence that the selection might be less positive for high-skill, high-wage mothers relative to high-skill, low-wage mothers.

Robustness

As before, it is useful to note that our key heterogeneity findings, $\beta_A < 0$ and $\beta_W \approx 0$, are true even when we only control for X under a standard OLS regression, as can be seen in column (ii) from Table 4. Thus, our main qualitative findings are also valid even without the use of our control function approach.

Moreover, under only the assumption from equation (4) (i.e., under no distributional assumption, as discussed in Caetano et al. (2021)), we can further infer that the true value of β_A is more negative than the one we find in column (ii) from Table 4, which matches our findings. Similarly, we can infer that the true value of β_W is more positive than the one we find in column (ii) from Table 4, which also matches our findings.

In Table 8 in the appendix, we also show that these results do not change when we allow for $\delta(X)$ to vary nonparametrically with X.

6 Conclusion

In this paper, we estimate the effect of maternal hours worked in the first three years of life on early childhood cognitive skills. We correct for the endogeneity of labor supply using a control function approach that leverages the bunching of some mothers at zero hours worked. We allow for heterogeneity in these effects by maternal skill, hours worked, and maternal pre-birth wages, which together improve our understanding of the trade-offs mothers and their families face when deciding whether and how many hours to work.

We find that working longer hours has a negative effect on the cognitive skills of the child, particularly for higher-skill mothers. Our results suggest that the presence of high-skill mothers in the home is particularly valuable for childhood skill accumulation, at least in the short run.

obtaining little evidence for heterogeneous effects by the mother's wage even when AFQT skills are not explicitly specified in the function $f(L, X; \theta)$.

Our results provide some useful insights for policies aimed at incentivizing mothers to work in the first years of their children's lives. We confirm the finding from previous literature that there is some scope for such policies to generate short-run, negative effects on children's skills. Our main contribution lies in providing a detailed analysis of potential heterogeneity in these effects along several dimensions, allowing us to assess for whom these unintended consequences are more likely to be important. We find that there is little scope for such unintended consequences among low-skilled mothers, even for those who work long hours. However, we also find that the time of high-skill mothers appears to be particularly hard to substitute for in early childhood, even given the higher earnings that such mothers typically obtain. On the one hand, our finding that these unintended consequences tend to be concentrated towards families with higher skilled mothers may be reassuring, since these are the families who are likely to have more resources to mitigate any potential downsides as the child grows. On the other hand, skill development is a dynamic process, and skill losses early in life may be particularly consequential. Understanding the long-run consequences of maternal labor supply and their implications for intergenerational inequality is an important research topic left for future work.

We conclude that policies aimed at increasing the flexibility of work arrangements are more likely to avoid these unintended negative consequences of maternal labor supply in early childhood. Such policies might allow mothers to maintain their work hours and career development while also allowing them to spend valuable time with their children. It may also allow their partners to be a better substitute for their absence. The need for social distancing during the Covid-19 pandemic has led to marked increases in the share of workers operating under more flexible work arrangements (e.g., work from home). In future work, it would be interesting to study the childhood development consequences of these recent changes, particularly if they prove persistent.

References

Adda, J., C. Dustmann, and K. Stevens (2017). The career costs of children. *Journal of Political Economy 125*(2), 293–337.

Agostinelli, F. and G. Sorrenti (2021). Money vs. time: Family income, maternal labor supply, and child development. *Working Paper* (273).

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy 113*(1), 151–184.

Anderson, P. M. and P. B. Levine (1999). Child care and mothers' employment decisions. Technical report, National bureau of economic research.

Andresen, M. E. and T. Havnes (2019). Child care, parental labor supply and tax revenue. *Labour Economics 61*, 101762.

Arcidiacono, P., P. Bayer, and A. Hizmo (2010). Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics 2*(4), 76–104.

Averett, S. L., H. E. Peters, and D. M. Waldman (1997). Tax credits, labor supply, and child care. *Review of Economics and Statistics 79*(1), 125–135.

Baker, M., J. Gruber, and K. Milligan (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy 116*(4), 709–745.

Bartik, A. W., Z. B. Cullen, E. L. Glaeser, M. Luca, and C. T. Stanton (2020). What jobs are being done at home during the covid-19 crisis? evidence from firm-level surveys. Technical report, National Bureau of Economic Research.

Baum II, C. L. (2003). Does early maternal employment harm child development? an analysis of the potential benefits of leave taking. *Journal of Labor Economics 21*(2), 409–448.

Baydar, N. and J. Brooks-Gunn (1991). Effects of maternal employment and child-care arrangements on preschoolers' cognitive and behavioral outcomes: Evidence from the children of the national longitudinal survey of youth. *Developmental Psychology 27*(6), 932.

Bernal, R. (2008). The effect of maternal employment and child care on children's cognitive development. *International Economic Review 49*(4), 1173–1209.

Bick, A., A. Blandin, and K. Mertens (2020). Work from home after the covid-19 outbreak. *CEPR Discussion Paper No. DP15000*.

Bick, A. and N. Fuchs-Schündeln (2017). Quantifying the disincentive effects of joint taxation on married women's labor supply. *American Economic Review 107*(5), 100–104.

Blau, D. and J. Currie (2006). Pre-school, day care, and after-school care: who's minding the kids? *Handbook of the Economics of Education 2*, 1163–1278.

Blau, D. M. (1999). The effect of income on child development. *Review of Economics and Statistics 81*(2), 261–276.

Blau, D. M. and A. P. Hagy (1998). The demand for quality in child care. *Journal of Political Economy 106*(1), 104–146.

Blau, F. D., A. J. Grossberg, et al. (1992). Maternal labor supply and children's cognitive development. *The Review of Economics and Statistics 74*(3), 474–481.

Blundell, R., M. Costa Dias, C. Meghir, and J. Shaw (2016). Female labor supply, human capital, and welfare reform. *Econometrica 84*(5), 1705–1753.

Bono, E. D., M. Francesconi, Y. Kelly, and A. Sacker (2016). Early maternal time investment and early child outcomes. *The Economic Journal 126*(596), F96–F135.

Bosch, N. and B. Van der Klaauw (2012). Analyzing female labor supply—evidence from a dutch tax reform. *Labour Economics 19*(3), 271–280.

Caetano, C. (2015). A Test of Exogeneity Without Instrumental Variables in Models With Bunching. *Econometrica 83*(4), 1581–1600. Available here.

Caetano, C., G. Caetano, and E. Nielsen (2021). Correcting Endogeneity Bias in Models with Bunching. *Working Paper*. Available here.

Carneiro, P., C. Meghir, and M. Parey (2013). Maternal education, home environments, and the development of children and adolescents. *Journal of the European Economic Association 11*(suppl_1), 123–160.

Carta, F. and L. Rizzica (2018). Early kindergarten, maternal labor supply and children's outcomes: evidence from italy. *Journal of Public Economics 158*, 79–102.

Chen, W., W. A. Grove, and A. Hussey (2017). The role of confidence and noncognitive skills for post-baccalaureate academic and labor market outcomes. *Journal of Economic Behavior & Organization 138*, 10–29.

Cortes, P. and J. Tessada (2011). Low-skilled immigration and the labor supply of highly skilled women. *American Economic Journal: Applied Economics 3*(3), 88–123.

Currie, J. and D. Almond (2011). Human capital development before age five. In *Handbook of labor economics*, Volume 4, pp. 1315–1486. Elsevier.

Dahl, G. B. and L. Lochner (2012, May). The Impact of Family Income on Child Achievement: Evidence from the Earned Income Tax Credit. *American Economic Review 102*(5), 1927–56.

Del Boca, D., C. Flinn, and M. Wiswall (2014). Household choices and child development. *Review of Economic Studies 81*(1), 137–185.

Desai, S., P. L. Chase-Lansdale, and R. T. Michael (1989). Mother or market? effects of maternal employment on the intellectual ability of 4-year-old children. *Demography 26*(4), 545–561.

Dingel, J. I. and B. Neiman (2020). How many jobs can be done at home? *Journal of Public Economics 189*, 104–235.

Eckstein, Z. and O. Lifshitz (2011). Dynamic female labor supply. *Econometrica 79*(6), 1675–1726.

Eissa, N. and J. B. Liebman (1996). Labor supply response to the earned income tax credit. *The Quarterly Journal of Economics 111*(2), 605–637.

Ettinger, A. K., A. W. Riley, and C. E. Price (2018). Increasing maternal employment influences child overweight/obesity among ethnically diverse families. *Journal of Family Issues 39*(10), 2836–2861.

Fogli, A. and L. Veldkamp (2011). Nature or nurture? learning and the geography of female labor force participation. *Econometrica 79*(4), 1103–1138.

Gathmann, C. and B. Sass (2018). Taxing childcare: Effects on childcare choices, family labor supply, and children. *Journal of Labor Economics 36*(3), 665–709.

Givord, P. and C. Marbot (2015). Does the cost of child care affect female labor market participation? an evaluation of a french reform of childcare subsidies. *Labour Economics 36*, 99–111.

Goux, D. and E. Maurin (2010). Public school availability for two-year olds and mothers' labour supply. *Labour Economics 17*(6), 951–962.

Grogger, J. (2003). The effects of time limits, the eitc, and other policy changes on welfare use, work, and income among female-headed families. *Review of Economics and Statistics 85*(2), 394–408.

Hill, M. A. and J. O'Neill (1994). Family endowments and the achievement of young children with special reference to the underclass. *Journal of Human Resources*, 1064–1100.

Hirsch, B. T. and E. J. Schumacher (1998). Unions, wages, and skills. *Journal of Human Resources*, 201–219.

Hsin, A. and C. Felfe (2014). When does time matter? maternal employment, children's time with parents, and child development. *Demography 51*(5), 1867–1894.

James-Burdumy, S. (2005). The effect of maternal labor force participation on child development. *Journal of Labor Economics 23*(1), 177–211.

Kalenkoski, C. M., D. C. Ribar, and L. S. Stratton (2009). The influence of wages on parents' allocations of time to child care and market work in the united kingdom. *Journal of Population Economics 22*(2), 399–419.

Løken, K. V., M. Mogstad, and M. Wiswall (2012). What linear estimators miss: The effects of family income on child outcomes. *American Economic Journal: Applied Economics 4*(2), 1–35.

Meyer, B. D. and D. T. Rosenbaum (2001). Welfare, the earned income tax credit, and the labor supply of single mothers. *The Quarterly Journal of Economics 116*(3), 1063–1114.

Milligan, K. and M. Stabile (2011). Do child tax benefits affect the well-being of children? evidence from canadian child benefit expansions. *American Economic Journal: Economic Policy 3*(3), 175–205.

Neal, D. A. and W. R. Johnson (1996). The Role of Premarket Factors in Black-White Wage Differences. *The Journal of Political Economy 104*, 869–895.

Nicoletti, C., K. G. Salvanes, and E. Tominey (2020). Mothers working during preschool years and child skills. does income compensate? *Working Paper*.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics 37*(2), 187–204.

Parcel, T. L. and E. G. Menaghan (1994). Early parental work, family social capital, and early childhood outcomes. *American Journal of Sociology 99*(4), 972–1009.

Polachek, S. W., T. Das, and R. Thamma-Apiroam (2015). Micro-and macroeconomic implications of heterogeneity in the production of human capital. *Journal of Political Economy 123*(6), 1410–1455.

Ruhm, C. J. (2004). Parental employment and child cognitive development. *Journal of Human Resources 39*(1), 155–192.

Ruhm, C. J. (2009). Maternal employment and child development. *Handbook of families and work*, 331–354.

Sayer, L. C. and J. C. Gornick (2012). Cross-national variation in the influence of employment hours on child care time. *European Sociological Review 28*(4), 421–442.

Todd, P. and K. Wolpin (2007). The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps. *Journal of Human Capital 1*(1), 91–136.

Vandell, D. L. and J. Ramanan (1992). Effects of early and recent maternal employment on children from low-income families. *Child development 63*(4), 938–949.

Waldfogel, J., W.-J. Han, and J. Brooks-Gunn (2002). The effects of early maternal employment on child cognitive development. *Demography 39*(2), 369–392.

Yamaguchi, S., Y. Asai, and R. Kambayashi (2018). Effects of subsidized childcare on mothers' labor supply under a rationing mechanism. *Labour Economics 55*, 1–17.

A Sensitivity Analysis Based on Oster (2019)

The estimates of δ in Table 2 are positive and significant, implying positive selection – a higher value of η increases both childhood skills and maternal labor supply. Positive selection is intuitive and is consistent with the discontinuities presented in Figure 3 in which the mothers bunched at zero hours were shown to have discontinuously lower levels of variables that are known to be positively correlated with cognitive skills.

In this appendix, we provide additional evidence that the degree of selection implied by our estimates is plausible. We do this by implementing the method proposed in Oster (2019), which itself builds on Altonji et al. (2005). The method requires as inputs the estimates of β and the R^2 from the regressions we ran in columns (i) and (ii) from Table 2. Given these inputs, under the assumptions from Oster (2019), we can infer the amount of selection-on-unobservables relative to selection-on-observables that are implied by the true value of β being identical to the estimate of β from column (v) in Table 2.

We report this implied ratio of selection-on-unobservables by selection-on-observables, which we denote δ_{Oster}, in Table 5. Following Oster (2019), we show these results for different potential values of R_{max}, which is the R-squared of a hypothetical regression of S on L, our observable controls, and all unobservable confounders (including some that we may have not have controlled for with our control function approach). For all possible values of R_{max}, our conclusions from Table 2 imply that selection on unobservables would be less pronounced than selection on observables, sometimes substantially less. For instance, if R_{max} is 0.70, then δ_{Oster} implies that our main estimates from Table 2 are compatible with selection on unobservables being about two-thirds as intense as selection on observables. If $R_{max} = 1$, which corresponds to the value of R_{max} suggested in Altonji et al. (2005), selection on unobservables need to be only about 40% as intense as selection on observables.

These results suggest that one does not need particularly strong selection on unobservables to rationalize our results. Importantly, Oster (2019)'s method requires completely different assumptions than ours. In particular, it does not use bunching in L and it does not make the distributional assumption we make. Thus, we view the results in Table 5 as providing independent confirmation of the plausibility of our corrected estimates.

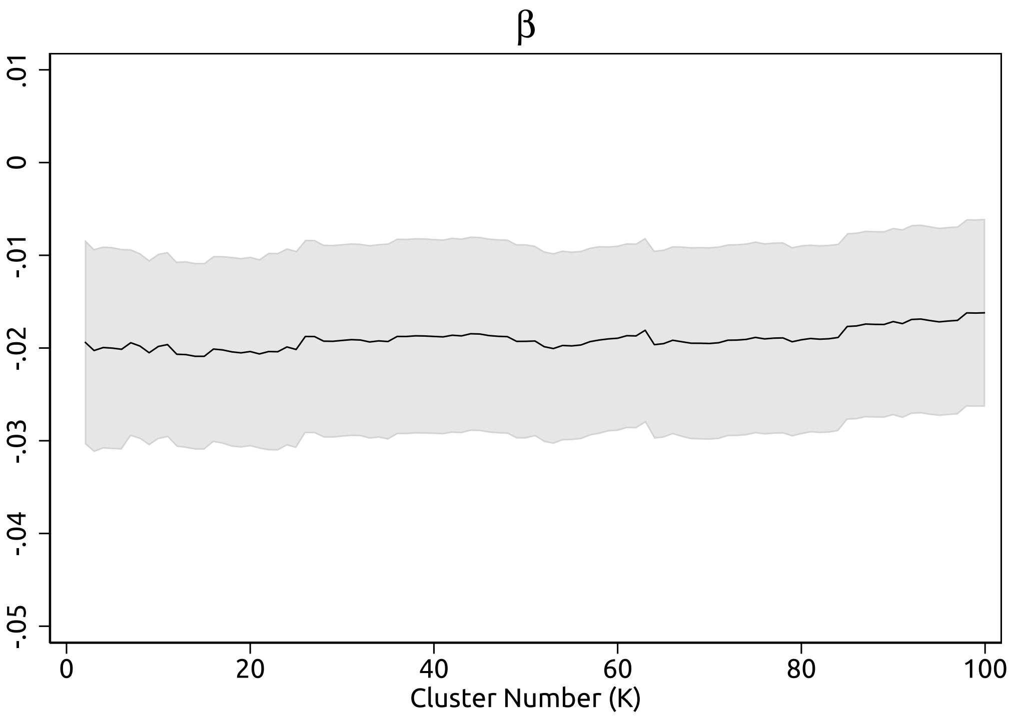
Table 5: Proportional selection of observables and unobservables, Oster (2019).

	True $\beta = -0.019$					
R_{max}	0.50	0.60	0.70	0.80	0.90	1.00
δ_{Oster}	1.12	0.82	0.65	0.54	0.46	0.40

Note: The table shows the values of δ_{Oster} as in Oster (2019) for different values of R_{max} when the true effect is $\beta = -0.019$, our estimate from Table 2. δ_{Oster} can be interpreted as the degree of selection on unobservables relative to observables.

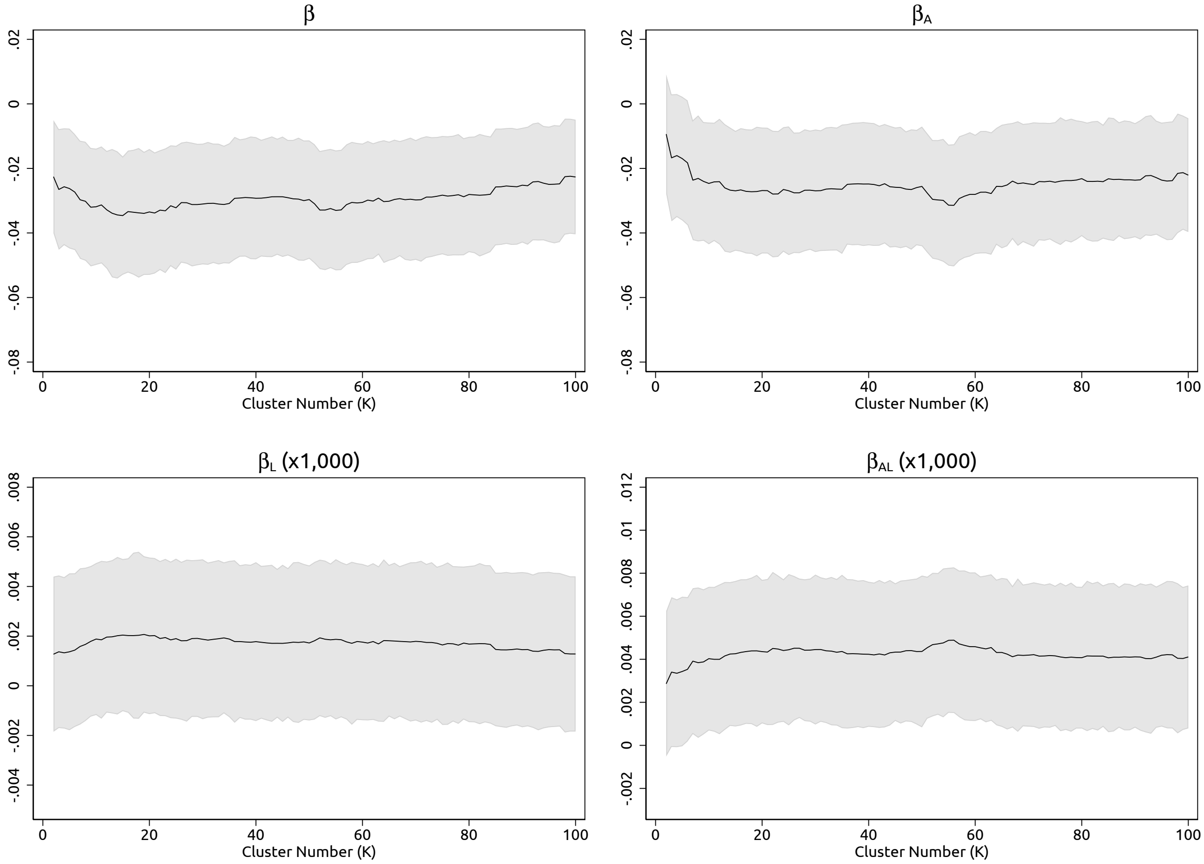
B Robustness Analysis

Figure 4: The Effect of Maternal Hours Worked on Early Childhood Skills as in Table 2 – Different Cluster Numbers (K)



Note: Estimates correspond to the Het. Symmetric estimate in Table 2 but with different numbers of clusters used in the analysis. 95% confidence intervals based on 1,000 bootstrap iterations. $N = 6,924$ for each K.

Figure 5: The Effect of Maternal Hours Worked on Early Childhood Skills by the AFQT Score and Labor Supply of the Mother as in Table 3 – Different Cluster Numbers (K)



Note: Estimates correspond to the Het. Symmetric estimate in Table 3 but with different numbers of clusters used in the analysis. 95% confidence intervals based on 1,000 bootstrap iterations. $N = 6,924$ for each K.

Table 6: The Effect of Maternal Hours Worked on Early Childhood Cognitive Skills, $\delta(X)$ by Cluster

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
β	0.014**	0.000	-0.014**	-0.019**	-0.017**
	(0.001)	(0.001)	(0.005)	(0.007)	(0.006)

Note: This table shows estimates of the effect of an additional 100 hours per year working in the three years following the child's birth on the child's early cognitive skills for specifications where $\delta(X)$ differs arbitrarily by cluster. N=6,924. Bootstrapped standard errors in parentheses (1,000 bootstrap samples). The corrected specifications use 50 clusters and include cluster indicators as controls. ** p<0.05, * p<0.1.

Table 7: The Effect of Maternal Hours Worked on Early Childhood Cognitive Skills by the AFQT Score and Labor Supply of the Mother, $\delta(X)$ by Cluster

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
β	0.018**	0.003	-0.023**	-0.031**	-0.030**
	(0.004)	(0.003)	(0.009)	(0.011)	(0.010)
β_A	0.047**	-0.010**	-0.019**	-0.024**	-0.026**
	(0.003)	(0.004)	(0.009)	(0.010)	(0.010)
β_L ($\times$1000)	-0.004**	-0.001	0.002	0.002	0.002
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)
β_{AL} ($\times$1000)	-0.015**	0.003**	0.004**	0.004**	0.004**
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
δ_A			0.008	0.013*	0.014**
			(0.005)	(0.007)	(0.006)

Note: This table shows estimates of the effect of an additional 100 hours per year working in the three years following the child's birth on the child's early cognitive skills for specifications where $\delta(X)$ differs arbitrarily by cluster. N=6,924. Bootstrapped standard errors in parentheses (1,000 bootstrap samples). The corrected specifications use 50 clusters and include cluster indicators as controls. ** p<0.05, * p<0.1.

Table 8: The Effect of Maternal Hours Worked on Early Childhood Cognitive Skills by the AFQT Score, Pre-Birth Wages, and Labor Supply of the Mother, $\delta(X)$ by Cluster

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Het. Tobit	(iv) Het. Uniform	(v) Het. Symmetric
β	-0.004	-0.002	-0.010	-0.021	-0.020
	(0.005)	(0.004)	(0.014)	(0.027)	(0.019)
β_A	0.039**	-0.016**	-0.024*	-0.040	-0.026
	(0.004)	(0.004)	(0.014)	(0.025)	(0.017)
β_W	0.007*	0.001	0.016	0.029	0.024
	(0.004)	(0.005)	(0.016)	(0.026)	(0.019)
β_L	0.002	0.001	0.001	0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
β_{AL} ($\times$1000)	-0.012**	0.005**	0.006**	0.006**	0.006**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
β_{WL} ($\times$1000)	-0.002	0.001	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
δ_A			0.006	0.021	0.009
			(0.010)	(0.022)	(0.014)
δ_W			-0.010	-0.024	-0.018
			(0.010)	(0.021)	(0.014)

Note: This table shows estimates of the effect of an additional 100 hours per year working in the three years following the child's birth on the child's early cognitive skills for specifications where $\delta(X)$ differs arbitrarily by cluster. N=3,994. Bootstrapped standard errors in parentheses (1,000 bootstrap samples). In addition to the controls in Table 2, these models also include pre-birth wages and the interaction of pre-birth wages and maternal AFQT. The corrected specifications use 50 clusters and include cluster indicators as controls. ** p<0.05, * p<0.1.