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

March 2023

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 mitigate short-term, unintended consequences of maternal labor supply through policies that promote more flexible work arrangements rather than through policies that increase the financial support to working mothers.

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 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.

*We would like to thank Joseph Altonji, Andrew Goodman-Bacon, Joseph Hotz, Josh Kinsler, Luiz Renato Lima, Jonathan Mansfield, Rodrigo Pinto, Christopher Ruhm, David Slichter, Christopher Taber, Hao Teng, 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.

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 (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 the total effect of maternal labor supply on childhood skill development is critical both for understanding the sources of childhood skill differences and as an input into various 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 – can alter mothers' labor supply choices.[1] Such policies may have unintended consequences on childhood skill development, with important implications for intergenerational mobility and inequality in general (Blau and Currie 2006, Currie and Almond 2011, Flood, McMurry, Sojourner, and Wiswall 2022).

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: higher-skilled mothers tend to work longer hours (Cortes and Tessada 2011, Adda, Dustmann, and Stevens 2017, Chen, Grove, and Hussey 2017), and this may be due in part to a higher return to working longer hours.

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 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

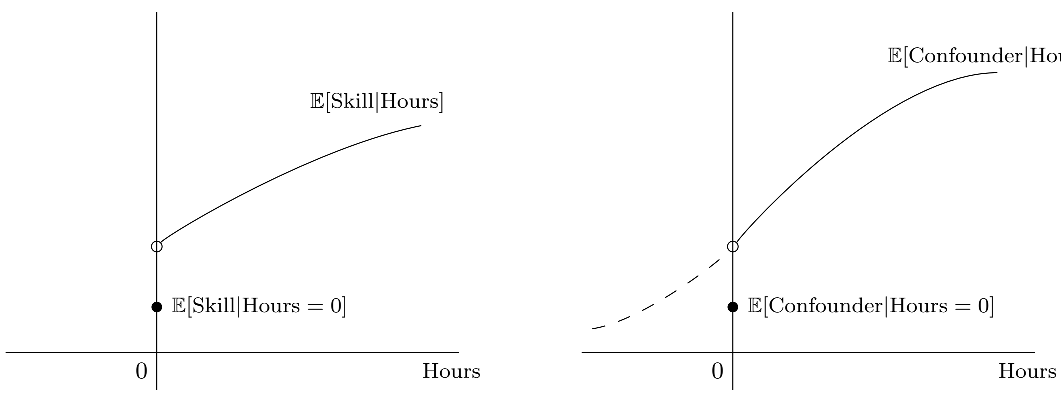
[1]The literature on tax credits effects on female labor supply is large; for some references, see work by Eissa and Liebman 1996, Averett, Peters, and Waldman 1997, Meyer and Rosenbaum 2001, Grogger 2003, Bosch and Van der Klaauw 2012, Blundell, Costa Dias, Meghir, and Shaw 2016, Bick and Fuchs-Schündeln 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; Flood et al., 2022), 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).

the observation that there is a concentration of mothers who work exactly zero hours (Caetano, Caetano, and Nielsen 2023). This control function approach helps us add value to this literature because it 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 important dimensions of heterogeneity, which enrich our understanding of the relationship between maternal labor supply and child development. It also allows us to recover average treatment effects using an entirely new source of variation relative to prior literature.

Figure 1 illustrates the source of variation in a simple context without controls. The left panel shows hypothetically how the outcome variable (child's skill) varies with the treatment variable (working hours of the mother).[4] The positive slope in this panel combines the treatment effect of maternal hours on the child's skill (what we want to identify) and the endogeneity bias, as confounders are not held constant when the treatment variable varies in the horizontal axis. The left panel also shows a discontinuity in the expected skill at zero: the average skill of children whose mothers work zero hours is sharply lower than the average skill of children whose mothers work just a few hours per year. It is not plausible that this discontinuity represents a treatment effect: working just a few hours over the first three years of the child's life, instead of zero, is unlikely to generate such a sharp increase in the skill of the child at age 6.

Figure 1: Isolating the Effect of Confounders



The right panel of Figure 1 offers an explanation for this discontinuity. It illustrates how a hypothetical confounder varies with the treatment variable. For concreteness, we can think of the confounder as positively correlated with the treatment variable, so that mothers with higher values of the confounder tend to work longer hours. This confounder likely reflects a combination of many unobserved factors that correlate to both maternal working hours and the skill of the child, such

[3]Another potential concern with understanding heterogeneous effects using IVs is that the "compliers" of a given IV – who cannot be directly observed – 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-at-zero group.

[4]See the left panel of Figure 5 in Section 5 for the empirical version of this figure in our setting.

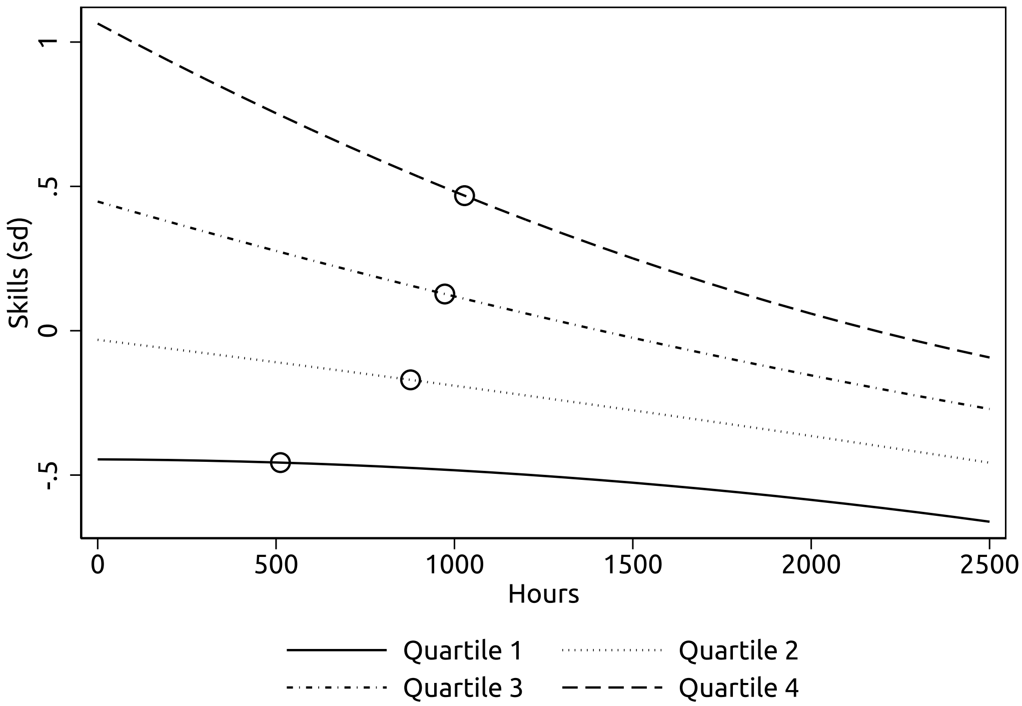
as the skill of the mother, the quality of her interactions with the child, and the overall quality of the child's environment. The plot shows that the average value of the confounder is comparable for mothers that work a similar number of hours. However, there is a discontinuity at zero hours: the average value of the confounder among the mothers who do not work is very different from the average among the mothers who work marginally positive hours. This happens because the mothers who do not work belong to two different groups: those with confounder values that led them to be exactly indifferent between not working and working, and those with confounder values that led them to be far from indifference, in favor of not working. Because the number of working hours cannot be negative, the first group of mothers choose an "interior solution," while the second group is constrained to choose a "corner solution". As the dashed line in the right plot suggests, the existence of the corner-solution group is what generates the discontinuity.

Summarizing the ideas above, (1) the non-negativity constraint leads confounders to be discontinuous at the bunching point, and (2) this discontinuity in confounders is the sole reason for the discontinuity in the outcome at the bunching point (i.e. the treatment did not cause that discontinuity). The ratio of these two discontinuities (i.e., the discontinuity in the outcome divided by the discontinuity in the confounder) can be interpreted as the effect of confounders at the bunching point. This effect can then be used to correct the endogeneity bias and identify the treatment effect. As this discussion suggests, our approach is somewhat related to "fuzzy" regression discontinuity designs, which also identify a causal effect with the ratio of two discontinuities (e.g. Hahn, Todd, and van der Klaauw 2001). This relationship is informally discussed in Appendix A.

Besides bunching, we need two assumptions in order to identify the effect of interest. First, because we do not directly observe the confounder, we need to make an assumption that allows us to obtain the size of the discontinuity in the right panel of Figure 1. Intuitively, this is equivalent to identifying how far from exact indifference between not working and working is the average mother who does not work. Second, because we can only identify the direct effect of confounders on the outcome at zero, we need to be able to extrapolate this effect to mothers who work positive hours. Intuitively, this is equivalent to assuming that the confounder enters linearly in the outcome equation. In practice, we show that both assumptions become much weaker as we add observed controls. Section 6 presents detailed sensitivity analyses of both assumptions and demonstrates that our key empirical conclusions do not depend on them.

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 2. 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

Figure 2: 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.

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 it, could in turn benefit all family members, including the child, in the long run. Thus, it would be valuable to investigate further whether it is feasible to design and implement work-promoting policies that mitigate 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, as higher-skill mothers disproportionately work longer hours. However, this potential explanation is ruled out by our finding that the effect is approximately linear, as shown in Figure 2. In fact, to the extent that we find some evidence of nonlinearity, Figure 2 indicates that the last hour might be particularly costly for children of mothers working longer hours only for mothers who have sufficiently low skills.

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 maternal 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 may 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 the time they spend 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. Nonetheless, 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's work on the child's cognitive skills in the short-run, even if the mother is also a high-earner.

These findings yield a number of important insights about the effects of maternal work during the first years of the children. First, using a new identification strategy leveraging different features of the data, we confirm what has typically been found in the 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-run 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-run unintended consequences for the children of higher-skill mothers: our results are consistent with the idea that it is safer to focus on policies that encourage greater flexibility in work arrangements for mothers, rather than focusing on policies that provide financial support to working mothers. 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.[5] Of course, it is difficult to draw direct implications for policy without a better understanding of the specific channels through which the causal effects we estimate are operating. More work is thus needed to assess the potential benefits of different policies. It is worth remarking that remote and hybrid work has recently expanded dramatically due to the necessity of social distancing during the COVID-19 pandemic, 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 impact of these changes on the skill of the children, which could complement the findings of our study.

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, while Section 6 provides a detailed sensitivity analysis and various assessments of our key identifying assumptions. Section 7 concludes. The appendix contains additional material that helps provide further context to our study.

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 support 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-

[5] 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.

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; and (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.

Remark 2.1. ***Censoring and Fixed Costs***
This paper is also related to the empirical literature estimating labor supply wage elasticities (see, e.g., Heckman, 1974; Gronau, 1974; Cogan, 1981; Mroz, 1987; Zabel, 1993). In addition to being concerned mainly with the estimation of different quantities, these papers differ from ours in at least two other important dimensions. First, they face a censoring problem: wages are only observed for those who are working. By contrast, we do not face a censoring problem because we observe childhood skills for everyone in our sample, including the children of mothers who work zero hours. Second, these papers often explicitly model both an intensive and an extensive margin of labor supply, and fixed costs are typically a relevant feature of the models they explore. The treatment variable used in

our paper and the whole related literature on child development discussed above is aggregated at the year level or higher. Following Cogan (1980) and Cogan (1981), we argue that, in this instance, a model with only variable costs approximates the distribution of the treatment variable well.[6] *Indeed, Figure 14 in Appendix C shows that we observe no "hole" around zero in the distribution of working hours, which is what would be expected if fixed costs were empirically important. In Section 6.2.2 (Remark 6.2), we consider a sensitivity analysis that allows for fixed costs, and we confirm in an independent way that fixed costs do not appear to be important in our context.*

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.[7] 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.[8] Throughout the analysis, we measure skills in standard-deviation (s.d.) units.

We measure our primary variable of interest, maternal labor supply, using the average number

[6] For instance, Cogan (1980) states, "It may be argued that the major sources of the costs of work are more properly treated as variable costs with respect to annual hours. This is true, especially if most of the variation in annual hours worked was the result of variations in days worked per year."

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

[8] 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.

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.[9] 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 the 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.[10]

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).[11]

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.[12]

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 reading recognition, and marginally below the average on math.

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

[10]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.

[11]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.

[12]For children born after 1994 in odd years, the survey was not conducted in their year of birth. In these cases, 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.

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 on 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 their 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 on average 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

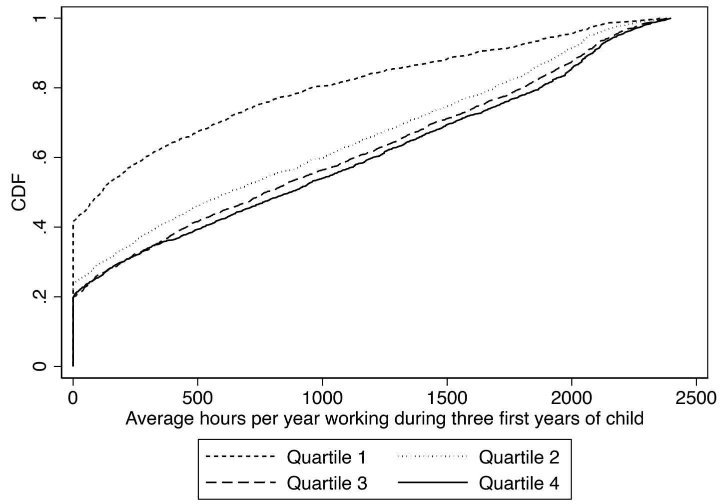
Here we detail the baseline list of controls used in the analysis. For the mother, we use variables meant to capture her human capital: AFQT, AFQT squared, and indicators for completed education at birth: less than high school, high school only, some college, college, and more than college. We also control 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. In some particular specifications, we also include as additional controls the pre-birth wage of the mother and the total pre-birth income of the mother's household from all sources other than her labor earnings.

4 Empirical Strategy

4.1 Bunching as a Corner Solution

Let L be our treatment variable, the average yearly working hours of the mother over the first three years of her child's life. Figure 3 shows L's cumulative distribution for each quartile of the maternal AFQT distribution. The distribution varies smoothly with L for $L > 0$, irrespective of the skill level of the mother. However, there is a notable bunching point at $L = 0$ for mothers of all skill levels. The degree of bunching at $L = 0$ tends to vary with the AFQT score of the mother, with lower-AFQT mothers bunching at higher rates, as expected.

Figure 3: 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.

Why is there such a concentration of mothers choosing exactly $L = 0$? To answer this question, it is helpful to conceptually separate the actual, realized number of working hours L from the *desired* number of working hours, which we denote L^*. The variable L^* results from a combination of factors, observed and unobserved, pertaining to the characteristics of the whole family which lead the mother to want to choose a given number of working hours. Thus, while L is the treatment variable, L^* is an index of the confounders that affect choice L, i.e. the "type" of the mother, for short.

The separation between L and L^* helps us understand the bunching in Figure 3. Mothers of type $L^* \geq 0$, choose exactly what their type leads them to choose: $L = L^*$. However, mothers of type $L^* < 0$ choose $L = 0$ because they cannot choose what their type would lead them to choose. Thus, there are two groups of mothers at $L = 0$: those who are exactly indifferent between working and not working ($L^* = 0$), and those who are away from exact indifference ($L^* < 0$). While mothers of type $L^* = 0$ choose $L = 0$ as an "interior solution", mothers of types $L^* < 0$ choose $L = 0$ as a "corner solution": the restriction that L cannot be negative is binding for them.

We can summarize the last two paragraphs with the following equation connecting L and L^*:

$$L = L^* \cdot \mathbf{1}(L^* \geq 0), \quad \text{with } \mathbb{P}(L^* < 0) > 0. \tag{1}$$

When $L > 0$, we have $L = L^*$. However, when $L = 0$, a break occurs between L and L^* for the observations of type $L^* < 0$.[13] As we show next, our identification approach uses this separation between L and L^* to identify the causal effect of interest.

[13] Figure 4 in Section 5.1 shows further evidence of the existence of mothers of type $L^* < 0$. See Footnote 23.

4.2 Control Function Approach

We want to estimate the effect of maternal working hours (L) on her child's skill (S). We write the outcome equation as

$$S = f(L, X; \boldsymbol{\beta}) + U, \tag{2}$$

where f is the treatment effect function of interest, with causal parameters $\boldsymbol{\beta}$. In its simplest form, $f(L, X; \boldsymbol{\beta}) = \beta L$ and β is the average effect of working hours on the child's skill across all mothers. However, we also consider interactions with some elements of the control vector X in order to report heterogeneous effects. U is the error representing all remaining factors that explain the outcome, which are likely correlated with L.

The next assumption imposes some structure on the relationship between U and L^*:

Assumption 1. *(Linearity in L^*) $U = m(X) + \delta(X)L^* + \epsilon$, where $m(X)$ and $\delta(X)$ are nonparametric functions and $\mathbb{E}[\epsilon|L^*, X] = 0$.*

Assumption 1 states that $\mathbb{E}[U|L^*, X] = m(X) + \delta(X)L^*$. Note that without loss of generality L^* is a sufficient index of all confounders of L because, by equation (1), L^* tracks any variation in L. Thus, Assumption 1 imposes a restriction only to the extent that it limits the functional form of how L^* affects the outcome. Specifically, while $\mathbb{E}[U|L^*, X]$ can be nonparametric in X, it must be multiplicative in L^*. This expression, which can be viewed as a random correlated coefficients model, is not structural; there need be no interpretation for $m(X)$ and $\delta(X)$. All that is required is that the nonparametric projection of U onto L^* and X be linear in L^*. In fact, the assumption can be stated equivalently as $\mathbb{E}[U|L^*, X] = a(X) + \delta(X)(L^* - \mathbb{E}[L^*|X])$, so only the part of L^* that is orthogonal to X must be multiplicative. We refer to Assumption 1 as the "linearity" assumption throughout the paper. Section 6.2 provides a detailed sensitivity analysis regarding this assumption.

Substituting the expression for U in Assumption 1 into equation (2), we obtain:

$$S = f(L, X; \boldsymbol{\beta}) + m(X) + \delta(X)L^* + \epsilon. \tag{3}$$

If we could observe L^*, then Assumption 1 would be enough to identify $\boldsymbol{\beta}$ directly in equation (3). We do not observe L^* when it is negative, though, so we need to first obtain a proxy of it. Specifically, note that we can write $L^* = L + L^*\mathbf{1}(L = 0)$, and therefore $\mathbb{E}[L^*|L, X] = L + \mathbb{E}[L^*|L = 0, X]\mathbf{1}(L = 0)$. We can then write

$$S = f(L, X; \boldsymbol{\beta}) + m(X) + \delta(X)[L + \mathbb{E}[L^*|L = 0, X]\mathbf{1}(L = 0)] + \varepsilon, \tag{4}$$

where $\varepsilon = \epsilon + (L^* - \mathbb{E}[L^*|L = 0, X])\mathbf{1}(X = 0)$, so the new error is simply the original error augmented by the proxy error. Equation (4) shows that, if we could identify $\mathbb{E}[L^*|L = 0, X]$, then we could add the proxy $L + \mathbb{E}[L^*|L = 0, X]\mathbf{1}(L = 0)$ to equation (3) as a control function in place of the unobservable L^*, allowing us to identify $\boldsymbol{\beta}$, as well as $m(\cdot)$ and $\delta(\cdot)$.[14]

[14] Note that $\mathbb{E}[\epsilon|L, X] = 0$ by the Law of Iterated Expectations ($\mathbb{E}[\epsilon|L, X] = \mathbb{E}[\mathbb{E}[\epsilon|L^*, X]|L, X] = 0$).

The term $\mathbb{E}[L^*|L=0,X]$ is the average type among mothers with observed covariates X who do not work. Since $P(L^*<0)>0$, we know that $\mathbb{E}[L^*|L=0,X]<0$. However, we do not know exactly how close to indifference (i.e., how close to $L^*=0$) is the average mother who does not work, yet this is precisely the quantity that is needed for identification.

Substantial progress can be made by considering all possible values of $\mathbb{E}[L^*|L=0,X]<0$. We can estimate $\boldsymbol{\beta}$ for a sequence of increasingly negative values in place of $\mathbb{E}[L^*|L=0,X]$, which gives us the full range of possible values of $\boldsymbol{\beta}$ that could be obtained under Assumption 1 alone. We implement this strategy in Section 6.1 to show that we can obtain all our qualitative findings without identifying $\mathbb{E}[L^*|L=0,X]$, that is, without making assumptions beyond Assumption 1.[15]

However, to keep the discussion of the results more concrete, we report point estimates of $\boldsymbol{\beta}$.. For this, we need to be able to pin down the value of $\mathbb{E}[L^*|L=0,X]$, which can be achieved by making assumptions about the shape of the distribution of $L^*|X$. Our headline results are obtained under the following assumption.

Assumption 2. *(Nonparametric Tail Symmetry) The tails of the distribution of $L^*|X$ below the $\mathbb{P}(L=0|X)$-th quantile and above the $(1-\mathbb{P}(L=0|X))$-th quantile are symmetric. Precisely, define $q(X)$ as the $(1-\mathbb{P}(L=0|X))$-th quantile of $L^*|X$. Then, for all $l\leq 0$,*

$$\mathbb{P}(L^*\leq l|X)=1-\mathbb{P}(L^*\geq q(X)-l|X).$$

This assumption requires that the lower tail of the distribution of $L^*|X$ be the mirror image of the upper tail of the distribution of L (since $L=L^*$ when $L>0$). We also consider nonparametric full symmetry. While this is a stronger condition than tail symmetry, it is testable.[16] We do not reject the full symmetry assumption in our data at standard levels of significance. Tail symmetry is weaker than full symmetry and is the weakest assumption we consider that is consistent with the data, which is why our headline estimates are obtained under Assumption 2. Nevertheless, we emphasize again that we focus our discussion only on findings that hold irrespective of the particular distributional assumption made (see Section 6.1).

Throughout the paper we also report results under the following stronger but more standard semiparametric assumptions.

Assumption 2′. *(Semiparametric Uniform)* $L^*|X\sim U[\kappa(X),\theta(X)]$.

Assumption 2″. *(Semiparametric Normal)* $L^*|X\sim\mathcal{N}(\mu(X),\sigma^2(X))$.

We consider these two alternative assumptions because they tend to yield very different estimates of $\mathbb{E}[L^*|L=0,X]$, further highlighting the robustness of the results to different distributional

[15] Implicit in the argument above is the rank condition that $f(L,X;\boldsymbol{\beta})$, $m(X)$ and the control function $L+\mathbb{E}[L^*|L=0,X]\mathbf{1}(X=0)$ are not perfectly multicollinear. The perfect multicollinearity of the control function and f is ruled out by the discontinuity of the control function at $L=0$ (because $\mathbb{P}(L^*<0)>0$) and the specific functional form of $f(L,X;\boldsymbol{\beta})$. For example, this is guaranteed if f is continuous in L at $L=0$, as in our case.

[16] To see how this assumption is testable, consider an example where $\mathbb{P}(L=0|X)=0.3$, so there is bunching of 30% of the observations at $L=0$. Then we can compare the empirical distribution of $L|X$ between percentiles 30 and 50 with the mirror image of the empirical distribution between percentiles 50 and 70.

assumptions. Intuitively, normality assumes a "long tail" of mothers who are far from indifference between working and not working. By contrast, the uniform distribution assumes that all types of bunched mothers are equally represented and that none have L^*'s too far from indifference. Consequently, our estimates of $\mathbb{E}[L^*|L=0,X]$ are more negative under the Semiparametric Normal assumption than under the Semiparametric Uniform assumption.

These three distributional assumptions yield simple expressions for $\mathbb{E}[L^*|L=0,X]$. For instance, under Assumption 2 (Nonparametric Tail Symmetry),

$$\mathbb{E}[L^*|L=0,X] = q(X) - \mathbb{E}[L|L \geq q(X), X]. \tag{5}$$

If $\mathbb{P}(L=0|X) \leq 0.5$, then we can easily obtain $q(X)$, the $1-\mathbb{P}(L=0|X)$-th quantile of the distribution of $L|X$. The control function in this case is $L + \mathbb{E}[L^*|L=0,X]\mathbf{1}(L=0)$ where $\mathbb{E}[L^*|L=0,X]$ is described in equation (5).[17]

Remark 4.1. ***Understanding the Role of Controls in the Linearity Assumption***

Although our approach can in principle be implemented without controls, their inclusion helps weaken the linearity assumption (Assumption 1). Recall that the linearity assumption is equivalent to $\mathbb{E}[U|L^*,X] = a(X) + \delta(X)(L^* - \mathbb{E}[L^*|X])$. Therefore, when we use controls, only the component of L^* that is orthogonal to the controls needs to have a linear effect on the outcome S. In other words, we still allow L^* to have an arbitrary nonlinear effect on S, as long as it is through some function of the controls X. Therefore, the richer the control set, the smaller the part of L^* that is constrained by the linear requirement.

Figure 5 in Section 5 shows direct evidence that controls weaken the linearity assumption. The left panel, which shows how S changes with L unconditionally, is clearly more nonlinear than the right panel, which shows how the residual of S changes with L after partialling out $m(X)$.

Remark 4.2. ***Understanding the Role of Controls in the Distribution Assumption***

The distributional assumption on L^* may likewise be relaxed with the use of controls. The tail symmetry requirement in Assumption 2, for example, is made on the conditional distribution of L^* given X, and so L^* can be affected by X in arbitrary ways. Indeed, the unconditional distribution of L^* may even be tail asymmetric provided that the residual variation in L^*, after accounting for all variation that is explained by X, is tail symmetric. To see this, note that without loss of generality we can write $L^* = \mathbb{E}[L^*|X] + \sigma(L^*|X)\eta$, where $\sigma(L^*|X)$ is the standard deviation of $L^*|X$ and $\eta = (L^* - \mathbb{E}[L^*|X])/\sigma(L^*|X)$. Assumption 2 therefore requires that η, which is the standardized variation in L^* not explained by X, is tail symmetric. This allows the unconditional distribution of L^* to be tail asymmetric due to any of the following reasons: (1) X has a tail asymmetric

[17] Under Assumption $2'$, $L|X$ is a truncated uniform distribution, and thus the expectation can be obtained as $\mathbb{E}[L^*|L=0,X] = -\mathbb{E}[L|L>0,X] \cdot \mathbb{P}(X=0|X)/(1-\mathbb{P}(L=0|X))$. Under Assumption $2''$, $L|X$ is a truncated normal distribution, and thus the expectation can be obtained with a standard Tobit model as $\mathbb{E}[L^*|L=0,X] = \mu(X) - \sigma(X)\lambda(-\mu(X)/\sigma(X))$, where λ is the inverse Mills ratio (i.e., the ratio of the standard normal PDF and the standard normal CDF). Under both Assumptions $2'$ and $2''$, the control function is $L + \mathbb{E}[L^*|L=0,X]\mathbf{1}(L=0)$ with the appropriate expression for $\mathbb{E}[L^*|L=0,X]$ substituted.

distribution; (2) the part of L^* predicted by X is tail asymmetric; and (3) the conditional standard deviation weighs observations in an unbalanced way, creating a tail asymmetry in L^*.

4.3 Estimation Details

Section 3 details our list of controls X, which contains both continuous and discrete variables. As discussed in Caetano et al. (2023), in order to maintain the predominantly nonparametric structure of the model, it is a good idea to "discretize" X before estimating the expectation $\mathbb{E}[L^*|L=0,X]$. Specifically, let $\{\hat{\mathcal{C}}_1,\ldots,\hat{\mathcal{C}}_K\}$ be a finite partition of the support of X into K 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 K-*th* dimensional vector of cluster indicators. In the estimation of the expectation, we substitute X with $\hat{C}_K$. 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 X is first discretized and then one of the distributional assumptions is applied separately for each cluster.

For example, in the Nonparametric Tail Symmetry case (Assumption 2), separately for each cluster k, we estimate the probability of bunching $\hat{p}_k=\hat{P}(L=0|X\in\hat{\mathcal{C}}_k)$ by calculating the frequency of mothers who do not work among all mothers in cluster k. Then, we calculate the $(1-\hat{p}_k)$-th quantile of L among the observations in cluster k, $\hat{q}_k$. Finally, $\hat{\mathbb{E}}[L^*|L=0,X\in\hat{\mathcal{C}}_k]=\hat{q}_k-\hat{\mathbb{E}}[L|L\geq\hat{q}_k,X\in\hat{\mathcal{C}}_k]$, where $\hat{\mathbb{E}}[L|L\geq\hat{q}_k,X\in\hat{\mathcal{C}}_k]$ is simply the average number of work hours among mothers in cluster k who work $\hat{q}_k$ hours or longer. This method is implemented for all clusters k such that $\hat{p}_k\leq 0.5$. If any cluster has $\hat{p}_k>0.5$, then Assumption $2''$ is made for that cluster, and the expectation for that cluster is calculated accordingly.[18]

The clusters are not arbitrary – rather, we employ a clustering algorithm in which two observations will be clustered together if they have similar Xs. As K grows, the observations in each cluster have increasingly similar values of X as the within-cluster variation in these controls decreases.[19] Thus, if $\mathbb{E}[L^*|L=0,X]$ is continuous, $\hat{\mathbb{E}}[L^*|L=0,X\in\hat{\mathcal{C}}_k]$ will approximate $\mathbb{E}[L^*|L=0,X]$ more and more closely as K grows.[20]

Following a similar logic, we also use the same clusters to make sure our specification of the function $m(X)$ in equation (3) approximates a nonparametric function of X. Specifically, we specify $m(X)=X'\tau+\sum_{k=1}^{K}\alpha_k\mathbf{1}(X\in\hat{\mathcal{C}}_k)$, so the cluster indicators control nonparametrically for differences across clusters, while the within-cluster differences due to X are controlled linearly. As the number of clusters increases, the nonparametric match improves, leaving progressively less unexplained variation within each cluster. In Section 6.2.2, we also consider a similar specification of $\delta(X)$ with K_δ clusters, thus allowing δ to vary nonparametrically with X.

[18]Estimation of the expectation under other assumptions is analogous. In the Semiparametric Uniform case (Assumption $2'$), separately for each cluster k, we use the estimator $\hat{\mathbb{E}}[L^*|L=0,X\in\hat{\mathcal{C}}_k]=-\hat{\mathbb{E}}[L|L>0,X\in\hat{\mathcal{C}}_k]\hat{p}_k/(1-\hat{p}_k)$, where $\hat{\mathbb{E}}[L|L>0,X\in\hat{\mathcal{C}}_k]$ is the average number of work hours among mothers in cluster k who work a positive amount. In the Semiparametric Normal case (Assumption $2''$), separately for each cluster k, we run a Tobit regression of L on a constant. Letting $\hat{\mu}_k$ and $\hat{\sigma}_k$ be the estimated coefficient (intercept) and standard deviation from this Tobit model, $\hat{\mathbb{E}}[L^*|L=0,X\in\hat{\mathcal{C}}_k]=-\hat{\mu}_k-\hat{\sigma}_k\lambda\left(-\hat{\mu}_k/\hat{\sigma}_k\right)$, where $\lambda(\cdot)$ is the inverse Mills ratio.

[19]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 (Hastie, Tibshirani, and Friedman 2009).

[20]See Caetano et al. (2023) for a more formal discussion of this approximation.

Summarizing, our estimation approach consists of the following steps. First, we create K clusters using our vector of pre-determined controls X. Next, we estimate $\mathbb{E}[L^*|L=0, X \in \hat{\mathcal{C}}_k]$ separately for each $k = 1, \ldots, K$ as explained above. Then, we construct the control function $(L + \hat{\mathbb{E}}[L^*|L = 0, \hat{C}_K]\mathbf{1}(L=0))$. Finally, we estimate our model of interest specifying $m(X)$ as explained above and including this control function as an additional control. For example, in our homogeneous model, in which $f(L, X; \boldsymbol{\beta}) = \beta L$ and $\delta(X) = \delta$, we estimate via OLS the following regression:[21]

$$\mathbb{E}[S|L, X] = \beta L + X'\tau + \sum_{k=1}^{K} \alpha_k \mathbf{1}(X \in \hat{\mathcal{C}}_k) + \delta[L + \hat{\mathbb{E}}[L^*|L=0, \hat{C}_K]\mathbf{1}(L=0)]. \tag{6}$$

5 Results

In this section, we present our estimates of $\boldsymbol{\beta}$ for different specifications of the function $f(L, X; \boldsymbol{\beta})$. In order to facilitate the interpretation of our results, we also specify $\delta(X)$ to be simple functions of the controls X, as these functions clarify how selection varies with some key observables. We show in Section 6.2.2 that our estimates of $\boldsymbol{\beta}$ are robust to specifications that allow $\delta(X)$ to depend nonparametrically on X.

5.1 Homogeneous Treatment Effects

We start from the most parsimonious specification of $f(L, X; \boldsymbol{\beta})$,

$$f(L, X; \boldsymbol{\beta}) = \beta L. \tag{7}$$

This specification of $f(L, X; \boldsymbol{\beta})$ is comparable to the specifications in the prior literature.

Table 2 presents the $\hat{\beta}$ 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 the estimates using our control function approach to correct for endogeneity, as outlined in Section 4. Each column differs only in the assumption made on the distribution of $L^*|X$. Column (iii) supposes that $L^*|X$ is normally distributed, with mean and variance that depend on the cluster to which X belongs (Assumption $2''$). Column (iv) supposes that $L^*|X$ is uniformly distributed with the upper and lower limits of the support depending on the cluster of X (Assumption $2'$). Finally, column (v) supposes that $L^*|X$ is symmetric in the tails, so that the unobserved portion of $L^*|X$ below 0 is the mirror image of the corresponding tail above 0, conditional on the cluster of X (Assumption 2). As discussed in Section 4, the latter are our preferred point estimates, as they rely on our weakest distributional assumption.

[21] Throughout the paper, we conduct inference via the bootstrap, which is proved in Caetano et al. (2023).

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

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Semip. Normal	(iv) Semip. Uniform	(v) Nonp. Tail 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 100 additional 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). This is the list of controls X: mother's AFQT and AFQT squared, mother's completed education at birth (<HS, HS, some college, college, >college), mother's age (≤ 19, $\in [20, 24]$, $\in [25, 29]$, $\in [30, 34]$, or ≥ 35), indicators for whether the mother's spouse is present in the household and the spouse's education at birth, ln(total family size), child's sex, race, birth order, Census geographic region, age at assessment date, and indicators for the year of the assessment. We specify $m(X)$ to enter both linearly and as indicators of each cluster of X, with a total of $K = 50$ clusters. $\mathbb{E}[L^*|L = 0, X]$ is also estimated with $K = 50$ clusters, and $\delta(X)$ is estimated with $K_\delta = 1$ cluster. See Figure 12 and Table 5 for the analogous results for different values of K and K_δ, respectively. ** 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 an additional 100 hours of maternal labor supply annually over the first three years of a child's life lowers the child's cognitive skills around age 6 by 0.019 standard deviations (s.d) on average. This effect is economically large given the sample variance in maternal labor supply – a one s.d. increase in maternal labor supply over the first three years of life (an increase in 838 hours per year, see Table 1) would translate to reductions in cognitive skills of about 0.16 s.d. The estimates assuming normal or uniform distributions for $L^*|X$ are quite similar to the symmetric case.

Table 2 also shows the estimates of δ, which is the average effect of the confounder L^* on the outcome Y. These estimates are positive and significant: mothers who work more hours tend to be positively selected relative to those who work fewer hours.

Selection on Observables vs. Selection on Unobservables

To provide more context to our findings, it is useful to note what happens to the estimates of β in Table 2 as controls are sequentially added to the regression. The estimate of β, which was large and positive without controls (column (i)), goes down to zero when the observable controls are added (column (ii)). Our negative results in columns (iii)-(v) simply show that, when we further add the control function (to control for unobservable confounders), the estimate of β goes down further, indicating that selection on unobservables tends to be in the same direction as selection on observables. In Appendix B, we formalize this idea by implementing the method developed in Oster (2019) to show that our main estimate from column (v) is consistent with the degree of selection on unobservables being smaller (but in the same direction) than the degree of selection on observables, which is generally considered plausible (Altonji, Elder, and Taber 2005). Importantly, Oster (2019)'s method relies on completely different assumptions and does not use bunching of the treatment variable in any way. We thus view these results as independent evidence that our findings are plausible and not merely an artifact of our method.

Next, we present independent evidence of positive selection on both observables and unobservables based on Caetano (2015)'s discontinuity test, which does not rely on Assumption 1 or Assumptions 2, 2′ or 2″.

Figure 4 provides evidence of positive selection on observables. It shows local linear regression fits of key covariates on L, estimated on the $L > 0$ sample.[22] We also show the average values of the covariates for the subsample of mothers bunched at $L = 0$. The left panel shows a clear positive relationship between maternal AFQT and L away from zero, with mothers who work similar positive hours tending to have similar AFQT scores. However, mothers who work exactly zero hours tend to have sharply lower AFQT scores than those who work a small, positive number of hours. The right panel shows a similar pattern for another key covariate: mothers at $L = 0$ are discontinuously less likely to have a spouse present in the year the child was born than mothers with marginally positive hours $L = l > 0$. We find positive discontinuities for many other observed covariates, suggesting positive selection on observables at $L = 0$. That is, mothers who work a small, positive number of hours per year tend to have discontinuously higher levels of observed covariates that are themselves positively correlated with children's cognitive skills, relative to those who work zero hours.[23]

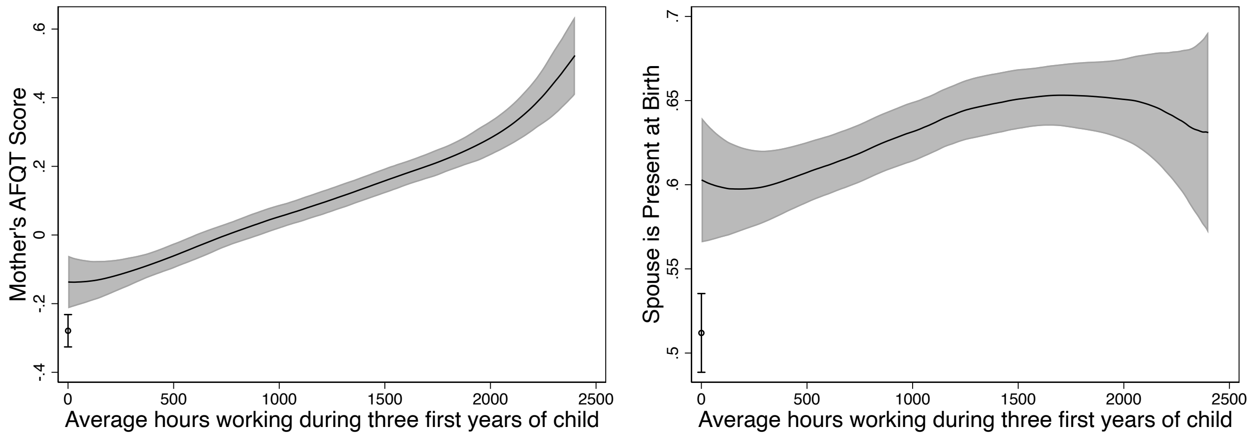
Figure 5 provides evidence of positive selection on unobservables. The left panel is analogous to those in Figure 4, but for the outcome variable (childhood cognitive skills S) in the vertical axis instead of a key covariate. A significant positive discontinuity is evident, confirming the result above that the effect of L^* on S is positive: i.e., on balance, the variables (observed or unobserved) that are positively correlated with L near zero tend to be positively correlated with S. The right panel in this figure shows that even after all controls X are held fixed, the child's skill S remains discontinuously higher as L increases from $L = 0$, suggesting that there are unobservable variables that are orthogonal to X which are themselves adding a positive bias to the results from column (ii) of Table 2.[24]

[22]We choose a bandwidth of 400 hours because standard approaches to obtain the optimal bandwidth, such as the rule-of-thumb (ROT), yield values that gravitate around 400. Versions of this figure with smaller bandwidths suggest similar discontinuities. We also conduct t-tests to confirm that there are no discontinuities anywhere other than at $L = 0$. For each of the covariates, we reject at 5% that the average at zero hours is equal to the average in the bin (0,20] hours. By contrast, comparing the bins (0,20] and (20,40], we never reject equality at 10%.

[23]Figure 4 provides further evidence that mothers of type $L^* < 0$ must exist in our sample, beyond the bunching of observations at $L = 0$. To see this, note that if types $L^* < 0$ did not exist, then all mothers at $L = 0$ would be of type $L^* = 0$, and on average this type should have plausibly similar characteristics to a type $L^* = l > 0$ for small l. That is, those exactly indifferent between working and not working ($L^* = 0$) and those nearly indifferent between working and not working ($L^* = l > 0$ for small l) should be on average similar to each other. In contrast, these discontinuities show that those at $L = 0$ are on average sharply different from those at $L = l > 0$ for small l, which suggest that types $L^* < 0$ must be also present in the sample of mothers choosing $L = 0$.

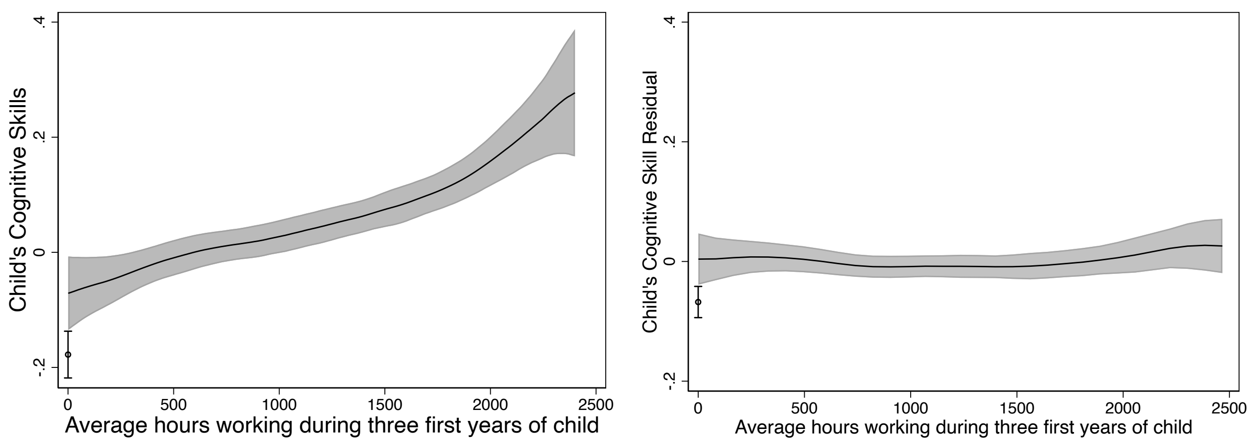
[24]As shown in Caetano (2015), in order to implement these discontinuity plots and infer that the selection on unobservables is positive, we need to make two assumptions similar to the ones made in regression discontinuity designs (RDDs): (a) L is a continuously distributed variable for $L > 0$, so we can arrive arbitrarily close to $L = 0$ from positive values of L; and (b) treatment effects are continuous at $L = 0$. These two assumptions are plausible in our context. For (a), note from Figure 3 that the slope of the CDF (which is the probability density function, PDF) is positive for small positive values of L around zero; indeed we observe several mothers working just a few hours per

Figure 4: Evidence of Positive Selection on Observables at $L = 0$



Note: This figure shows the local linear regression of two key observed covariates 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.

Figure 5: Evidence of Positive Selection on Unobservables at $L = 0$



Note: This figure is analogous to Figure 4, but instead of covariates the variables in the vertical axis represent the outcome variable S (children's cognitive skills) in the left panel, and the "residualized" outcome S in the right panel. Covariates X enter nonparametrically in the right panel, exactly as in Table 2.

Alternative Specifications

Table 7 in Appendix C shows that the results in Table 2 are robust to using alternative samples and controls. The first panel of the table reproduces the findings from Table 2. In the second panel, we restrict the sample to only include mothers who are married at the time of the birth. In the third panel, we restrict the sample to only include the unmarried mothers. In the fourth panel, we restrict

year on average during the first three years of the child's life, which can be directly seen in the histogram of Figure 14 in Appendix C, which zooms into the distribution of L near $L = 0$. For (b), it is plausible that mothers working on average just a few hours per year during the first three years of the child's life should not generate a large causal effect on the cognitive skills of the child at age 6, relative to those working zero hours per year. Thus, discontinuities in $\mathbb{E}[S|L = l]$ as l approaches 0 from the positive side cannot generate any discontinuity because of its causal effect, so it must be because selection varies discontinuously. See Appendix A for further details.

the sample to only include mothers who do not have a college degree. Finally, in the last panel we add as a control the total pre-birth income of the mother's household from all sources other than her labor earnings. Although the estimates change slightly for these different specifications, the key findings from Table 2 stand.

5.2 Heterogeneity by Maternal AFQT and Labor Supply

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

$$f(L, X; \boldsymbol{\beta}) = (\beta + \beta_A A + \beta_L L + \beta_{AL} AL)L. \tag{8}$$

Column (v) in Table 3 presents the results under our preferred nonparametric tail symmetry assumption. The results under alternative distributional assumptions (normality and uniformity) are very similar.

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) Semip. Normal	(iv) Semip. Uniform	(v) Nonp. Tail 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 100 additional 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). We specify $m(X)$ to enter both linearly and as indicators of each cluster of X, with a total of $K = 50$ clusters. $\mathbb{E}[L^*|L = 0, X]$ is also estimated with $K = 50$ clusters, and $\delta(X)$ is estimated with $K_\delta = 1$ cluster. See Figure 15 and Table 8 in Appendix C for the analogous results for different values of K and K_δ, respectively. ** p<0.05, * p<0.1.

Table 3 reveals a few interesting patterns. First, $\beta < 0$, so the estimated effect on cognitive skills of an additional 100 working hours for a mother with average skills (i.e. $A = 0$) is negative. Second,

$\beta_A < 0$, so mothers with a higher AFQT skill tend to have a more negative effect of working hours on children skills. This effect is statistically significant and quite large: each additional standard deviation of mother's AFQT lowers her estimated effect by 0.026 s.d. Third, β_L is positive, but small and insignificant, so the return of working hours for the average AFQT skill mother is approximately constant in hours.[25] Fourth, $\beta_{AL} > 0$, so there is some evidence that the return of working hours increases with hours for mothers with higher AFQT skill, but the degree of convexity is small. These results are illustrated in Figure 2 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 with higher skills tend to work longer hours and also tend to have children with higher skills. As their mothers work longer hours, the children's skills go down, but more quickly for the children with higher skilled mothers.[26] 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 compared to -0.019 in the homogeneous case.

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

5.3 Further Treatment Effect Heterogeneity by Pre-Birth Wages

The results so far suggest a misalignment between the work incentives mothers face and the short-term benefits of this work 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 of 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 in terms of the foregone interactions with their child. In order to investigate this hypothesis, we expand equation (8) by including a third dimension of heterogeneity: the pre-birth wage rate of the mother, W:[27]

$$f(L, X; \boldsymbol{\beta}) = (\beta + \beta_A A + \beta_W W + \beta_L L + \beta_{AL} AL + \beta_{WL} WL)L. \tag{9}$$

We allow the slope and convexity of the effect of hours to depend on maternal skills and wage

[25] Note that L^2 is part of the terms multiplying both β_L and β_{AL}. Since L^2 is a very large number for most working mothers, the estimates of these coefficients are much smaller than the other estimates. For readability, we present these estimates multiplied by 1,000.

[26] For some children in our sample whose mothers have sufficiently low skills, Table 3 actually implies positive effects. In fact, 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.

[27] Specifically, W denotes a residualized wage measure in which the effect of the age of the 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 comparable to the estimates of β_A and β_{AL}.

rates in a separable way.[28] 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 gains to skills flowing from the additional money received by the mother due to this work.[29] We assume that the heterogeneous effects via A (AFQT skills) holding W (pre-birth wages) constant tend to incorporate mostly the loss of interaction between the mother and the child, while the heterogeneous effects via W holding A constant 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 maternal 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 the life of the child.

Table 4 presents the estimates of the function $f(\cdot)$ as specified in equation (9). For this table, we restrict the sample to women who were working prior to giving birth, and include the pre-birth wage rate and its interaction with AFQT score as additional controls. The results from the more restricted functions $f(L, X; \boldsymbol{\beta})$ estimated in Tables 2 and 3 are similar when we impose this sample restriction and add these controls. 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 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 some evidence of stronger nonlinearity by mother's skill (β_{AL}), and there is no evidence of nonlinearity by wages (β_{WL}). 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. Nonetheless, it seems that there is much less heterogeneity across different values of W than across different values of A.[30] To see this, compare the values of the point estimates of β_W and β_A in Table 4 in the context of the joint distribution of A and W in the sample. Note that a mother who is 1 s.d. above the average in terms of A tends to be only one

[28] We also experimented with allowing for further heterogeneity by also including the interaction terms AWL and AWL^2. Including these additional interaction terms barely change the main estimates, while the estimates of these terms themselves are imprecisely estimated.

[29] 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.

[30] We also estimate specifications analogous to Table 3 but with pre-birth wages taking the place of AFQT, also obtaining little evidence for heterogeneous effects by the mother's wage even when AFQT skills are not explicitly specified in the function $f(L, X; \boldsymbol{\beta})$.

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) Semip. Normal	(iv) Semip. Uniform	(v) Nonp. Tail 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 100 additional 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 mentioned in Tables 2 and 3, these models also include pre-birth wages (W) and the interaction of pre-birth wages and maternal AFQT ($W \cdot A$) as elements of X. We specify $m(X)$ to enter both linearly and as indicators of each cluster of X, with a total of $K = 50$ clusters. $\mathbb{E}[L^*|L = 0, X]$ is also estimated with $K = 50$ clusters, and $\delta(X)$ is estimated with $K_\delta = 1$ cluster. See Figure 16 and Table 9 in Appendix C for the analogous results for different values of K and K_δ, respectively. ** p<0.05, * p<0.1.

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), we expect that only $0.018/4 = 0.0045$ s.d. would 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 that the selection is somewhat positive, and is disproportionately positive 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.

6 Sensitivity and Robustness

In this section, we show that our key findings in Section 5 are robust to violations of the identifying assumptions. Recall that we make two identifying assumptions: a linearity assumption (Assumption 1), and a distributional assumption on $L^*|X$ (Assumption 2, 2′ or 2″). We consider violations of each of these assumptions in turn, and jointly.

6.1 The Distributional Assumption

Here we show that the key findings from the point estimates reported in Section 5 are robust to the failure of the distributional assumption (i.e. Assumption 2″, for the results in column (iii), Assumption 2′, for the results in column (iv), and Assumption 2, for our headline results in column (v)). We consider how our estimates of $\boldsymbol{\beta}$ vary when the estimator of $\mathbb{E}[L^*|L = 0, X]$ is biased, which is what could happen if the distributional assumption does not hold.

We begin with the homogeneous treatment effect results from Section 5.1. Let $\tilde{\mathbb{E}}$ represent the value of the expectation identified under a specific distributional assumption, and $\tilde{\beta}$ the corresponding treatment effect identified using our control function approach. If the distributional assumption does not hold, $\tilde{\mathbb{E}}$ may be different from $\mathbb{E}$, and therefore $\tilde{\beta}$ may be different from the true treatment effect β. Caetano et al. (2023) show that the mistake in the identification of β ($\mathbb{B}_\beta = \tilde{\beta} - \beta$) can be written as a function of the mistake in the identification of the expectation ($\mathbb{B}_{\mathbb{E}} = \tilde{\mathbb{E}} - \mathbb{E}$), given by[31]

$$\mathbb{B}_\beta = -\frac{\mathbb{B}_{\mathbb{E}}}{\tilde{\mathbb{E}}} \cdot \delta. \tag{10}$$

This formula reveals an asymmetry: for a given value of δ, and for a given magnitude of the mistake in the expectation, $|\mathbb{B}_{\mathbb{E}}|$, it is preferable to err towards a $\tilde{\mathbb{E}}$ that is larger in magnitude than $\mathbb{E}$. This means that, all else the same, it is generally less consequential for our estimator of β if we err towards our estimator of $\mathbb{E}[L^*|L = 0, X]$ being too negative rather than not negative enough.

This asymmetry is confirmed in Figure 6. The thick black curve shows the estimates of β we would have obtained using every possible mistaken value for the expectation.[32] Specifically, the horizontal line shows the value of $\tilde{\mathbb{E}}$, and the thick black curve shows the resulting $\tilde{\beta}$ estimated using the model $Y = \tilde{\beta}L + \tilde{m}(X) + \tilde{\delta}(L + \tilde{\mathbb{E}} \cdot \mathbf{1}(X = 0)) + \varepsilon$. We considered also cases where the plugin $\tilde{\mathbb{E}}$ is allowed to vary with X. These cases are more cumbersome to report, and conclusions were similar.[33] The vertical lines are added for reference, and are positioned at the value $\hat{\mathbb{E}}[L^*|L = 0] = \frac{1}{n}\sum_{i=1}^{n} \hat{E}[L^*|L = 0, X = X_i]$, where $\hat{E}[L^*|L = 0, X = X_i]$ is estimated as described in Section

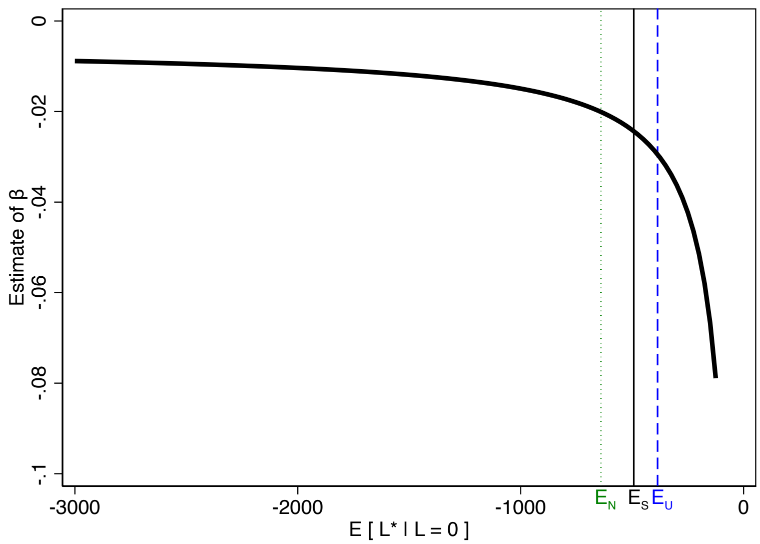
[31] We show here the formula in the case without controls. The formula in the general case is more cumbersome, and can be seen in Remark 2.3 in that paper. Since the controls absorb some of the endogeneity that is leftover by the use of a mistaken control function $L + \tilde{\mathbb{E}}[L^*|L = 0, X]\mathbf{1}(L = 0)$, the magnitude of $\tilde{\beta} - \beta$ tends to be smaller when controls are used than when controls are not used. Therefore, the formula we show represents a worst case scenario for the failure of the distributional assumption.

[32] We restrict the range of $\tilde{\mathbb{E}}$ in this plot purely for formatting reasons, but we confirmed that the thick curve behaves as one would expect based on the plot for extremely large values of $\tilde{\mathbb{E}}$ and for extremely low values of $\tilde{\mathbb{E}}$.

[33] It is easy to see that this simplification should not be consequential to our conclusions, since our estimates of β for $K = 1$ cluster (i.e. when no controls are used in the estimation of the expectation) are very similar to our estimates of β for $K = 50$ clusters. See Figure 12 in Section 6.3.

4.3 using 50 clusters. The estimates $\hat{E}[L^*|L = 0, X = X_i]$ used for the green dotted line assume normality (Assumption 2″), as do our estimates in column (iii). Analogously, the blue dashed line and the solid black line correspond to the uniform (Assumption 2′) and tail symmetry (Assumption 2) assumptions, as do our estimates in columns (iv) and (v), respectively.[34]

Figure 6: Model from Section 5.1: $\hat{\beta}$ for each counterfactual value of $\tilde{\mathbb{E}}[L^*|L = 0]$



Note: The thick black curve shows what would be $\hat{\beta}$ obtained from regression equation (4) in the homogeneous case of Section 5.1 for different counterfactual values of $\tilde{\mathbb{E}}[L^*|L = 0]$. The vertical lines represent the weighted average of the estimates of $\mathbb{E}[L^*|L = 0, X]$ across all $K = 50$ clusters, obtained from Assumptions 2 (solid black line), 2′ (dashed blue line), and 2″ (dotted green line). $N = 6,924$.

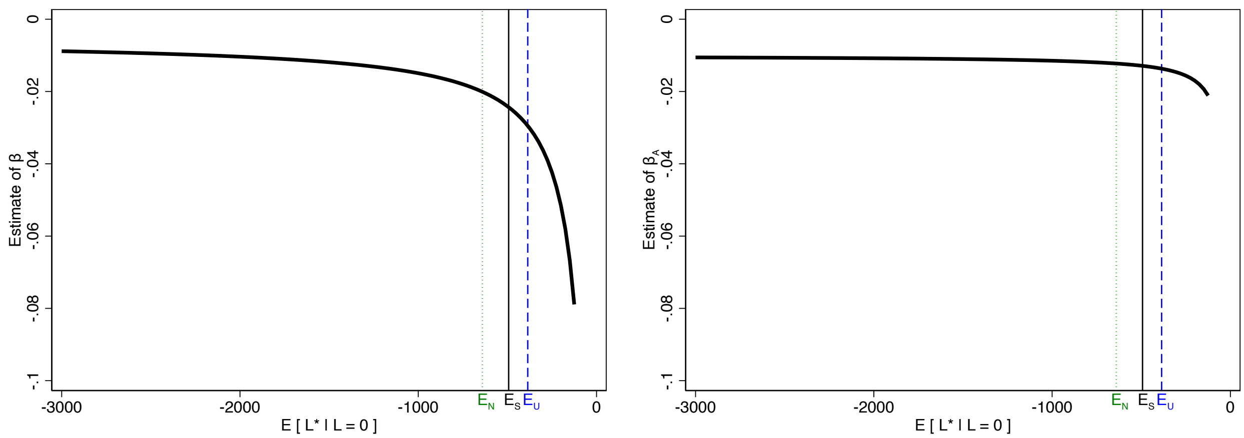
The two main takeaways from Figure 6 are as follows. First, the qualitative finding that $\beta < 0$ does not depend on the value of $\tilde{\mathbb{E}}$, and thus does not depend on the distributional assumption we make. Second, there is more scope for our estimates from columns (iii)-(v) from Table 2 to be underestimating the magnitude of β than overestimating it. Indeed, if we reported results using more negative estimates of $\mathbb{E}[L^*|L = 0, X]$ than the ones we obtained under our distributional assumptions, then the estimate of β would have changed little. However, had we used instead less negative estimates of $\mathbb{E}[L^*|L = 0, X]$ than ours, then the estimate of β would have been more negative than our reported results, and potentially substantially more so. The conclusion is therefore that the qualitative findings from Section 5.1 would be the same for any distributional assumption we could have made, and the quantitative value of $\hat{\beta}$ would have gone from slightly less negative to substantially more negative than the values we reported.

Figure 7 allows us to draw analogous conclusions for the main parameters of the heterogeneous treatment effect model from Section 5.2. It is clear that our key qualitative conclusions that $\beta < 0$ and that $\beta_A < 0$ do not depend on the estimated value of the expectation. Moreover, note that our

[34] The small discrepancy between the values of $\tilde{\beta}$ implied by the crossing of the thick black line with the vertical lines and the estimates in Table 2 is due to the fact that the crossing value corresponds to the estimate using the control function $L+\hat{E}[L^*|L = 0]\mathbf{1}(L = 0)$, while our estimates in Table 2 use the control function $L+\hat{E}[L^*|L = 0, X]\mathbf{1}(L = 0)$. The difference reflects the small impact on the estimates of allowing the expectation to change with X, as discussed in Section 6.3.

quantitative conclusions about β_A depend less on the estimated expectation than our quantitative conclusions about β.

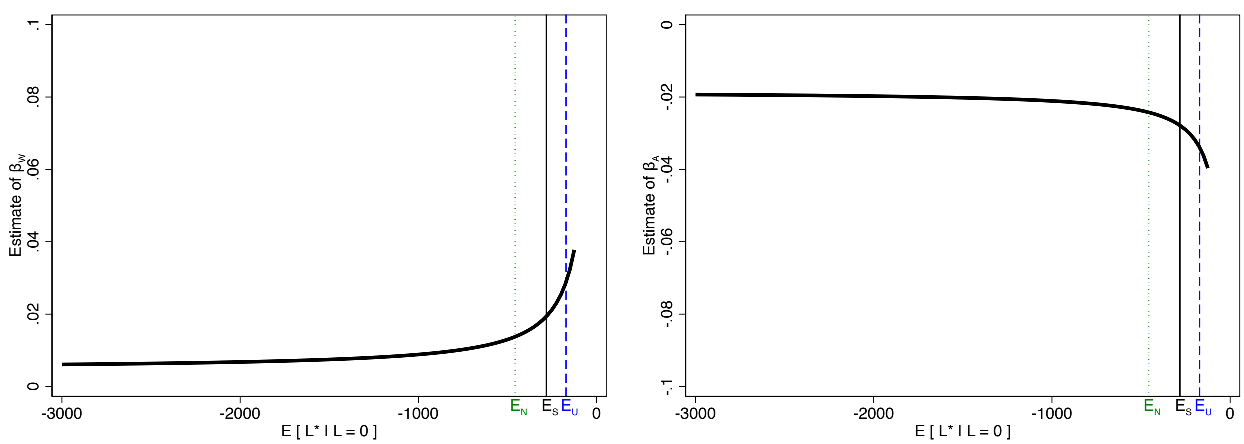
Figure 7: Model from Section 5.2: $\hat{\beta}$ and $\hat{\beta}_A$ for each counterfactual value of $\tilde{\mathbb{E}}[L^*|L=0]$



Note: This figure is analogous to Figure 6, but for β and β_A from Section 5.2 rather than β from Section 5.1. Note that the value of $\tilde{\mathbb{E}}[L^*|L=0]$, whatever it may be, must be the same in both panels of the figure. $N = 6,924$.

Finally, Figure 8 shows analogous results for the main parameters of the heterogeneous treatment effects model from Section 5.3.[35] It is clear that our key qualitative conclusions that $\beta_W \geq 0$, $\beta_A < 0$ and that $|\beta_A| > \frac{1}{4}|\beta_W|$ again do not depend on the estimated value of the expectation.[36] Moreover, our estimates of β_W and β_A have a similar degree of sensitivity to different values of $\tilde{\mathbb{E}}$, which is much smaller than the degree of sensitivity of β.

Figure 8: Model from Section 5.3: $\hat{\beta}_W$ and $\hat{\beta}_A$ for each counterfactual value of $\tilde{\mathbb{E}}[L^*|L=0]$



Note: This figure is analogous to Figure 7, but for β_W and β_A from Section 5.3 rather than β and β_A from Section 5.2. Note that the value of $\tilde{\mathbb{E}}[L^*|L=0]$, whatever it may be, must be the same in both panels of the figure. $N = 3,994$.

[35] Note that the vertical lines in Figure 8 are closer to zero, relative to the vertical lines from Figures 6 and 7. This is intuitive, as Figure 8 restricts the sample to mothers who worked prior to the birth of the child. The full sample of mothers used in Figures 6 and 7 include those who have never worked, and thus might be farther away from indifference between working and not working (i.e., more negative L^*).

[36] Recall the discussion in Section 5.3 that a mother who is 1 s.d. above the average in A tends to have only a 1/4 s.d. above the average W.

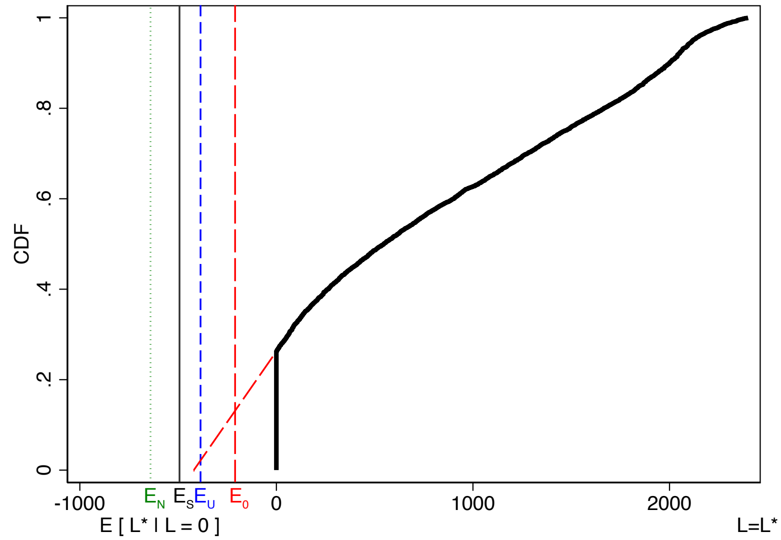
In summary, the four main qualitative conclusions in this paper are that (a) $\beta < 0$, (b) $\beta_A < 0$, (c) $\beta_W \geq 0$ and (d) $|\beta_A| > \frac{1}{4}|\beta_W|$, and these do not depend on the particular distributional assumption we make. Regarding our quantitative results, the estimate of β is more sensitive to the particular distributional assumption than the estimates of β_A and β_W.

Remark 6.1. ***Quantitative Robustness***
To make further progress on understanding the robustness of our quantitative findings on β to mistakes in the distributional assumption, one may consider restricting the range of possible values of $\mathbb{E}[L^*|L = 0]$.[37] Figure 9 presents a useful benchmark method for restricting the range of $\mathbb{E}[L^*|L = 0]$. The right side of the figure shows as a thick black curve the cumulative density function (CDF) of L, with bunching of 25% of mothers at $L = 0$. The left side of the figure shows different vertical lines representing different estimates of the average type of these bunched mothers. Besides the normal, tail symmetry and uniform estimates which were already shown in Figures 6-8, we also show a long-dashed vertical red line, positioned at the expectation level $\mathbb{E}_0$, which we explain next.

Recall that a mother of type $L^* = 0$ is exactly indifferent between working and not working. Thus, $\mathbb{E}[L^*|L = 0]$ represents how far from indifference is the average mother who chooses $L = 0$. Suppose that among the mothers who do not work ($L^* \leq 0$), no type is more common than those that are indifferent between working and not working ($L^* = 0$). Thus, although there could be many more mothers of type $L^* < 0$, each specific type, $L^* = l < 0$, is at most as common as the type with exact indifference, $L^* = 0$. If this assumption holds, what is the largest expectation (i.e., closest to 0) that is consistent with the data? The answer is E_0, represented by the long dashed red vertical line in Figure 9. This line is obtained by using the slope of the CDF (i.e., the PDF) of L as it approaches $L = 0$, which represents the proportion of mothers at $L^* = 0$. The diagonal long dashed red line is the CDF of $L^* \leq 0$ that would be implied by assigning all mothers of type $L^* < 0$ to a type that is as close to zero as possible (intuitively, "stacking" all these mothers to a "pile" as close to zero as possible), while making sure no type $L^* < 0$ is more common than the type $L^* = 0$. This means that E_0 is a conservative upper bound of the true value of $\mathbb{E}[L^*|L = 0]$, under this assumption.

[37] We abstract from controls X here for simplicity. The argument that follows could be made separately for each cluster of X, and therefore implemented in the case with controls. Because our results are very similar when we estimate the expectation without controls X (i.e., with $K = 1$ clusters) versus with controls (i.e., $K = 50$ clusters), this simplification is not consequential for this discussion.

Figure 9: A useful benchmark for $\mathbb{E}$



Note: The right side of the figure shows the CDF as a black thick curve, while the left side shows as vertical lines the different estimates of $\mathbb{E}[L^*|L=0]$ from Figures 6-8 plus a new benchmark estimate, $\mathbb{E}_0$ (long-dashed red vertical line). This benchmark estimate comes from the assumption that mothers with negative types around $L^*=0$ are as common as mothers of type $L^*=0$. The implied distribution of types of the mothers at $L=0$ under this assumption is also shown as a long-dashed red line. $N=6,924$.

Under the assumption above, we can conclude that $\hat{\beta} \geq -0.06$, which substantially narrows down the range of possible values of $\hat{\beta}$. In Figure 8, this strategy helps to narrow down the ranges of $\hat{\beta}_A$ and $\hat{\beta}_W$ as well, since it implies that $\hat{\beta}_A \geq -0.04$ and $\hat{\beta}_W \leq 0.04$.

6.2 The Linearity Assumption

The linearity assumption (Assumption 1) restricts the effect of the confounder L^* on outcome S to be the same for $L > 0$ as for $L = 0$. To gather more intuition about this assumption, it may be helpful to think of L^* as an index which combines the effects of several confounders. Ideally we would want to allow for the effect of L^* to vary with L because the relative weights of the confounders that compose the index L^* could change with L. For instance, some confounders may only affect S when $L = 0$, while others may only affect S when $L > 0$. If these weights are sufficiently different when $L = 0$ versus when $L > 0$, then the linearity assumption would not be a good approximation.

In this section, we study how our main findings would change under violations of this linearity assumption. We consider two complementary sensitivity analyses. First, we reduce the extent of the extrapolation of the effect of L^* to be local only to the smaller values of L, in order to understand whether the estimates we obtain change as we increase the degree of extrapolation towards larger values of L. Second, we check how our main estimates change as we allow for nonlinearities by allowing for the effect of the confounder on the outcome, δ, to change across observations with different values of L.

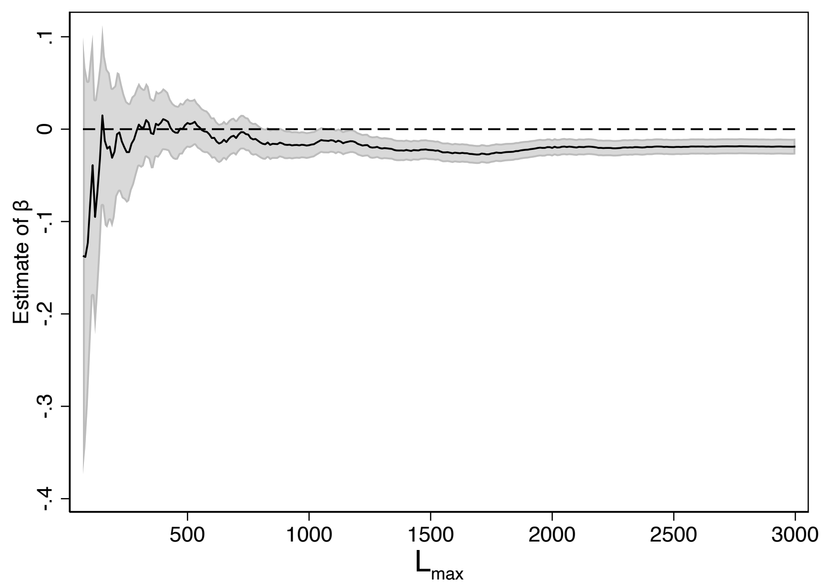
6.2.1 Local vs. nonlocal extrapolation

In the first sensitivity analysis, we restrict the sample to $L \leq L_{\max}$ for progressively larger values of $L_{\max}$. Under this restriction, we are making the "global" linearity assumption (Assumption 1) more local: $\mathbb{E}[U|X, L^*] = m(X) + \delta(X)L^*$ for $L^* \leq L_{\max}$ only. If our main conclusions are an artifact of the linearity assumption, then we should expect that $\hat{\boldsymbol{\beta}}$ for lower values of $L_{\max}$ would lead to different conclusions than $\hat{\boldsymbol{\beta}}$ for higher values of $L_{\max}$.[38] Intuitively, this check relies on the idea that the linearity tends to be a weaker assumption locally versus globally.

Figure 10 plots the homogeneous estimates of β corresponding to column (v) of Table 2 for different values of $L_{\max}$. Overall, the truncated estimates are generally indistinguishable from the full-sample estimates (which are the ones in the far right of the plot, when the restriction $L \leq L_{\max}$ is not binding). For local extrapolations around zero (i.e., for low values of $L_{\max}$), the point estimates for β are generally more negative than the full-sample estimate, although they are also imprecise. As we increase the degree of extrapolation, allowing $L_{\max}$ to become larger, the estimates become more precise and less negative, converging to the estimate shown in Table 2.

The results in Figure 10 are too noisy to allow us to assess whether going from a local to a global linearity assumption generates any bias. Still, there is strong evidence that going from an extrapolation in the range $0 \leq L \leq 1000$ to a global extrapolation (i.e., full sample) does not change the results. Thus, provided that linearity in the range $0 \leq L \leq 1000$ is valid, nonlinearities for values of L beyond 1000, if existent, do not seem to be consequential to our main findings.

Figure 10: Model from Section 5.1: Does the nonlocal linear extrapolation matter?



Note: Figure plots as a solid line estimates of β for restricted samples with $L \leq L_{\max}$. Bootstrapped 95% confidence intervals based on 250 iterations shown in gray. Estimates shown for $L_{\max} \geq 70$.

Figures 17 and 18 in Appendix C show analogous results for the main estimates from Sections

[38] We keep the expectation the same as we vary $L_{\max}$, so any change in $\hat{\boldsymbol{\beta}}$ as $L_{\max}$ varies can only be attributed to nonlinearities as we go from a local to a nonlocal linearity assumption.

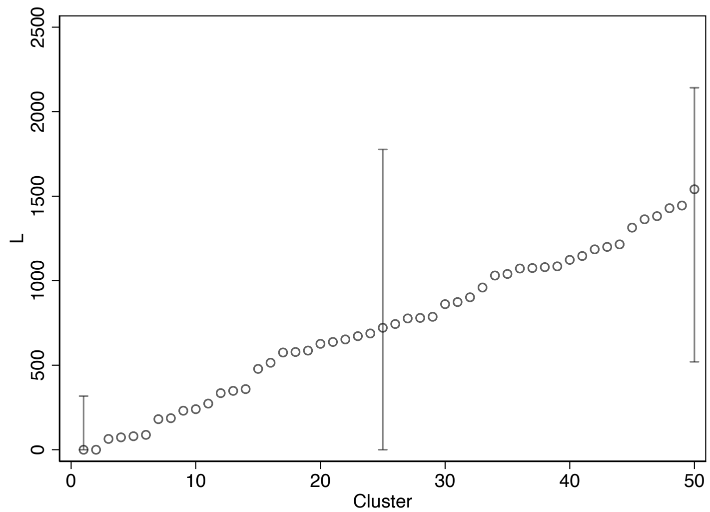
5.2 and 5.3, respectively. Although the estimates are very noisy for small values of $L_{\max}$, by $L_{\max} \approx 1000$, confidence intervals become narrower and a signal emerges. As in the homogeneous case, nonlinearities for values of L beyond 1000, if existent, do not seem to be consequential to our main findings about β_A and β_W either. However, we have not ruled out the possibility that there are nonlinearities for $L \leq 1000$ which might be biasing our results, so we turn next to a different analysis to consider this case.

6.2.2 Allowing for nonlinearities

Ideally, we would like to allow the projection in Assumption 1 to vary freely with L^* (i.e. $\mathbb{E}[U|X, L^*] = \varphi(X, L^*)$ for an arbitrary function φ), but we cannot. However, instead of assuming $\mathbb{E}[U|X, L^*] = m(X) + \delta L^*$ as we did in the results from Section 5, we can allow for the projection $\mathbb{E}[U|X, L^*] = m(X) + \delta(X)L^*$, which may partially absorb the nonlinear variation of $\varphi(X, L^*)$ in L^*, if indeed there is such nonlinearity. To clarify, because X can partially predict L^*, the variation of $\delta(X)$ across values of X may pick some of the variation in L^*, thus $\delta(X)L^*$ may fit some of the nonlinearity in L^*, if existent.

Specifically, we consider K_δ clusters of X, and specify $\delta(X) = \sum_{k=1}^{K_\delta} \delta_k \mathbf{1}(X \in \hat{\mathcal{C}}_k)$, where $\mathbf{1}(X \in \hat{\mathcal{C}}_k)$ is the indicator of whether X belongs to cluster k (See Section 4.3 for details). Figure 11 shows that the cluster indicators can indeed predict some of the variation in L (and therefore in L^*). The figure shows the median value of L within each of the $K_\delta = 50$ clusters, sorted from the lowest to the highest median. The interquartile ranges for a few selected clusters are also shown, to provide a sense of the distribution of L (and thus of L^*) within each cluster. The median L varies a great deal across clusters, from zero to over 1,500 hours. The bunching rates vary a lot too, from 57% (for the first cluster) to 11% (for the last cluster). Because of this variation, we can fit some of the nonlinearity in L^* when we allow $\delta(X)$ to vary across clusters. For instance, δ_1 is the coefficient of the first cluster, and is a weighted average of the effects of the confounder for mothers that disproportionately work few or no hours. Conversely, δ_{50} is the coefficient of the 50th cluster, and is a weighted average of the effects of the confounder for mothers who disproportionately work long hours.

Figure 11: Heterogeneity in the Distribution of L across Clusters



Note: This plot shows the median within cluster for each of the clusters when $K_\delta = 50$. Clusters are sorted from left to right based on the median of L within cluster. The interquartile ranges within cluster for selected clusters are also shown.

Thus, we assess whether nonlinearities are important in our context by comparing the results with $K_\delta = 50$, which assumes $\mathbb{E}[U|X, L^*] = m(X) + \delta(X)L^*$, with the results with $K_\delta = 1$ from Section 5, which assume $\mathbb{E}[U|X, L^*] = m(X) + \delta L^*$. If the results are similar, then the linearity assumption with $\mathbb{E}[U|X, L^*] = m(X) + \delta L^*$ might already be a good approximation.

Table 5 below shows that this is indeed the case. The table shows the estimates of β when $K_\delta = 50$, which are remarkably similar to the estimates of β when $K_\delta = 1$, shown in Table 2. This suggests that the linearity assumption is a reasonable approximation in the homogeneous model from Section 5.1.

Table 5: Model from Section 5.1: Allowing for $\delta(X)$ to vary by cluster

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Semip. Normal	(iv) Semip. Uniform	(v) Nonp. Tail 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 analogous to the estimate from column (v) from Table 2, but allowing the coefficient of the control function, $\delta(X)$, to vary with each of the 50 clusters ($K_\delta = 50$ instead of $K_\delta = 1$). N=6,924. Bootstrapped standard errors in parentheses (1,000 bootstrap samples). ** p<0.05, * p<0.1.

Tables 8 and 9 in Appendix C show the analogous results for the heterogeneous analyses from Sections 5.2 and 5.3. The main estimates when we allow δ to vary across $K_\delta = 50$ clusters are very similar to the estimates when $K_\delta = 1$ (Table 3 and Table 4, respectively). These results suggest that the linearity assumption is also a reasonable approximation in the heterogeneous models.

Remark 6.2. ***Fixed Costs and Nonlinearities***

The results above provide independent evidence in favor of the claim in Remark 2.1 that fixed costs in the decision of mothers to work should not bias our results. Specifically, we argue below that, if fixed costs were to bias our results, then this would manifest through a nonlinearity of $\mathbb{E}[U|L^*, X]$ *in* L^* *that would lead the estimates of* $\boldsymbol{\beta}$ *to vary with* K_δ*. As the results above show, this does not seem to be the case.*

To see this, note that under fixed costs, equation (1) is replaced by $L = L_1^* \cdot \mathbf{1}(L_1^* \geq FC)$*, where* L_1^* *can be understood as desired working hours, as before, and* $FC \geq 0$ *can be understood as the minimum number of hours each mother must desire to work in order for her to work at all.* FC *may be positive because of fixed costs. In general, different mothers may have different values of both* L_1^* *and* FC*.*

In this scenario, for $L > 0$*, we have* $L^* = L_1^*$*, so* L^* *continues to be a sufficient index of all confounders, since* L^* *tracks any variation in* L*, as before. However, for* $L = 0$*,* L^* *must now reflect the influence of two "primitive" confounders,* L_1^* *and* FC*.*[39] *If fixed costs matter, and the effects of the "primitive" confounders* L_1^* *and* FC *are sufficiently different, then the effect of* L^* *at* $L = 0$ *should be sufficiently different from the effect of* L^* *for* $L > 0$*. If this nonlinearity is not absorbed by controls (either by* $m(X)$ *or* $\delta(X)$*), then Assumption 1 would be violated.*

When $K = 50$*, the proportion of observations at* $L = 0$ *for each of the 50 clusters changes substantially, from 57% in cluster 1 down to 11% in cluster 50 (Figure 11). This means that effectively much of this potential nonlinearity is captured by* $m(X)$*, since this term absorbs any average difference in the effect of confounder* L^* *across clusters. To the extent that controlling for different intercepts might not be enough to handle all the nonlinearity, in this section we considered* $K_\delta = 50$ *instead of* $K_\delta = 1$*, thus allowing for* $\delta(X)$ *to capture further nonlinearity with different slopes across clusters. We do not find evidence that this last relaxation of the nonlinearity assumption is important for our results, suggesting that if fixed effects are important at all in our context, controls seem sufficient to handle any nonlinearity generated by their presence. The next section confirms this result by relaxing the linearity assumption in another way – via the function* $m(X)$ *instead – and showing that the estimates of* $\boldsymbol{\beta}$ *do not change from* $K = 50$ *all the way to* $K = 100$*.*

6.3 Controlling for X flexibly

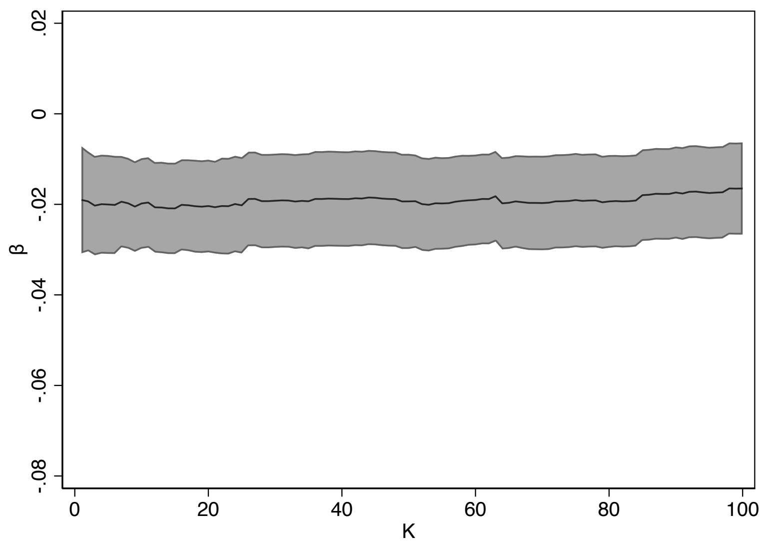
In this section, we examine whether the number of clusters ($K = 50$) we use in the results from Section 5 is appropriate. Specifically, the role of the clusters is twofold (see Section 4.3 for details): (1) to add nonparametric flexibility to the $m(X)$ expression. For $K = 1$, the specification of $m(X)$ would have been simply $X'\tau$, but for $K = 50$, we add 50 cluster indicators to the linear specification to allow for additional nonparametric variation at the cluster level; (2) to allow for heterogeneity in $\hat{\mathbb{E}}[L^*|L = 0, X]$ in X at the cluster level, i.e. the expectation is the same for every observation in the same cluster, but can vary freely across clusters. As the number of clusters K grows, we

[39] For instance, L^* might be interpreted as a weighted average between two unobservables, L_1^* and $L_2^* = L_1^* - FC$, with weights that relate to the proportion of mothers of type L_1^* inside the intervals $(-\infty, 0]$ and $(0, FC]$, respectively.

are better able to nonparametrically approximate $m(X)$ and $\mathbb{E}[L^*|L = 0, X]$. Moreover, if our results stabilize beyond a certain number of clusters $\bar{K}$, this suggests that the variation across $\bar{K}$ clusters approximately captures the necessary nonparametric flexibility needed to estimate our main coefficients. In this case, using a number of clusters $K > \bar{K}$ is unnecessary.

In Figure 12, we replicate the analysis in column (v) of Table 2 for $K = 1, \ldots, 100$. We notice that the results are very stable as we increase the number of clusters from $K = 1$ to $K = 100$. This shows that 50 (the number we use in the tables) is a sufficiently large number of clusters in order to guarantee all the nonparametric flexibility our method requires. In fact, $K = 50$ clusters is a very conservative choice, since it is far beyond the amount of flexibility the method seems to require in our context.

Figure 12: Model from Section 5.1: Different Cluster Numbers (K)



Note: Estimates correspond to the estimate in column (v) from Table 2 (based on Assumption 2), but with different numbers of clusters used in the analysis. 95% confidence intervals based on 1,000 bootstrap iterations. $N = 6,924$.

The analogous findings for our results from Tables 3 and 4 are shown, respectively, in Figures 15 and 16 in Appendix C. There too it is clear that the estimates of $\boldsymbol{\beta}$ stabilize at very low values of K.

These findings help allay further concerns about both the linearity and the distributional assumptions. As discussed in Remark 4.1, a nonparametric $m(X)$ guarantees that the linearity assumption is only made with respect to $L^* - \mathbb{E}[L^*|X]$, that is, the part of the confounder that is not predicted by controls. The stability of the estimates from low values of K all the way to $K = 100$ shows that allowing for X to enter more nonlinearly does not seem to affect the estimate of $\boldsymbol{\beta}$. Moreover, as discussed in Remark 4.2, the relative stability of our main estimate ($\hat{\boldsymbol{\beta}}$) when K grows (i.e., as we effectively change $\hat{\mathbb{E}}[L^*|L = 0, X]$) provides further evidence that our main conclusions do not seem to be dependent on the distributional assumption.

7 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 use this novel identification strategy to estimate the average treatment effect of maternal labor supply for our entire sample. We also 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 especially valuable for childhood skill accumulation, at least in the short run.

Our findings provide some useful insights for policies aimed at incentivizing mothers to work in the first years of their children's lives. We confirm the typical finding from the 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 the 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. We do find clear evidence of such unintended consequence among high-skilled mothers, even for those that are high earners, suggesting that additional earnings are not enough to compensate for their absence. On the one hand, our finding that these unintended consequences tend to be concentrated in 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, based on our results, work-promoting policies aimed at increasing the flexibility of work arrangements seem the best bet to avoid the 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. Flexible work schedules may also allow partners to be a better substitute for the mother's absence. The need for social distancing during the Covid-19 pandemic has led to marked increases in the share of workers operating under 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.

Additional research is also needed to understand the exact mechanisms through which the effects we uncover operate. For instance, maternal labor supply may affect the child's short-run skills not only through the amount of time the mother interacts with her child, but also through the average quality of the interaction. In turn, the average quality of the interaction may change in either direction with increased maternal labor supply, either because of increased work stress or because of reduced money stress. There are also two distinct but complementary channels worth disentangling. First, the hours worked by the mother in the first three years of the child's life may have a direct effect on the child's skills measured around age 6. Second, the mother's labor supply in these early years might have a causal effect on her subsequent labor supply in years 4-6, which in turn might have a direct effect on the child's skills around age 6. Separately identifying these effects would add value to this literature by improving our understanding of the specific times in early childhood when longer hours of mother-child interaction are particularly valuable.

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.

Angrist, J. D. and M. Rokkanen (2015). Wanna get away? Regression discontinuity estimation of exam school effects away from the cutoff. *Journal of the American Statistical Association 110*(512), 1331–1344.

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. *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 (2023). Correcting Endogeneity Bias in Models with Bunching. *Working Paper*. Available here.

Calonico, S., M. Cattaneo, M. Farrell, and R. Titiunik (2019). Regression-Discontinuity Designs Using Covariates. *The Review of Economics and Statistics 101*(3), 442–451.

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.

Cogan, J. (1980). Labor supply with costs of labor market entry. In *Female labor supply*, pp. 327–364. Princeton University Press.

Cogan, J. F. (1981). Fixed costs and labor supply. *Econometrica 49*(4), 945–63.

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.

Flood, S., J. McMurry, A. Sojourner, and M. Wiswall (2022). Inequality in early care experienced by us children. *Journal of Economic Perspectives 36*(2), 199–222.

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.

Gronau, R. (1974). Wage comparisons–a selectivity bias. *Journal of political Economy 82*(6), 1119–1143.

Hahn, J., P. Todd, and W. van der Klaauw (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica 69*, 201–09.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning: Data mining, inference, and prediction.* Springer Science & Business Media.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica: Journal of the Econometric Society*, 679–694.

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.

Mroz, T. A. (1987). The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica 55*(4), 765–799.

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? *Forthcoming at Journal of Labor Economics, https://www.journals.uchicago.edu/doi/10.1086/719688*.

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.

Zabel, J. E. (1993). The relationship between hours of work and labor force participation in four models of labor supply behavior. *Journal of Labor Economics 11*(2), 387–416.

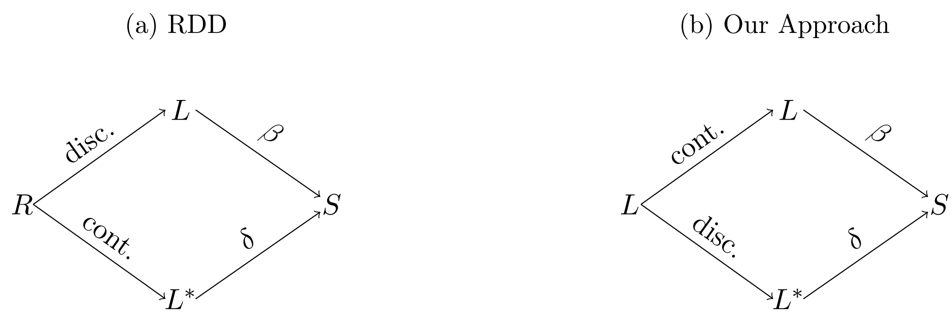
A Relationship to Regression Discontinuity Designs

In this Appendix, we clarify the relationship between our method and regression discontinuity designs (RDDs), as suggested in the Introduction. To streamline the discussion, we consider the homogeneous case without controls X. In that case, equation (3) simplifies to

$$S = m_0 + \beta L + \delta L^* + \varepsilon, \tag{11}$$

where m_0 is a constant. Using this model, Figure 13 provides a graphical representation of the parallel between the two methods, where the arrows represent causal relationships. We are interested in the effect of L on S, denoted β, and if $\delta \neq 0$, we have an endogeneity problem, since L^* is an index of all potential confounders.

Figure 13: Relationship between RDD and Our Approach



The left panel of this figure shows the standard "fuzzy" RDD case, e.g. Hahn et al. (2001). Typically, in a fuzzy RDD there is a "running variable" R such that the treatment variable L is discontinuous in R at a known threshold, while the confounder L^* is continuous in R at the same threshold. Thus, whenever the outcome variable S is discontinuous in R at the threshold, we can infer that this discontinuity happened through the L channel and not through the L^* channel. We can identify β by simply dividing the intention-to-treat (reduced-form) effect, which is the discontinuity of S in R at the threshold, by the first stage, which is the discontinuity of L in R at the threshold.

In contrast, our approach is shown in the right panel of the figure. It can be seen as an "upside down" fuzzy RDD that aims to identify the effect of L^*, rather than the effect of L. To see this, consider that in our case L is both the treatment and the "running variable", using the language of RDD. The treatment L is continuous in the running variable L, as they are the same variable. However, as discussed in Figure 1 in the Introduction, L^* is discontinuous in the running variable L at the threshold $L = 0$. Thus, any discontinuity in the outcome S at $L = 0$ can be attributed solely to the L^* channel. Using the language of RDD, we can identify δ as the ratio of two discontinuities: the "intention-to-treat" or "reduced-form" discontinuity of $\mathbb{E}[S|L]$ at $L = 0$ and the "first stage" discontinuity of $\mathbb{E}[L^*|L]$ at $L = 0$. More formally, the "intention-to-treat" or "reduced-form" discontinuity is[40]

$$\begin{aligned} ITT &:= \lim_{l \downarrow 0} \mathbb{E}[S|L=l] - \mathbb{E}[S|L=0] \\ &= \lim_{l \downarrow 0} \mathbb{E}[m_0 + \beta L + \delta L^*|L=l] - \mathbb{E}[m_0 + \beta L + \delta L^*|L=0] = -\delta \mathbb{E}[L^*|L=0], \end{aligned}$$

which holds because $\lim_{l \downarrow 0} \mathbb{E}[\beta L + \delta L^*|L=l] = \lim_{l \downarrow 0} \mathbb{E}[\beta L + \delta L|L=l] = 0$. Moreover, the "first stage" discontinuity in this case is

$$FS := \lim_{l \downarrow 0} \mathbb{E}[L^*|L=l] - \mathbb{E}[L^*|L=0] = -\mathbb{E}[L^*|L=0].$$

Figure 1 in the Introduction illustrates this interpretation of the method: the left panel is analogous to the reduced form RDD plot, while the right panel is analogous to the first stage RDD plot. Unlike an actual RDD, our approach cannot directly identify the "first stage" discontinuity, since L^* is not observed. This is why we need to first guarantee that there is a "first stage" with the assumption $\mathbb{P}(L^* < 0) > 0$ (which implies $\mathbb{E}[L^*|L=0] < 0$), and then make a distributional assumption to identify the size of the discontinuity, $\mathbb{E}[L^*|L=0]$. Once this "first stage" is obtained, we can identify δ simply by dividing the "intention-to-treat" effect by the "first stage" effect, as in an RDD. Once the confounder effect is identified, we can easily obtain the treatment effect by subtracting the endogeneity bias from the observed naive difference in outcomes. Our control function approach implements this idea by identifying β and δ in the same step, rather than in two steps as discussed above, and by using the data from the full range of L, not only a local neighborhood of $L = 0$ as in this analogy. See Remark A.1 for further details.

Remark A.1. ***Why do we need to make the Linearity Assumption?***
The parallels between our method and the RDD also clarify why we need the linearity assumption (Assumption 1). In the calculation of the ITT above, we only used the linearity assumption locally (i.e., the assumption that $\mathbb{E}[U|L^*]$ is linear in $L^* - \mathbb{E}[L^*|X]$ in a neighborhood of $L = 0$). We can therefore forego the global linearity assumption in favor of a local linearity assumption only around $L = 0$. In this case, our RDD-like strategy of dividing the intention to treat by the first stage discontinuities identifies a meaningful treatment effect: the average treatment effect of L for those at $L^* = 0$. The reason we make the global linearity assumption in the paper is that we want to identify the average treatment effect among all mothers, not only among those at $L^* = 0$. Thus, we need to make an extrapolation by assuming that the effect of L^* we identify at $L = 0$ is the same as the effect of L^* for $L > 0$. This extrapolation is where the parallel with the standard RDD ends: RDD studies most often focus on identifying the effect only at the threshold. There is still a parallel with less standard RDD studies that attempt to extrapolate in order to identify nonlocal effects, as they typically make similar assumptions to our global linearity assumption. For instance, Angrist and Rokkanen (2015) identify effects away from the threshold by using an assumption of ignorability of the running variable conditional on other predictors of the outcome. This ignorability assumption

[40] Unlike in the RDD, the threshold is at the extreme of the support of the distribution of the running variable, at $L = 0$, so we do not need to take the limit from the other side.

is closely related to the global linearity assumption we make: for $L > 0$, L^* has no effect on the outcome except through $m(X)$ or the control function $\delta(X)[L + \mathbb{E}[L^*|L = 0, X]\mathbf{1}(L = 0)]$. For more context, Section 6.2.1 shows the estimates of β when we only make a linearity assumption locally (with different bandwidths, as done in local linear estimation for RDDs) in order to estimate the effect around the bunching point.

Remark A.2. ***Role of Covariates***
Another important difference between RDDs and our approach is the use of covariates. Typically, covariates are added in RDDs only for efficiency purposes (e.g. Calonico, Cattaneo, Farrell, and Titiunik 2019). In contrast, as discussed in Remarks 4.1 and 4.2, covariates are useful in our approach to weaken or test both the linearity and the distributional assumptions. Covariates are also useful in both methods to provide supporting evidence in favor of the identification strategy. In RDDs, one typically indirectly argues that unobservables are continuous in the running variable at the cutoff by showing that observables are continuous at the cutoff. Analogously, in Figure 4 we indirectly argued that unobservables are discontinuously at the bunching point by showing that observables are discontinuous at the bunching point.

B Sensitivity Analysis Based on Oster (2019)

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 6. 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 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 relies on 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 6 as providing independent confirmation of the plausibility of our main estimates.

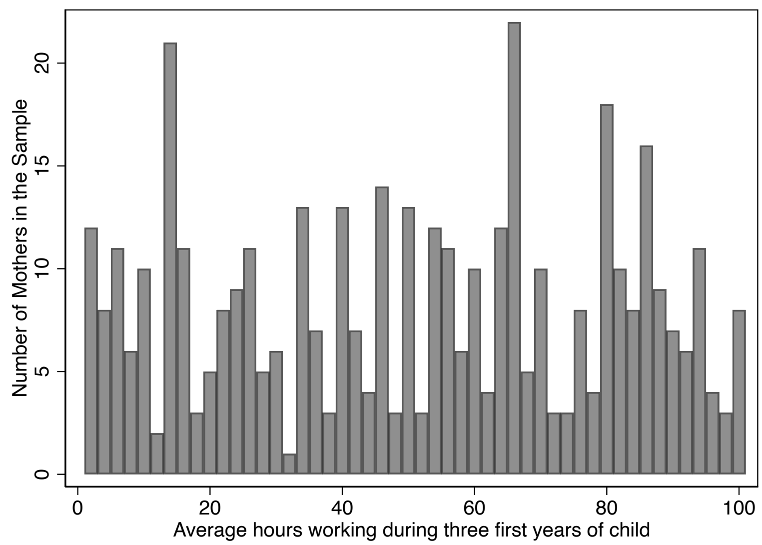
Table 6: Proportional selection of observables and unobservables, Oster (2019).

	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 column (v) of Table 2. δ_{Oster} can be interpreted as the degree of selection on unobservables relative to observables.

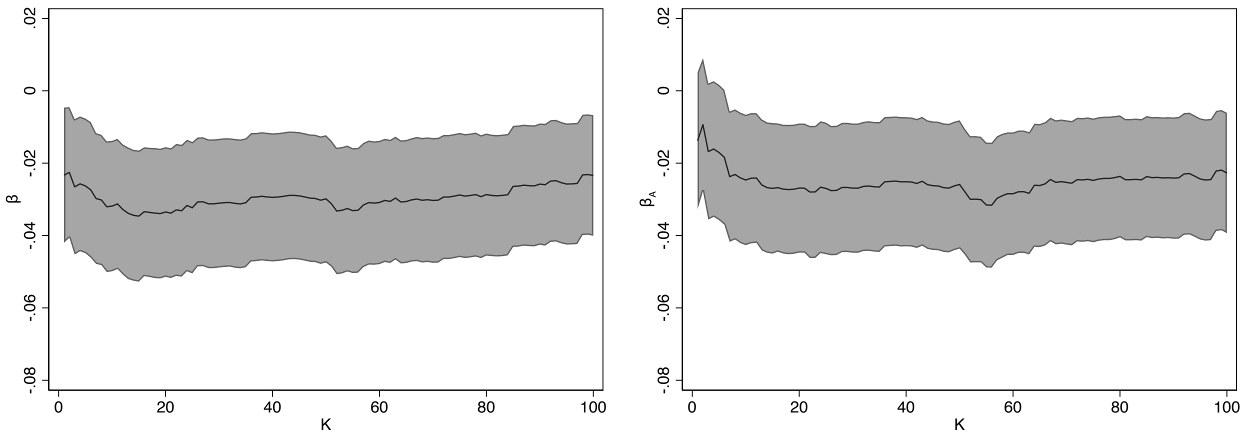
C Other Tables and Figures

Figure 14: Distribution of L near $L = 0$



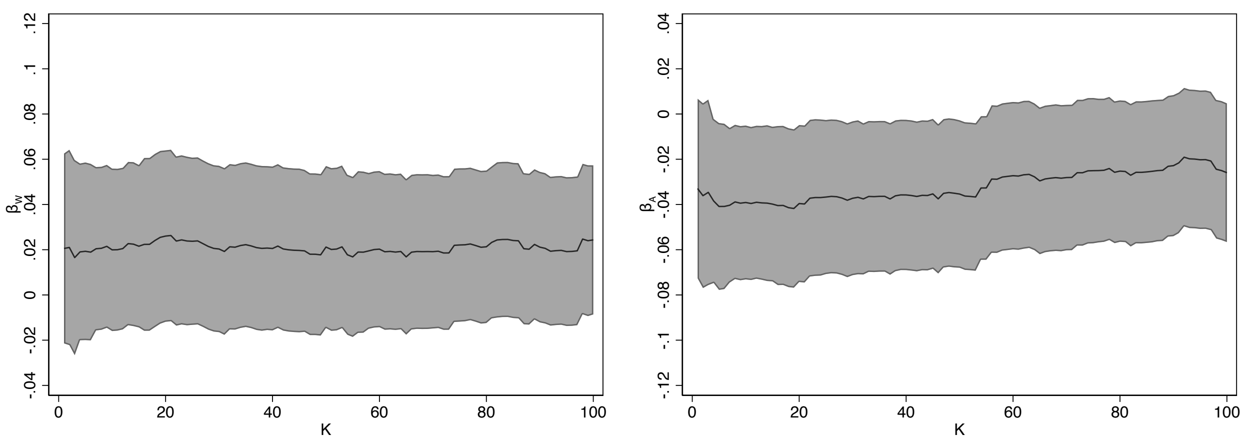
Note: This Figure shows the histogram of $0 < L \leq 100$ for the full sample, using a bandwidth equals to 2 hours.

Figure 15: Model from Section 5.2: Different Cluster Numbers (K)



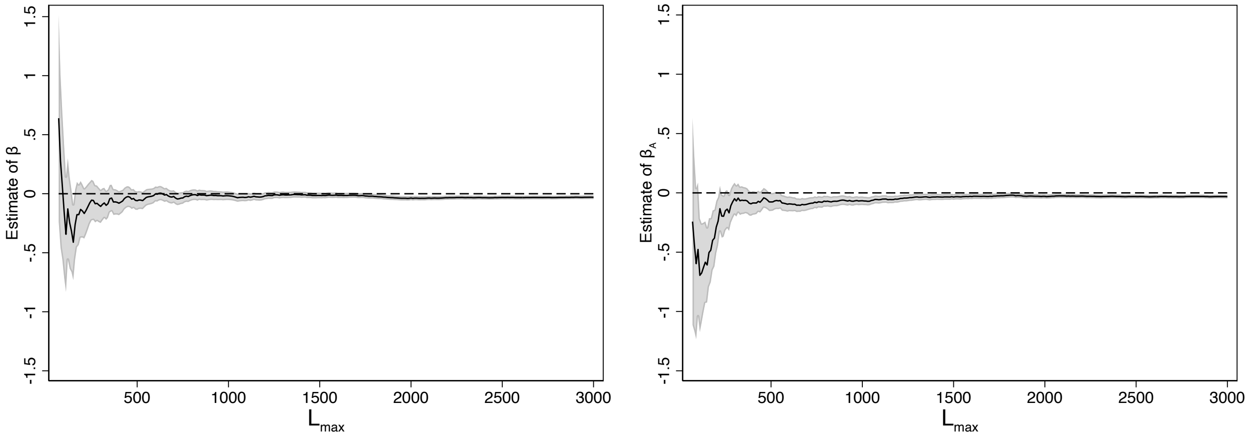
Note: Estimates correspond to the estimate in column (v) from Table 3 (based on Assumption 2), but with different numbers of clusters used in the analysis. 95% confidence intervals based on 1,000 bootstrap iterations. $N = 6,924$.

Figure 16: Model from Section 5.3: Different Cluster Numbers (K)



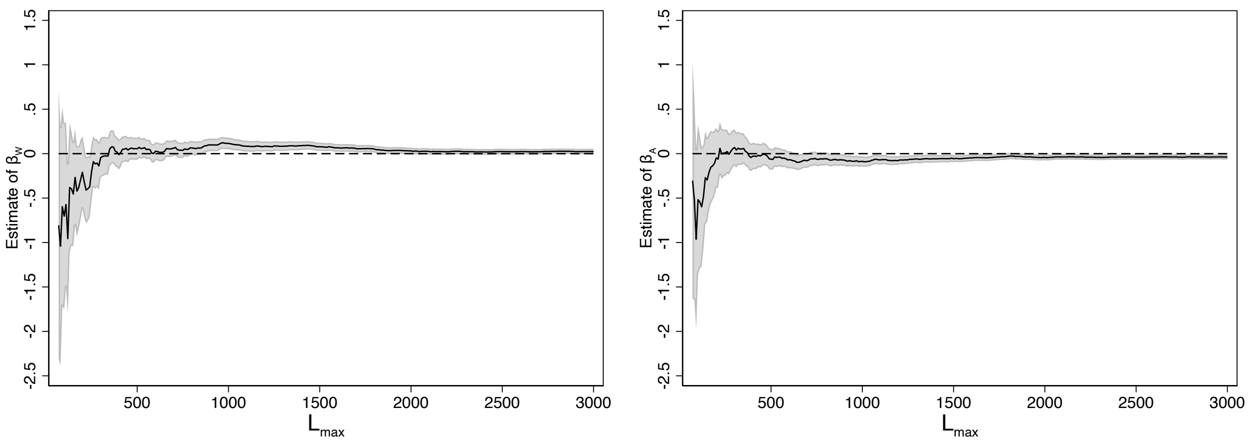
Note: Estimates correspond to the estimate in column (v) from Table 4 (based on Assumption 2), but with different numbers of clusters used in the analysis. 95% confidence intervals based on 1,000 bootstrap iterations. $N = 3,994$.

Figure 17: Model from Section 5.2: Does the nonlocal linear extrapolation matter?



Note: Figure plots as a solid line estimates of β (left panel) and β_A (right panel) for restricted samples with $L \leq L_{\max}$. Bootstrapped 95% confidence intervals based on 250 iterations shown in gray. Estimates shown for $L_{\max} \geq 70$.

Figure 18: Model from Section 5.3: Does the nonlocal linear extrapolation matter?



Note: Figure plots as a solid line estimates of β_W (left panel) and β_A (right panel) for restricted samples with $L \leq L_{\max}$. Bootstrapped 95% confidence intervals based on 250 iterations shown in gray. Estimates shown for $L_{\max} \geq 70$.

Table 7: Model from Section 5.1 – Alternative Samples and Specifications

		(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Semip. Normal	(iv) Semip. Uniform	(v) Nonp. Tail Symmetric
From Table 2	β	0.014**	0.000	-0.016**	-0.019**	-0.019**
($N = 6,924$)		(0.001)	(0.001)	(0.005)	(0.006)	(0.005)
	δ			0.014**	0.017**	0.017**
				(0.004)	(0.005)	(0.005)
Married	β	0.007**	-0.002	-0.011*	-0.014*	-0.012*
($N = 4,203$)		(0.002)	(0.001)	(0.006)	(0.008)	(0.007)
	δ			0.008	0.011	0.009
				(0.005)	(0.007)	(0.006)
Not Married	β	0.021**	0.005**	-0.012	-0.010	-0.021**
($N = 2,721$)		(0.002)	(0.002)	(0.008)	(0.008)	(0.009)
	δ			0.014**	0.012*	0.022**
				(0.006)	(0.007)	(0.008)
Not College	β	0.012**	0.001	-0.015**	-0.019**	-0.018**
($N = 5,944$)		(0.001)	(0.001)	(0.005)	(0.006)	(0.006)
	δ			0.014**	0.018**	0.017**
				(0.004)	(0.005)	(0.005)
Other Income	β	0.013**	-0.000	-0.017**	-0.023**	-0.017**
($N = 5,176$)		(0.001)	(0.001)	(0.006)	(0.007)	(0.007)
	δ			0.014**	0.020**	0.014**
				(0.005)	(0.006)	(0.006)

Note: This table repeats the specification from Table 2 using different samples and/or controls. The first panel reproduces the results from Table 2. The "Married" results use the sub-sample of women who report being married at the birth of their child, while the "Not Married" results use the complementary sub-sample. The "Not College" results re-run the analysis using the sub-sample of women who report not having a college degree. The "Other Income" results use the total pre-birth income of the mother's household from all sources other than her labor earnings as an additional control. ** p<0.05, * p<0.1.

Table 8: Model from Section 5.2: Allowing for $\delta(X)$ to vary by cluster

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Semip. Normal	(iv) Semip. Uniform	(v) Nonp. Tail 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 analogous to the estimate from column (v) from Table 3, but with $K_\delta = 50$ instead of $K_\delta = 1$. N=6,924. Bootstrapped standard errors in parentheses (1,000 bootstrap samples). ** p<0.05, * p<0.1.

Table 9: Model from Section 5.3: Allowing for $\delta(X)$ to vary by cluster

	(i) Uncorrected No Controls	(ii) Uncorrected w/ Controls	(iii) Semip. Normal	(iv) Semip. Uniform	(v) Nonp. Tail 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 analogous to the estimate from column (v) from Table 4, but with $K_\delta = 50$ instead of $K_\delta = 1$. N=3,994. Bootstrapped standard errors in parentheses (1,000 bootstrap samples). ** p<0.05, * p<0.1.