Towards Causal Estimates of Children's Time Allocation on Skill Development

Gregorio Caetano, Josh Kinsler and Hao Teng*

April 2017

Abstract

Cognitive and noncognitive skills are critical for a host of economic and social outcomes as an adult. While there is broad agreement that a significant amount of skill acquisition and development occurs early in life, the precise activities and investments that drive this process are not well understood. In this paper we examine how children's time allocation affects their accumulation of skill. Children's time allocation is endogenous in a model of skill production since it is chosen by parents and children. We apply a recently developed test of exogeneity to search for specifications that yield causal estimates of the impact time inputs have on child skills. We show that the test, which exploits bunching in time inputs induced by a non-negativity time constraint, has power to detect endogeneity stemming from omitted variables, simultaneity, measurement error, and several forms of model misspecification. We find that with a sufficiently rich set of controls, we are unable to reject exogeneity in our most detailed production function specifications. The estimates from these specifications indicate that active time with adult family members, such as parents and grandparents, are the most productive in generating cognitive skill.

1 Introduction

There is a growing consensus among economists that skills acquired during childhood have an important influence on later life outcomes.[1] This view stems from extensive research

*Gregorio Caetano: Economics Department, University of Rochester. Josh Kinsler: Economics Department, University of Georgia. Hao Teng: Economics Department, University of Rochester. We would like to thank Carolina Caetano, Vikram Maheshri, Daniel Ringo, and David Slichter for helpful discussions. All errors are our own.

[1]See Almond and Currie (2011) for a comprehensive review of the literature.

linking childhood cognitive and noncognitive skills with education, labor market, health, and criminal outcomes.[2] Additional analyses indicate that adult labor market outcomes are largely determined by skills already in place by age 16.[3] In light of the evidence linking child and youth cognitive and noncognitive skills to adult outcomes, it is imperative that we understand the determinants of these skills such that parents and policy-makers can invest wisely.

A common approach for estimating the impact of child and youth investments is to relate skill measures to an activity or input of interest, such as time spent in childcare, active time spent with the mother, or time spent studying.[4] The challenge in estimating the impact of a specific activity on skill development is three-fold.[5] First, the time allocation of child and youth activities is endogenous since it reflects choices made by parents and children. Second, even when exogenous variation in an activity of interest is available, it is difficult to interpret the resulting coefficient without information on the substitution among all potential activities. Finally, in the extreme scenario that all relevant inputs are observed, sample size constraints force researchers to aggregate inputs and impose parametric restrictions. The concern is that economic theory provides little guidance about how to aggregate inputs and which specific parametric restrictions to impose. If the researcher's choices are inconsistent with the data generating process, the parameters of interest can be biased.

Going forward, these three issues are unlikely to be jointly solved by an instrumental variables approach since each input requires its own instrument. Moreover, running an experiment is also infeasible, as the treatment arms would have to consist of fully prescribed inputs for each child. If only some inputs are manipulated, parents can optimize over the remaining ones to either reinforce or negate the intended effect. Either way, it will not be possible to estimate the ceteris paribus causal impact of substituting one time input for a well-defined alternative time input.

In this paper, we take an alternative approach aimed at handling these three issues. First, we consider a large space of skill production specifications, where each specification contains many restrictions made for tractability that are common in the literature. For example, we group certain inputs together, restrict complementarities among various inputs, and so on. For each specification, we then assess the extent to which the corresponding restrictions lead to endogeneity using a recently developed exogeneity test (Caetano (2015) and Caetano and

[2]McLeod and Kaiser (2004) and Currie and Thomas (1999) link cognitive skill measures at age 7 with future educational attainment and employment outcomes respectively. Cunha et al. (2006), Deming (2009), and Heckman et al. (2013) highlight the importance of non-cognitive skill formation both for further cognitive skill development and outcomes such as teen pregnancy, crime, and educational attainment.

[3]See Keane and Wolpin (1997) and Cameron and Heckman (1998) as examples.

[4]See for example Bernal and Keane (2010), Del Boca et al. (2013), and Del Boca et al. (2016).

[5]See Todd and Wolpin (2003) and Keane (2010).

Maheshri (2016)). The test provides an objective statistical criterion to determine whether the parameters of interest can be interpreted as causal. Thus, the test serves as a guide for model selection, leading to a rejection of some, but not all specifications. Of course, failing to reject a specification does not guarantee that the parameters of that specification can be interpreted as causal. Indeed, a specification can suffer from endogeneity and our test may fail to detect it. The key methodological contribution of this paper lies in systematically studying the power of the test in the context of time allocation and skill production. If we find that the test is sufficiently powerful in this context, then failing to reject a given specification allows us to reasonably interpret its main parameters as causal.[6]

The exogeneity test we utilize exploits the idea that unobservable confounders tend to vary discontinuously when a time input is zero. For instance, consider the time input "father reading a book to the child". Families that perform this activity sixty minutes a day are similar to families that perform this activity forty-five minutes a day. Furthermore, families that perform this activity forty-five minutes a day are similar to families that perform this activity thirty minutes a day, and so on. However, the notion of similarity breaks down at zero minutes a day. For instance, families that spend no time in this activity may have an unavailable father or may significantly devalue reading. Additionally, because reading time cannot be negative, these unobserved traits tend to accumulate at zero. As a result, families in which the father spends zero minutes reading to the child are likely discontinuously different from families in which the father spends even a small amount of time reading to the child. To test for whether such unobserved heterogeneity exists, we exploit the idea that paternal reading time varies continuously from sixty minutes down to zero minutes, while unobservables correlated with this variable vary discontinuously at zero minutes. If any of these discontinuous unobservables are incorrectly omitted from the specification, the dependent variable (in our case a skill) will vary discontinuously at zero, leading us to reject the specification.

While rejected specifications are discarded, a specification that passes the exogeneity test may not necessarily be accepted for causal inference, as the test may not be sufficiently powerful to detect endogeneity. We investigate the power of the exogeneity test in our setting using a number of complementary tools. First, we provide evidence that each activity of interest has a discontinuous mass of observations at zero minutes, providing direct evidence of

[6] The test was originally developed by Caetano (2015) as a way of *rejecting* specifications, not *validating* them. Thus, she focused on showing that the test had some power to detect endogeneity. In contrast, in our approach we use this test to validate a specification. For that, we not only need to show that the test has some power; we need to show that the test is so powerful that failing to detect endogeneity in this test reasonably points to the conclusion that the specification has no endogeneity. Doing this requires a systematic study of the power of the test in our context. See Caetano and Maheshri (2016) for further discussion on this topic.

bunching (McCrary (2008)). The existence of many different activities where families bunch at zero minutes contributes to a substantial increase in the power of the test. Indeed, different unobservables may vary discontinuously at zero minutes depending on the activity. We also show that over 90% of all our observed socio-demographic variables are discontinuous at the zero minute threshold for at least one activity, suggesting that unobserved confounders are also discontinuous at these thresholds.[7] Next, we provide both theoretical and empirical evidence that our test is capable of detecting endogeneity from a comprehensive list of sources, including omitted variables, simultaneity, measurement error, and all conceivable types of misspecification. Further, in practice we find that the test is able to detect discontinuities in the most parsimonious skill specifications, consistent with our power analysis. Finally, we pursue a series of additional robustness exercises designed to detect confounders that might have previously evaded detection, and compare estimated time input productivities across specifications that reject and fail to reject exogeneity. Throughout, we track the cumulative properties a confounder must have to bias our preferred estimates and be undetectable by both the test of exogeneity and the further robustness checks we perform. Ultimately, we conclude that interpreting our preferred estimates as causal is plausible, in the sense that it is unlikely a confounder will possess all these properties at the same time.

We implement our approach using skill assessments and time diaries from the Child Development Supplements of the Panel Study of Income Dynamics (PSID). With the help of their primary caregiver, children fill out a detailed 24 hour time diary to record all of their activities during the day, where each activity took place, and with whom they did the activity. These time diaries are collected in 1997, 2002, and 2007, and cover one weekday and one weekend day for each survey year. Cognitive and noncognitive skills are also assessed during each wave of the survey. In addition to time use and skill assessments, the PSID also includes a detailed list of child demographics, family background characteristics, and other measures of the environment in which the child is raised. Our estimation sample comprises the 2002 and 2007 survey years since we utilize a value-added production model in our baseline specification. Children in our sample are between 10 and 18 years of age. This is a particularly interesting point in time in a child's life as their choices regarding time use become increasingly more autonomous. Yet, parents can still steer their children towards activities and activity partners that tend to be productive.

Our search for a skill production specification whose parameters can be interpreted as causal proceeds in the following manner. In our baseline specifications, we categorize child ac-

[7] This logic is analogous to the one used in regression discontinuity (RD) designs. Support for the RD identifying assumption is based on the idea that if observables vary continuously at a threshold, then unobservables are also likely to vary continuously at that same threshold.

tivities according to the level of engagement – active (e.g., reading) or passive (e.g., watching TV) – and with whom the activity is completed – mother, father, siblings, friends, grandparents, others, or no one. We then relate the time devoted to these activities to skill measures (math, vocabulary, comprehension, and noncognitive) using standard production functions, such as value-added. We also consider models containing many other subclassifications of activities as suggested in the previous literature. The activity excluded from all specifications is sleeping, so that the results should be interpreted as a substitution between a given activity and sleeping time. Every specification also includes a series of indicator variables that reflect whether the time devoted to each activity is zero. These indicator variables are included to absorb any discontinuous change in the outcome variable when inputs are zero, conditional on controls. If we reject the null hypothesis that the coefficients on the zero time input indicators are jointly equal to zero, then we conclude that this particular specification suffers from endogeneity. In our context, a "specification" is defined by the outcome variable (skill), the main explanatory variables (time inputs), controls (ranging from none to a detailed list of child, family, and environmental observables) and a particular functional form establishing the relationships between these variables.

Our search ultimately identifies specifications where we fail to reject exogeneity of children's time inputs. Part of our success in this endeavor is due to the detailed nature of the time use data. When we estimate the impact of a particular time input, we are able to control for all other time inputs of the same child. These alternative inputs absorb much of the endogeneity, as they elicit heterogeneity in preferences and constraints across children and activity partners in the sample. However, the time use data alone is not sufficient to account for all endogeneity. Lagged skills in combination with different categories of controls are important in absorbing endogeneity for different skill measures. For vocabulary skill specifications, child characteristics are crucial, while for comprehension specifications, child as well as mother characteristics are important. For math skill specifications, child and mother characteristics are not enough, as other family member's characteristics are necessary to eliminate endogeneity. For noncognitive skill, we find that even in the most parsimonious specifications we are unable to reject exogeneity. This finding could stem from either a lack of power to detect endogeneity in the production function of noncognitive skill, or that time inputs are truly exogenous in that case. We provide evidence that our test has much less power for non-cognitive skill than for cognitive skills, and thus believe the former interpretation is both consistent with our analysis and more conservative. As a result, we do not report time input estimates for noncognitive skill formation in the main text.[8]

However, for the cognitive skill specifications that fail to reject exogeneity, we are more

[8] For completeness, we provide these results in the appendix.

confident that the key parameters of interest can be interpreted causally. Our estimates indicate that active time with adult family members, such as parents and grandparents, most promote cognitive skill formation.[9] For example, one additional hour per week spent in active time with grandparents leads to 3.2% of a standard deviation increase in comprehension scores.[10] The large, positive impact that time with grandparents has on child skill accumulation has not been documented in this literature previously.

Overall, we find that active time is not necessarily better than passive time at improving skills. Skill productivity depends substantially on the details of the activity, which in this paper are partially encapsulated by the activity partners. For instance, compared to everyone else, mothers tend to spend a higher proportion of their active time with the child on educational activities, and grandparents tend to be more engaged during activities. This could explain why active time with grandparents tends to be most productive. Future research utilizing larger samples is needed to disaggregate these production functions even further to identify the specific activities that are most productive. Finally, we present a simple model of time allocation to help illustrate why selection on observables might be a valid assumption in this context. The model is also useful as a lens through which to interpret our findings, place them in the literature, and to discuss promising avenues for further research.

There is a prolific literature estimating the impact of parental investments and child activities on skills. Dustmann and Schönberg (2012) and Bernal and Keane (2010) utilize quasi-experimental policy variation that enables them to study the impact of an increase in maternal time on child cognition. Todd and Wolpin (2007) and Fiorini and Keane (2014) estimate the impact of a more comprehensive list of child inputs on skills, and discuss selection of models using a criterion related to goodness of fit.[11] Cunha et al. (2010) focus primarily on studying the impact of investments performed at different moments of the child's life, rather than the impact of a specific investment on skills. They collapse many measures of parental inputs into a scalar labeled parental investment, enabling them to estimate complex models with dynamic complementarities, but only address endogeneity as it pertains to this scalar investment measurement. The approach developed here is complementary to these studies. We systematically study the impact of a variety of child inputs on skill using a model selection criterion related to exogeneity. Going forward, our approach can be used to

[9]Non-time inputs, such as mother's education and family income, are associated with skill development as expected, though interpreting these associations as causal is outside of the scope of this paper.

[10]This result aligns with studies in developmental psychology which emphasize the irreplaceable role of grandparents in the development of grandchildren (see Smith (2003) for more details).

[11]Todd and Wolpin (2007) choose their most preferred model using root mean-squared error as the selection criterion. Fiorini and Keane (2014) approach the identification problem by estimating multiple production functions that rely on different exogeneity assumptions, and focusing on the stability of the productivity ranking of inputs.

assess whether skill production estimates are biased as a result of restrictions that contradict the data generation process. Further, our approach provides useful insight into the types of data researchers need to gather in order to allay endogeneity concerns.

The rest of the paper is organized as follows. In Section 2, we describe the PSID data. Section 3 presents our approach and provides both theoretical and empirical evidence in support of this approach. In Section 4, we present our main results and robustness checks. In Section 5, we discuss the interpretation of our results, before we conclude in Section 6. An online appendix provides more detailed evidence regarding Sections 3 and 4.[12]

2 Data

To estimate the effect of time inputs on skill we use data from the Panel Study of Income Dynamics (PSID) and the three waves of the Child Development Supplements (CDS-I, CDS-II, and CDS-III).[13] In 1997, the PSID started collecting data on a random sample of the PSID families that had children under the age of 13. About 3,500 children aged 0-12 residing in 2,400 households were interviewed in 1997, and then followed in two subsequent waves, 2002 and 2007. Rows 1-3 in Table 1 illustrate the age range and average age for each wave, respectively.

Data collected in the CDS include measures of children's cognitive and noncognitive skills, time use diaries, and information about child and family characteristics, such as parent-child relationships, child health, and home environment. We match the CDS children with their PSID families to get additional information such as family annual income, mother and father's ages, mother and father's education levels, and so on. We pool CDS children across the 2002 and 2007 waves.[14] Row 4 in Table 1 illustrates the age range and average age of children in our sample.

To the best of our knowledge, the only other data set combining information on child skills and family background with time use diaries is the Longitudinal Study of Australian Children (LSAC). Compared to the LSAC, the PSID-CDS has the advantage of focusing on a larger age range of children (0-22 years old) and has richer time use data in terms of the number of child activities and with whom these activities were performed. As an example, the PSID-CDS allows us to separate the time children spend with mothers and fathers, who

[12]This appendix is available at http://bit.ly/1KOy1aj.

[13]Panel Study of Income Dynamics is a US longitudinal survey of a nationally representative sample of individuals and families, started in 1968 with a sample of 4800 families. It is funded by National Institute of Child Health and National Development (NICHD).

[14]Since we estimate a value-added model, we only use data from the 1997 wave to construct lag scores for the 2002 wave of the data.

according to Del Boca et al. (2013), have differential impacts on skill development. The ability to split activities according to detailed partner categories is also helpful in mitigating endogeneity, an issue we discuss further in Section 5.1.

2.1 Time Use Diaries

The time use diary from the CDS collects the details of child activities for two random days of a week (one weekday and one weekend). Diary forms are mailed to each child's address, and each child (with the help of her primary caregiver if needed) fills out a detailed 24 hour time diary to record all of her activities during the day, such as where each activity took place and with whom they did the activity.[15] An interviewer then visits the household to check/edit the diary that has been completed.[16]

The PSID-CDS classifies child activities according to the type of activity (215 in CDS I, 317 in CDS II, and 315 in CDS III), where the activity took place (14), and with whom (11) the activity was completed. In CDS I, most diaries (80%) are completed by the child's primary caregiver or the child and her primary caregiver together. Sampled children are considerably older in CDS II and CDS III, and as a result approximately half of the children in these rounds complete the time diaries on their own.

We clean the time use data so that the diaries are as representative as possible. Time diaries may have limited reliability since they are only a very small sample of a given child's days.[17] To allay this concern, we first exclude cases where either the weekday or the weekend diary is not reported. Second, we exclude diaries that describe a non-typical day.[18] Third, we keep only complete diaries and do not impute unassigned slots, with one exception: time periods between 10 p.m. and 6 a.m. that are missing are recoded as *sleeping or napping*, as in Fiorini and Keane (2014).[19] As a result of the above restrictions, we drop 15.3% of time diaries in CDS I, 14.7% in CDS II, and 13% in CDS III. Thus, we are left with complete diaries – those such that the duration of all the activities add up to 24 hours for one typical weekday and one typical weekend. The numbers of observations in our samples are 2,807 in CDS I, 2,520 in CDS II, and 1,424 in CDS III.

Since we have over 200 variables corresponding to the type of activity the child performs and 11 variables corresponding to with whom the child performs the activity, it is not feasible

[15] 89% of the primary caregivers are mothers.

[16] 10% of the interviews were done via phone.

[17] Researchers have found that young children's parents enjoy working with their child to complete the child's time diaries, and these diaries can adequately represent the child's day (Timmer et al. (1985)), but it is not clear whether that particular day is a representative sample.

[18] Respondents are asked whether each reported day is typical.

[19] We also follow them by recoding as *sleeping or napping* the time periods between 10 p.m. and 6 a.m. originally filled as "refused to answer".

to estimate the effect of every single combination of these two variables given the available sample size. Ideally, all else constant, more disaggregated categories of activities are preferred, as it better exploits the heterogeneity in the data and estimates more interpretable parameters. However, as the categories of activities become increasingly disaggregated, the estimated parameters of interest become less precise. We choose to categorize children's activities into two general types of activities, namely active and passive.[20] Active (passive) activities include all activities in which the child actively (passively) participates. The activities that we recode as active are *taking lessons (e.g., dancing), reading, socializing, active leisure, household chores, jobs, school/day care, and organizational activities.* In contrast, the activities we recode as passive are *obtaining goods and services, traveling/waiting, using computers, watching TV, passive leisure, and personal needs and care.*[21]

We categorize with whom the child performs the activities into seven groups of people: mother, father, grandparents, siblings, friends, others (i.e., someone other than the first five groups), and self. In reality, a child could perform an activity with many different people at the same time. Whenever the child is with more than one person within the same time slot, we assign the slot to the partner in the following order: mother, father, grandparent, sibling, friend, and others.[22] Finally, we also add two other categories: *refuse to answer or do not know* and *sleeping or napping.*[23] We choose *sleeping or napping* as the omitted time input in our estimation, so that all our reported results should be interpreted as a substitution between a given activity and *sleeping or napping.*

Our decision to categorize child activities primarily according to activity partners reflects three ideas. First, it is not possible for parents to micromanage the precise activities that 10 to 18 year old children participate in on a daily basis. For example, if a child is sent to their room to do homework, parents typically do not monitor every minute expended. However, parents can readily manipulate with whom their children spend time, whether it be family, friends, or no one at all. Thus, it is useful for parents to understand what these partner decisions may mean for the production of skill. Second, the productivity of any specific activity will likely depend upon the activity partner. Indeed, there are many potentially important intangibles in an interaction with a child (e.g., altruism, expectations, trust, power asymmetries, etc.), and these intangibles are likely to be different depending

[20]We choose these labels rather than "quality" / "non-quality" or "productive" / "non-productive" because they seem to describe more objectively the type of task the child is performing, and do not necessarily reflect our expectations about how the production function should look like.

[21]A full description of our recoding rules is available upon request.

[22]Our results are robust to changes in this order.

[23]Following Fiorini and Keane (2014), we distinguish "refused to answer or do not know" (which we include in our sample) from the case where an activity is missing (which we exclude from our sample).

on who is interacting with the child.[24] Finally, differentiating active and passive time by activity partners also allows us to capture unobserved heterogeneity within activity type, as much of it could be endogenous. Indeed, each of these partners has different preferences and opportunity costs of time, which may influence activity choices both within and across input categories. Our data suggests that a substantial amount of unobserved heterogeneity is controlled this way. Figure 1 shows the actual activity composition of active/passive time spent with parents and grandparents in our data, suggesting a large difference within categories.[25] For example, mothers spend 10.2% of their active time with child on sports, compared to 24.3% for fathers. In contrast, mothers spend 29.1% of their active time on educational activities with the child, compared to 21.7% for fathers. Further, grandparents spend a higher proportion of active time with the child in socializing and organizational activities relative to parents. Figure 2 shows that adult family members also tend to engage with the child in the activities differently too.[26] For instance, grandparents are more likely to participate with the child in the activities, while parents are more likely to just be around the child while she does the activities.[27]

An alternative approach to modeling heterogeneity in time inputs would be to aggregate less along the activity margin and more along the partner dimension. However, in Section 3.4 we provide suggestive evidence that more potential confounders are absorbed with this specification of inputs than with other specifications in the literature. In Section B of the appendix, we attempt to bridge our specification of inputs with other specifications in the literature, showing that our results are not sensitive to the unobserved heterogeneity absorbed by alternative categorizations of inputs, such as the ones implemented by Fiorini and Keane (2014) and Del Boca et al. (2013).

2.2 Summary Statistics

2.2.1 Children's Time Allocation

In this section, we describe children's time allocation using the recoded activity categories as described above. We construct a weekly measure for each time input by multiplying the weekday hours by 5 and the weekend day hours by 2, and then adding up the total hours.

Table 2 shows the weekly distributions of time (in hours). *Sleeping or napping* is the most popular activity in our sample, as expected. Children also spend a lot of passive time

[24] For instance, Del Boca et al. (2013) find that maternal and paternal time have heterogenous impacts on skill, particularly when children are young.

[25] Figures D.1 and D.2 in the appendix illustrate the activity composition for all other activity partners.

[26] Activity engagement for other partners is presented in Figure D.3 in the appendix.

[27] The definitions of participating and being around are the same as in Del Boca et al. (2013).

with their mother, and a lot of active time by themselves. This is not surprising as a ten- to eighteen-year-old child is likely to spend a large amount of time with her mother on activities such as watching TV, and a lot of time at school.[28] Across the 16 time categories, active time with grandparents has the smallest mean, which is half an hour per week. Children in our sample are more likely to spend active and passive time with their mother than with their father. Importantly, almost every input category has a sizable mass of respondents reporting zero minutes.

2.2.2 Children's Skills, Demographics, and Parental Background

In this section, we discuss other variables in the data that are relevant to our analysis. We start by describing the children's skill variables that we use as outcomes in our models. PSID-CDS children aged 3 and older are evaluated using the Woodcock-Johnson Revised Tests of Achievement (WJ-R), Form B (Woodcock and Johnson (1989)). In 1997, children aged 3-5 are administered Letter-Word Identification and Applied Problems sub-tests. Children aged 6 and above receive Letter Word and Passage Comprehension sub-tests as well as Applied Problems and Calculation sub-tests. In the 2002 and 2007 waves, these tests are re-administered, with the exception of the Calculation sub-tests. Since the Calculation sub-test is only administered for the 1997 wave, we do not include it as one of our skill measures. Thus, we use standardized versions of Letter Word, Applied Problems, and Passage Comprehension as our child cognitive skill measures.[29] In the following sections we refer to these scores as Vocabulary, Math, and Comprehension.

Noncognitive skills are measured through parental assessment. In all three waves, the primary caregiver is asked questions about the child's behavioral problems. Twenty-six questions are used to measure the child's behavioral problem scale, and ten other questions are asked about the positive aspects of children's lives, including obedience/compliance, social sensitivity, persistence and autonomy. With these thirty-six questions, we construct a measure of noncognitive skills by using iterated principal factor analysis, similar to Cunha and Heckman (2008) and Fiorini and Keane (2014). In Table D.1 in the appendix we show the rotated factor loadings. The factor loadings are all above 0.26 and stable across the two waves. The constructed measure is standardized to have mean zero and standard deviation one and is ordered so that a higher score means better noncognitive skills.

[28] 88% of self active time consists of time at school. We incorporate school time into self active time because it is difficult to determine with whom the child spends the bulk of their time at school. As discussed in Section B of the appendix, our results are robust to splitting self active time into two categories of inputs, one incorporating school activities and the other incorporating the other activities comprising of self active time.

[29] We do not use the standardized scores provided by the PSID-CDS. Instead we standardize the raw score of each skill measure to have mean 0 and standard deviation 1.

The PSID-CDS collects extensive information on the child, her household, as well as her school environment. In Table 3, we present demographic and parental background statistics for a few selected variables. Child characteristics are presented in rows 1 to 4, parental characteristics are presented in rows 5 to 12, and environmental characteristics are presented in rows 13 to 16. On average, children in the sample are the second child to her mother, and more than 50% live with both biological parents. Children's annual family income is above the US median income in 2007.[30]

3 Empirical Approach

In this section, we discuss our empirical approach, paying special attention to how we implement the test of exogeneity in our setting. See Caetano (2015) for the formal description of the test in the univariate context.

We are interested in assessing whether we can consistently estimate β via OLS in the following equation:

$$\text{Skill}_i = \text{Input}_i\beta + \text{Control}_i\pi + \text{Error}_i, \tag{1}$$

where i denotes a child. Skill_i refers to a particular skill of the child (e.g., mathematics skill), as measured by standardized assessment scores. Input_i refers to a vector of all activities done by the child in hours per week, whose jth element is denoted as Input_i^j (e.g., active time spent with the mother). Control_i refers to covariates added to absorb confounding factors. These covariates may or may not be actual additional inputs in the production function (Todd and Wolpin (2003) termed such an equation a "hybrid production function"). Finally, Error_i refers to the unobserved determinants of Skill_i that are not absorbed by covariates.

In this context, a "model" is defined as a unique combination of $(\text{Skill}, \text{Input}, \text{Control})$ in equation (1) for precise definitions of Skill, Input and Control.[31] We can consistently estimate $\beta := (\beta^1, ..., \beta^J)'$ via OLS in model $(\text{Skill}, \text{Input}, \text{Control})$, described in equation (1) if:

Assumption 1. *$Cov\left(Error_i, Input_i^j | \underbrace{Input_i^{-j}, Control_i}_{Covariates_i^j}\right) = 0$, for all j, where $Input_i^{-j} := (Input_i^1, ..., Input_i^{j-1}, Input_i^{j+1}, ..., Input_i^J)$.*

[30] Household annual income is adjusted to 2007 dollars.

[31] For robustness, we also consider different types of models in the paper, including ones where skills are a non-linear function of inputs.

Our approach consists of testing Assumption 1 (jointly for all j) in all feasible models. In models that survive the test, we conclude that, at the same time for all j, all confounding factors that would bias β^j are absorbed by $\text{Covariates}_i^j := \left(\text{Input}_i^{-j}, \text{Control}_i\right)$. Thus, we can plausibly interpret the $\hat{\beta}^{OLS}$ estimated from the models that survive the test as causal effects of time inputs on skills. Of course, the credibility of this approach depends crucially on the capability of the test to detect potential endogeneity. In the rest of this section, we explain the test and discuss in detail the types of endogeneity the test can and cannot detect in our context.

3.1 Testing the Exogeneity Assumption

The test of exogeneity relies on the observation that many children choose to spend zero minutes doing certain activities. When they choose to spend zero minutes in an activity, they may be in a "corner solution": they may desire to choose negative amounts of that input but cannot. Because the time inputs that should matter in the production function of skills are the inputs **actually chosen**, not the **desired** inputs, we can formulate a test of exogeneity by exploiting this excess variation in the desired time input holding constant the actually chosen input at zero minutes.

We explain the intuition of this test in Figure 3, which illustrates the correlation between a generic time input and a generic child skill across all children in the sample. The goal of the test is to understand whether part of this correlation can be interpreted as causal. The discontinuity shown in Panel (a) must be the result of either the observed covariates or unobserved confounders varying discontinuously when the time input is equal to zero.[32] As shown in Panel (b), conditional on all observed covariates $\left(\text{Input}_i^{-j}, \text{Control}_i\right)$, the discontinuity remains. The remaining discontinuity in Panel (b) must be the result of unobserved confounders that are not absorbed by the covariates, so Assumption 1 is rejected for this model.

The statistical power of the test comes from the assumption that unobservables vary discontinuously when a time input is zero. Below we show empirical evidence supportive of that; but first, we discuss why this is the case. Unobservables are likely discontinuous at zero in our context because observations are bunching at a threshold, leading to a "corner solution". For instance, consider a generic unobservable "mother type", which helps determine the skills of a child. Figure 4 illustrates how the average mother type varies depending on

[32] Of course, we rule out the possibility that the main effect is discontinuous at zero in equation (1). This implicit assumption, also made in all papers in the literature, is plausible in our context (e.g., a minute of reading a book should not affect the child's skill). Another reason this assumption seems innocuous in our context is that we cannot reject the null hypothesis of continuity for models with a detailed enough list of covariates.

the level of time spent reading books to the child. Mothers who read less to their child tend to have a lower type, as illustrated in the figure. However, something unique happens at zero. The mothers who read zero minutes to their child are discontinuously different from the mothers who read a little to their child. The reason is that among mothers who read zero, there are some whose type is so low that if possible they would have read negative amounts of time to their child. In this example, if mother type is not fully absorbed by covariates, then $E\left[\text{Skill}_i|\text{Input}_i^j = x, \text{Covariates}_i^j\right]$ will be discontinuous at $x = 0$, which explains the discontinuity found in Panel (b) of Figure 3. Each time input can elicit many such unobservable confounders; if covariates do not fully absorb them, then endogeneity will be detected.

We implement the test by adding the vector D_i to equation (1):

$$\text{Skill}_i = \text{Input}_i\beta + \text{Control}_i\pi + D_i\phi + \text{Error}_i', \tag{2}$$

where $D_i := \left(d_i^1, ..., d_i^J\right)$, $d_i^j := 1_{\{\text{Input}_i^j=0\}}$. The vector D_i allows for a discontinuity in $E\left[\text{Skill}_i|\text{Input}_i^j = x, \text{Covariates}_i^j\right]$ at $x = 0$ for each j. We implement an F-test for whether $\phi = 0$, which tests for the null hypothesis that $E\left[\text{Skill}_i|\text{Input}_i^j = x, \text{Covariates}_i^j\right]$ is continuous at $x = 0$ for all j jointly. This test is equivalent to testing whether Assumption 1 holds (Caetano (2015); Caetano and Maheshri (2016)).

3.2 Evidence of Bunching

The test described above exploits the potential bunching of observations resulting from a non-negative time constraint. Here we show empirically that observations do indeed bunch at zero time input threshold. Figure 5 shows the cumulative distribution function (CDF) of various activities. The fact that the CDFs cross the vertical axis away from the origin is direct evidence of bunching, as it shows that the probability density function is discontinuous at zero (McCrary (2008)). Moreover, the CDFs are smooth away from zero, suggesting that there is no bunching elsewhere. Table 2 shows the proportion of observations with zero time inputs for all inputs, providing evidence that other activities have similar distribution functions. In fact, each child in our sample chooses to spend zero minutes in at least two of our 16 aggregated activities. As discussed below, the fact that the non-negative time constraint binds so often is key to perform a powerful test to detect endogeneity.

Note that the bunching of inputs is not necessary for the test to work; it is sufficient that unobservables are discontinuous at the threshold for whatever reason. Nevertheless, the evidence of bunching is suggestive of the "corner solution" intuition developed above: unobservables should be discontinuous because people cannot choose negative amounts of time for any child related activity.

3.3 Sources of Detectable Endogeneity

While the bunching evidence above indicates that the proposed test should have power, the test may not have power to detect all sources of endogeneity. To structure the discussion, we write the following general model of skill production:

$$\text{Skill}_i = f(\widetilde{\text{Input}}_i, \text{Other}_i) \tag{3}$$

where $\widetilde{\text{Input}}_i$ is a vector of $\tilde{J}$ activities, defined at a very detailed level, and Other_i is a vector that includes all other inputs in the production function. Elements of $\widetilde{\text{Input}}_i$ in this generic framework are defined precisely by a unique combination of all its features. For instance, reading different books, or reading different pages of the same book, refer to different time inputs. The production function $f(\cdot)$ is unrestricted. This is a "general" model in the following sense: if we observed all elements of $\widetilde{\text{Input}}_i$ and Other_i, and if we were able to estimate a non-restrictive $f(\cdot)$, then we would be able to identify the causal partial effect of $\widetilde{\text{Input}}_i$ on Skill_i. Although data limitations preclude estimation of this general model, it is nevertheless useful because it provides a framework for discussing a comprehensive list of endogeneity concerns that might arise when we deviate from this ideal scenario.

Assumption 1 essentially combines all of the simplifying assumptions that are needed to go from the general production function, equation (3), to the OLS specification we aim to estimate, equation (2). For example, Assumption 1 includes assumptions about linearity, additive separability, and that Control_i is sufficient to account for Other_i. A failure of any of these assumptions will imply the existence of a variable w_i that may bias our main estimates if it is not absorbed by covariates. If this variable w_i is discontinuous when $\text{Input}_i^j = 0$ for some j and is not fully absorbed by covariates, then this discontinuity will be captured by D_i, leading to a rejection of the model.

Figure 6 shows a few examples of *potentially* omitted variables w_i that are likely elements of Other_i, where $E[w_i|\text{Input}_i^j = x]$ is discontinuous at $x = 0$ for some j. Each plot shows $E[w_i|\text{Input}_i^j = x]$ along with a local cubic fit, where w_i is denoted in the title, and Input_i^j is denoted in the horizontal axis.[33] These plots indicate that the exogeneity test has power to detect whether we incorrectly omitted w_i: if w_i affects skill conditional on covariates then

[33] At $x = 0$, we also show the 95% confidence interval. For $x > 0$, the scatter plot aggregates to the next hour of the time input. The shaded region represents the 95% confidence interval for the fit with an out-of-sample prediction at zero minutes. For the local fit, we use the Epanechnikov kernel with the rule-of-thumb bandwidth for the kernel and 1.5 times the rule-of-thumb bandwidth for the standard-error calculation. Results are robust for different choices of kernel and bandwidths.

skill will be discontinuous at $\text{Input}_i^j = 0$. As an example, Panel (a) of Figure 6 indicates that lagged math score is statistically different across children who spend zero and positive amounts of passive time with friends. If lagged math score affects child skill conditional on covariates and lagged math score is excluded from the model, then the indicator variable for zero passive time with friends will absorb the effect. In this case, the exogeneity test would lead to a rejection of the model. Panels (b)-(d) in Figure 6 indicate discontinuities in the number of siblings, number of children born to mother, and household income when various time inputs are zero, expanding the list of potential elements of Other_i whose omission the test has power to detect. Additional examples are provided in Figures A.5 and A.6 in the appendix.

The w_i we consider in Figure 6 are observable, so they can in principle be incorporated in Control_i to avoid any endogeneity stemming from their exclusion. However, if observables are discontinuous at zero, then unobservables are also likely to be discontinuous at zero. For instance, the discontinuity in household income suggests that unobservables such as child good expenditures, child's schooling environment, etc, might also be discontinuous.[34] If these variables are not fully absorbed by covariates, then their discontinuity will be captured by D_i, leading to a rejection of the model.

Other sources of endogeneity that arise when moving from the general production function to our OLS specification include simultaneity, measurement error, and model misspecification. However, these sources of endogeneity can essentially be framed as an omitted variable bias problem. To the extent that these omitted variables are discontinuous at zero for some input j, then our test has power to detect them, similar to the logic described above. We formalize these ideas in detail in Section A of the appendix and provide evidence that unobservables related to simultaneity, measurement error, and misspecification are likely discontinuous when a subset of our time inputs equals zero. Thus, we believe our test is capable of detecting endogeneity from a comprehensive list of potential sources in the context of child skill development.

3.4 Which confounders cannot be detected by the test?

While we are confident that we can detect endogeneity arising from omitted variables that are discontinuous at zero for some input j, there remains a set of confounders that the test is unable to detect. Consider the set of all potentially endogenous variables w, characterized by being correlated to both Input_i and Skill_i. If any such w is not absorbed by covariates,

[34] The argument that similar patterns of discontinuity in observables should also be found in unobservables is analogous to the one made by researchers about continuity when implementing regression discontinuity designs.

Assumption 1 in the context of Equation (1) will be violated. These variables fall under two categories: (a) those that vary discontinuously at $\text{Input}_i^j = 0$, and (b) those that vary continuously at $\text{Input}_i^j = 0$. Thus far, we have focused our discussion on confounders of type (a) and our ability to detect them with our test. However, endogenous variables of type (a) can actually be further subdivided in two types: (a1) those that are correlated with Skill_i when $\text{Input}_i^j = 0$, and (a2) those that are uncorrelated with Skill_i when $\text{Input}_i^j = 0$. These three types of confounders, (a1), (a2), and (b), form a partition: any confounder w_i must be of one and exactly one of these three types. The exogeneity test described above can detect all confounders of type (a1), but cannot detect confounders of types (a2) or (b). We now discuss how likely are confounders to be of each of these types in our context.

Conceptual Argument

To help frame our argument, Figure 7 illustrates examples of confounders of types (a) and (b). The solid black lines (and points) in the figure correspond to Input_i^j, while the dashed black lines correspond to $\text{Input}_i^{j\star}$, where $\text{Input}_i^{j\star}$ is the optimal choice of input j by individual i when not bounded by a non-negativity time constraint. Of course, $\text{Input}_i^{j\star} = \text{Input}_i^j$ for $\text{Input}_i^j > 0$, but $\text{Input}_i^{j\star} \leq \text{Input}_i^j$ for $\text{Input}_i^j = 0$. This occurs because people cannot spend a negative amount of time on an activity.[35]

Panel (a) of Figure 7 distinguishes between confounders of types (a1) and (a2), both of which vary discontinuously at $\text{Input}_i^j = 0$. In this example, the red range along the right side of the vertical axis is the support region of the confounder for the whole sample, while the blue range along the left side of the vertical axis is the support region of the confounder for the subsample of observations such that $\text{Input}_i^j = 0$. A confounder, by definition, must be correlated to Skill_i in the red range. A confounder is of type (a1) if it is correlated to Skill_i in the blue range, while it is of type (a2) if it is not correlated to Skill_i in the blue range. The evidence of bunching shown above suggests that type (a2) is unlikely, since a significant portion of the sample is such that $\text{Input}_i^j = 0$. Moreover, the redundancies implied by the multivariate test are particularly helpful in this case because the blue range of the same confounder will vary across inputs, allowing the test to cover more of the support of any confounder.

Panel (b) of Figure 7 depicts a confounder of type (b). In this case, the average of the confounder when $\text{Input}_i^{j\star} \leq 0$ has to be equal to the corresponding average for observations where $\text{Input}_i^{j\star} = 0$. This is implausible because the confounder is by definition correlated

[35]Note that Input_i^j, rather than $\text{Input}_i^{j\star}$, should be included as inputs in the production function, since we want to identify the effect of the *actual (not the desired)* time spent on activities. Thus, we do not have a censored model, we only have a corner solution model. Wooldridge (2002) discusses this distinction in detail.

to Input_i^j and there are many observations such that $\text{Input}_i^{j\star} < 0$ (as per the bunching evidence shown in Section 3.2). Indeed, the discontinuity plots discussed in the previous section suggest that confounders of type (a) are much more likely to occur.

We now provide further evidence that confounders of type (a1) should be norm in our context.

Empirical Evidence

Below, we measure the relative frequencies of different types of confounders using an intuitive diagnostic procedure. First, we construct an exhaustive set of observed variables as our pool of potential confounders.[36] For each candidate confounder, we test whether it is an actual confounder by performing an omitted variable test. Specifically, for each candidate observed confounder, we compare the estimates of all time input coefficients through an F-test when we include the variable versus when we exclude the variable in a model with only time inputs as regressors (i.e. no controls).[37] In turn, each variable in the subset of actual confounders is assigned to type (a1) if it is discontinuous at zero for some time input, or type (b) if it is not.[38] Table 4 shows the results of this procedure. At a 5% (10%) significance level, 23%-24% (26%-27%) of 98 variables are empirically found to be confounders in cognitive skill models. Moreover, almost all confounders are of type (a1) (i.e. on average, 96% at 5% significance level and 95% at 10% significance level). This exercise provides strong empirical evidence that confounders of types (a2) and (b) are unlikely in our case.[39] In contrast, our results for noncognitive skill suggest that the test has little power to detect endogeneity in noncognitive skill production function specifications. At a 5% (10%) significance level, only 2% (3%) of 98 variables are empirically found to be confounders in noncognitive skill models (although all of them are of type (a1)).

Our multivariate setting adds redundancies that contribute to almost all confounders being of type (a1), since a confounder of type (a2) or (b) for input j can be of type (a1) for input j'. All observations with $\text{Input}_i^j > 0$ are such that $\text{Input}_i^{j'} = 0$ for some j', for all j, reducing the possibilities of confounders of type (a2). Similarly, all observations with

[36]The set of variables contains all potential confounders that are observed in our data, including lagged test scores, lagged time inputs, child characteristics, parental characteristics, family environmental characteristics, school environmental characteristics, school experience as well as variables related to misreporting of time diaries.

[37]In doing so, we include the variable not only linearly, but also in many other forms. For instance, for a variable w we include as controls w, w^2, w^3, w^4, $\frac{1}{1+w}$, $\sqrt{w}$, and $log(1+w)$.

[38]In practice we do not find any type (a2) confounders. This is due to the fact that for each observation in our sample, at least two of the 16 time inputs are equal to zero.

[39]Of course, this statement only refers to observables. Our results in Section 3.3 implicitly suggest that our set of observables are somewhat representative of the set of unobservables in an important way: for each source of endogeneity we can conceive, we find examples of observable confounders reflecting that source.

$\text{Input}_i^j = 0$ are such that $\text{Input}_i^{j'} = 0$ for some $j' \neq j$, for all j, reducing the possibility of confounders of type (b). Thus, it is not surprising that almost all confounders are detectable in our context.

Despite the improbability of confounders of types (a2) and (b), we pursue in Section B of the appendix an extensive set of robustness checks designed to detect them.

Remark 1. Our approach is not more helpful than standard approaches in verifying whether there is important unobserved heterogeneity in a given model. Rather, our approach improves over standard approaches only in verifying whether this unobserved heterogeneity leads to endogeneity. For instance, a lack of endogeneity due to heterogeneous effects (e.g., Input_i and Other_i are non-separable) does not imply a lack of heterogeneous effects. It only implies that our estimate should be an unbiased estimate of the weighted average of different heterogeneous effects, where weights are given by the distribution of Other_i in the data. For instance, it may be that an additional hour of passive time alone has a positive effect for a wealthier child and a negative effect for a poorer child (e.g., the TV program the child is actually watching may be completely different for these two types of children). In this case, the estimates of a model where endogeneity cannot be detected by the test of exogeneity could be zero. Indeed, they should be an unbiased weighted average of one additional hour on passive time alone across all children, where weights are given by the proportion of wealthier and poorer children.

4 Results

We start by proposing a set of regression models that we can plausibly estimate given the data described in Section 2. As explained, a "model" is defined as a unique combination of (Skill, Input, Control) in equation (1), where Skill $\in$ {math, vocabulary, comprehension, noncognitive skills}, Input $\in$ {sleeping or napping, active time with "companion", passive time with "companion", don't know or refuse to answer}, and companion $\in$ {mother, father, grandparents, siblings, friends, others, self}. We now describe the models we consider and present our results.

4.1 Exogeneity Tests and Skill Production Estimates

We consider only value-added models, which have been standard in the literature, by including in all specifications the value of Skill_i observed in the previous wave. Thus, our estimation sample includes children between the ages of 10 and 18 from the 2002 and 2007 waves of the CDS. We consider a sequence of six specifications of the value-added model, where each specification includes a richer set of controls than the previous one. We take this approach for two reasons. First, it illustrates that our exogeneity test has power to detect endogeneity in the most parsimonious models. Second, it helps identify the key controls that absorb important sources of endogeneity.

The details of each specification are as follows. Specification (1) has no controls other than the corresponding lagged skill. Specification (2) adds child characteristics, such as age, gender, and race. Specification (3) adds mother demographic characteristics, such as age, education level, and age at child's birth. Specification (4) adds family demographic characteristics, such as father's age, whether the child lives with biological parents, and household annual income. Specification (5) adds family environmental characteristics, such as whether the child has a musical instrument at home and whether the child's neighborhood is safe. Specification (6) adds school characteristics and school experience, such as whether the child is in a public or private school, and whether the child has ever attended a gifted program.[40]

Table 5 shows the exogeneity test F-statistic and corresponding p-value for each model we consider. The F-statistics and p-values in bold represent the surviving specifications, i.e., specifications that we are not able to reject exogeneity at the 10% significance level. In specification (1), we reject exogeneity for all cognitive skills, which provides direct evidence that our test has power to detect endogeneity in the basic value-added model we consider, complementing the evidence shown in Section 3.3. For noncognitive skills, we fail to reject exogeneity in specification (1). This is consistent with the analysis in Table 4 showing that there are essentially no observable confounders in the noncognitive skill model. One interpretation of these findings is that time inputs are not endogenous in a model of noncognitive skill production. Alternatively, we may simply lack power to detect such endogeneity. Because Table 4 indicates that our test lacks power for noncognitive models, we take the

[40]Here is a full list of the control variables included for each category. Child characteristics: child's age, child's age squared, child's gender, child's race indicators, birth order to mother, born in the US indicator, child's grade indicator, and child's BMI. Mother demographic characteristics: mother's education in years, mother's current age, mother's current age squared, and mother's age at child birth. Family demographic characteristics: father's education in years, father's current age, father's age at child birth, mother's marital status at child birth, household annual income (in \$10,000s), number of siblings child lives with, indicators of whether child lives with biological parents, and indicator of whether child lives with grandparents. Family environmental characteristics: spending on tutoring programs (in \$100s), spending on extra-curricular lessons (in \$100s), spending on school supplies for the child (in \$100s), spending on clothes for the child (in \$100s), indicator for whether child has a musical instrument at home, indicator for whether child has a desk at home, indicator for whether child has a working TV at home, rating of neighborhood safety, rating of neighborhood quality, number of books mother read in the previous year, mother's working hours per week, and mother's working days per week. School characteristics: indicator for whether child ever attended a private school, indicator for whether child has ever attended a gifted program, and number of school changes since the beginning of the school year.

conservative interpretation and do not present the coefficient estimates for noncognitive skill formation in the text (see Table D.2 in the appendix).

For different cognitive skill measures, the specification of Control that results in a failure to reject exogeneity is different. For example, the child's observed characteristics, together with their lagged skill and all time inputs, are enough to absorb any confounder in the production function of **vocabulary** skills. To absorb the confounders for **comprehension** skills, mother's demographic characteristics are necessary. In contrast, **math** seems to be a more complex production process, as we fail to reject only models that include observed child and family demographic characteristics. Family demographic characteristics (i.e. specification (4)) lead to a jump in p-value for **math** skills (i.e. p-value goes from 0.064 to 0.177), which is suggestive of the importance of family demographic characteristics in absorbing endogeneity. Thus, Table 5 suggests that different groups of control variables are playing different roles in absorbing endogeneity depending on the skill in question.[41] We are unable to reject exogeneity in specifications (4), (5) and (6) for all four skills.

Our identification strategy is based on the premise that any confounder will be absorbed as we add controls, otherwise the test of exogeneity will detect its presence. However, this might not be the case if the standard errors of $\hat{\phi}$ also increase with the addition of controls. In that case, a discontinuity in $\mathrm{E}\left[w|\text{Input}_i^j = x\right]$ at $x = 0$ would be wrongly interpreted as continuous; i.e., confounders of type (a) would be erroneously understood to be confounders of type (b). To check if this is the case, we present the distribution of the standard errors of all elements of $\hat{\phi}$ for all combinations of specifications and skills in Figure D.4 in the appendix. In practice, the standard errors do not seem to increase as more detailed controls are added in the specifications that we consider. This is not surprising since the addition of controls is simply an addition of incidental parameters to the regression, so it does not necessarily affect inference on the parameters of ϕ, which remain fixed across all specifications.

Table 6 presents the estimated coefficients of time inputs from a surviving specification (specification (6)) for the three cognitive skill measures.[42] We find that for **math** skills, active time with mother, active time with father, active time with grandparents, active time with friends, self active time, passive time with mother, passive time with siblings, and self passive time are statistically significant. Active time with grandparents is the most productive input: one more hour a week spent on active time with grandparents rather than

[41]Note that in exercises not reported here, we vary the order in which we add controls. The importance of each group of variables in accounting for endogeneity is similar to what is observed in Table 5. A full description of the permutation exercises we performed is available upon request.

[42]Note that we also fail to reject in specifications (4) and (5) across the three cognitive skill measures. The estimated effects of time use are nearly identical across specifications (4), (5), and (6) for each skill type. Moreover, all subsequent robustness exercises are consistent across specifications.

on sleeping or napping would increase the math test score by 2% of a standard deviation, while one more hour a week spent on active time with mother rather than sleeping or napping would increase test score by about 0.5% of a standard deviation. It is also noteworthy that active time with friends is as productive as active time with mother. Although we find that parental inputs have an impact on math skills, there is little to no effect on vocabulary skills. In contrast, we find that active time with grandparents has a statistically significant effect on child **cognitive** skills generally **(i.e. math, vocabulary and comprehension**).

The coefficients in Table 6 indicate the impact of each input on skills relative to sleeping. However, by comparing the coefficients with each other we can comment on the relative effectiveness of various inputs. For example, substituting an additional hour per week of active time with the father for active time with others would increase math scores by 1.4% of a standard deviation (with a standard error of 0.7%).

Although plausibly causal, it is difficult to compare these estimates across partners because of our need to aggregate time inputs. Our pie charts in Figure 1 (also Figures D.1 and D.2 in the appendix) allow us to better interpret these estimates. These heterogeneous effects by partner may reflect the different activities and quality of engagement of partners for a given category. Indeed, among the active activities, grandparents tend to spend more time with children playing, socializing, engaging in organization activities (e.g. volunteer work), and doing arts and crafts, compared to parents.[43] Grandparents are also more likely to actively participate in an activity rather than just being around the child (see Figure 2).[44] These compositional differences within aggregated categories may explain why active time with grandparents is found to be more productive than active time with parents.

While the estimated impact of time inputs on skill development is our main focus, it is important to note that the coefficient estimates for the control variables are consistent with the rest of the literature. Maternal education and family income are associated with higher cognitive skill levels. Additionally, boys tend to score higher on math, while girls tend to score higher in vocabulary and comprehension. The full set of model coefficients can be found in Section D of the appendix. Note that we cannot interpret these estimates causally since we are unable to formally address any endogeneity concerns associated with the control variables.

Our main specifications assume that the effects of time inputs on child skills are linear, but there can be interesting hidden heterogeneity in the results. In Tables C.1 and C.2 in

[43]For example, grandparents on average spend 15.8% of active time with child on socializing, whereas mother spends 7% and father spends 5.1%.

[44]Grandparents participate in activities 59% of their active time with the child. In contrast, mother participate only 36% of her active time with the child and father participate 43.7% of his active time with the child.

the appendix, we present estimates from a linear B-spline specification in order to allow for non-linear treatment effects:

$$\text{Skill}_i = \sum_j f^j(\text{Input}_i^j) + \text{Control}_i\pi + D_i\phi + \text{Error}_i', \tag{4}$$

where $f^j(\cdot)$ is a linear B-spline function of Input_i^j with parameters β^{jk}, $k = 1, 2, 3$, representing the linear effect within equally frequent intervals of the distribution of Input_i^j. Generally, we find that the specifications that survive the exogeneity test in the linear model also tend to survive the exogeneity test in the B-spline model, and vice-versa. Thus, most of the power of the test seems to stem from discontinuous unobservables rather than model misspecification, otherwise the B-spline models would fail to reject in even the most parsimonious specifications.

4.2 Comparing Surviving and Non-surviving Specifications

Thus far, we have chosen appropriate models for causal inference purely based on the exogeneity test described in Section 3. However, there can be confounders that are not detectable by the test. As discussed in Section 3.4, there are two potential categories of confounders: (a) confounders that are discontinuous at $\text{Input}_i^j = 0$, and (b) confounders that are continuous at $\text{Input}_i^j = 0$. Among type (a) confounders, there are two subtypes: (a1) those that are correlated with skill at $\text{Input}_i^j = 0$, and (a2) those that are not. The exogeneity test introduced in Section 3.1 is capable of detecting all unobservables of type (a1), but is incapable of detecting unobservables of types (a2) or (b).

As discussed in Subsection 3.4, there are a number of reasons to believe that the class of variables included in types (a2) and (b) is small in our context. Regardless of how implausible the existence of these variables might be, this subsection provides one common robustness check that can in principle detect them if they exist.

We compare estimates of β across specifications, irrespective of whether the specification survives or does not survive the test, as shown in Section 4. This comparison is often done in empirical studies, where, heuristically, a good model is one that provides estimates that are robust to added controls (which might be omitted variables in the model).[45] This "test of stable coefficients" is in principle capable of detecting endogeneity from the two undetectable sources of endogeneity discussed above. Indeed, added controls may partly absorb (both at $\text{Input}_i^j = 0$ and at $\text{Input}_i^j > 0$) confounders of type (a2) or (b), leading to a change in the main estimates. If a model survives the test of exogeneity, but does not survive this

[45] For instance, Fiorini and Keane (2014) implement a somewhat weaker version of this test whereby they compare whether the ranking of the magnitude of each coefficient is the same across specifications.

test, then it is evidence that the test of exogeneity did not detect some important source of endogeneity.

We test for whether the fifteen elements of β in each specification (1)-(5) from Section 4 are jointly significantly different from the corresponding coefficients in specification (6), our preferred model. We present the p-value of this test for each skill measure in Table 7. Numbers in bold refer to those specifications that survive the exogeneity test at the 10% level of significance. In general, specifications that survive the exogeneity test (in bold) also survive the test of stable coefficients (p-value $> 10\%$). Across all models, only one model that survives the exogeneity test is rejected by the other test: specification (3) for comprehension in Table 5. This suggests that confounders from the undetectable sources of endogeneity discussed above are only controlled for after family demographic characteristics are added as controls (specification (4)). Conversely, no models do not survive the exogeneity test but survive the other test. From specification (4) onwards, all specifications survive both tests for all skills. Overall, these results are consistent with the idea that, as we add controls from specifications (1) to (6) in Section 4, we converge to the true causal estimates.[46]

In Section D of the appendix, we present the actual estimates for specifications (1)-(6) for each skill, for both the linear and the B-spline cases, illustrating more explicitly how the estimates are virtually unchanged for the surviving specifications but often change for the non-surviving ones. Finally, in Section B of the appendix, we perform an additional sequence of robustness checks aimed at detecting endogeneity resulting from confounders of type (a2) and (b). We again find little evidence of their existence.

5 Discussion

5.1 Why Does Selection on Observables Seem to Work in This Context?

The results for the linear and non-linear models discussed in the prior sections indicate that with rich enough controls we are able to arrive at specifications for which we fail to reject exogeneity. Moreover, as discussed in detail in the past sections, this does not appear to result from a lack of power with the exception of noncognitive skills. A natural question to ask at this point is why a selection on observables approach seems to be appropriate in the context of this application.

[46] Table C.3 in the appendix shows analogous results for the non-linear models discussed at the end of the previous subsection. All surviving specifications according to the exogeneity test also survive the test of stable coefficients with one exception.

While the richness of the available controls in the PSID is certainly helpful for mitigating endogeneity, incorporating the full set of inputs into the production function is also quite useful. To see this, consider the following simple model of input choices and skill formation where, for simplicity, we treat the child as the sole decision-maker. Skill for individual i is determined according to

$$\text{Skill}_i = f(\text{Input}_i, \theta_i),$$

where Input_i is a vector of J time inputs and θ_i is a vector of other inputs (i.e., Other_i in equation (3)) impacting skill which reflects any heterogeneity in the production function across children (e.g., how much attention the child pays when reading). Children choose Input_i to maximize utility

$$U_i = g(\text{Input}_i, \theta_i, \omega_i)$$

subject to $\text{Input}_i^j \geq 0$ and $\sum_{j=1}^{J} \text{Input}_i^j = T$, where T is the total available time (i.e., 24 hours per day). ω_i is a vector denoting heterogeneity in utility that is not associated with heterogeneity in skill production (e.g., how much the child enjoys reading). In this general formulation, skill and time inputs can in principle affect utility directly, as can the other inputs influencing the production of skill, θ_i.

Given this maximization problem, the chosen vector of time inputs, Input_i^*, is implicitly defined by the levels of θ_i and ω_i:[47]

$$\text{Input}_i^* = h(\theta_i, \omega_i)$$

so that individuals with different levels of (θ_i, ω_i) tend to choose different levels of the vector of inputs. For a given θ_i, the variation in inputs due to ω_i is not endogenous and is in fact precisely the type of variation we want to exploit when estimating the production function. Of course, although the component of ω_i that is orthogonal to θ_i would make ideal instruments to identify the effect of interest, it is difficult to know *ex ante* which source of variation is included in ω_i and which source of variation is included in θ_i, hence our need to develop an alternative identification strategy in this paper.

We can write $\text{Input}_i^{*,j}$ as

$$\text{Input}_i^{*,j} = h^j(\theta_i, \omega_i, \text{Input}_i^{*,-j}).$$

[47] Input_i^* represents $\widetilde{\text{Input}}_i$ in equation (3). For simplicity in the exposition, we assume no measurement error in this section.

In our context, endogeneity arises if an input is correlated with θ_i across individuals, conditional on covariates: $Cov\left(\text{Input}_i^{*,j}, \theta_i | \text{Input}_i^{*,-j}, \text{Control}_i\right) \neq 0$, i.e., if $h^j(\cdot, \text{Input}_i^{*,-j}, \text{Control}_i)$ varies with θ_i.

We conjecture that we are able to eliminate endogeneity and identify the effects of interest with our data for two reasons. First, to the extent that $\text{Input}_i^{*,-j}$ absorb elements of θ_i, adding them as covariates can substantially reduce the potential for endogeneity, requiring less of the vector Control_i. Second, as we add Control_i we are able to shut down any correlation between θ_i and $\text{Input}_i^{*,j}$ (conditional on $\text{Input}_i^{*,-j}$) before we shut down the correlation between ω_i and $\text{Input}_i^{*,j}$. The full set of controls incorporated in the empirical model must be unable to thoroughly absorb ω_i, otherwise there would be no independent variation remaining in Input_i to estimate the production function. ω_i reflects tastes and household constraints, which are likely quite heterogeneous across people, while θ_i is bound by technical features of the skill production technology. Thus, it is not surprising that covariates can fully control for θ_i without fully controlling for ω_i.

The above discussion illustrates a largely under-appreciated benefit of modeling the full vector of inputs in skill production. The inclusion of a comprehensive list of time activities not only enhances the interpretability of the production parameters, but can also substantially allay endogeneity concerns. Indeed, all else constant, $\text{Input}_i^{*,-j}$ helps absorb more confounders the more disaggregated inputs are. This is evident in the exercise conducted in Section 3.4, where we show that less than 30% of the observable variables we considered as potential confounders end up being confounders in a omitted variable test controlling for all inputs. In contrast, if we only include one time input, for example, active time with mother, the number of confounders essentially doubles.

5.2 What Can (and Cannot) be Inferred from Our Estimates?

In this paper, we estimate the average marginal productivity of each input on each skill. It is useful to interpret these estimates with the aid of the framework described above. We estimate $E[f_j(\text{Input}_i^*, \theta_i)]$ for each j, where f_j refers to the first derivative of the production function f with respect to its jth input, and the expectation is taken across all children i.

When $E[f_j(\text{Input}_i^*, \theta_i)] > 0$, we conclude that on average children will see an improvement in skill if they decide to spend more time on activity j (relative to sleeping), in comparison to their current time. However, that does not necessarily imply that children *should* spend more time on activity j. Indeed, children and their families likely make time allocation choices in order to maximize utility, not skill. To illustrate the implications of this, we show how different children and their parents might choose different levels of time inputs, and how these different choices might lead to different estimates of $f_j(\text{Input}_i^*, \theta_i)$. Assume that children and their parents care about skill (f), non-skill (u), and costs (c) such that

$$U_i = f(\text{Input}_i, \theta_i) - nc(\text{Input}_i, \theta_i, \omega_i)$$

where $nc(\text{Input}_i, \theta_i, \omega_i) := c(\text{Input}_i, \theta_i, \omega_i) - u(\text{Input}_i, \theta_i, \omega_i)$ represent the utility cost net of non-skill benefits, which is allowed to be heterogenous across different time investments. Intuitively, one can think of c as representing the component of utility related to "costs" and u as representing the component of utility related to "fun", although u can be interpreted more generally to also encompass any mistake in optimization.[48] The first order conditions for an optimum in the interior imply

$$f_j(\text{Input}_i^*, \theta_i) - nc_j(\text{Input}_i^*, \theta_i, \omega_i) = f_{j'}(\text{Input}_i^*, \theta_i) - nc_{j'}(\text{Input}_i^*, \theta_i, \omega_i)$$

where nc_j is defined analogously to f_j. In words, there should be a one-to-one relationship between differences in marginal productivity across two positive inputs j and j' and their corresponding net costs. If time input j is observed to have a greater marginal product than input j', the reason must be that input j is commensurately more costly (net of non-skill utility benefits). In addition, consider a situation where $\text{Input}_i^{*,j} = 0$ and $\text{Input}_i^{*,j'} > 0$. Then it must be the case that

$$f_j(\text{Input}_i^*, \theta_i) - nc_j(\text{Input}_i^*, \theta_i, \omega_i) \leq f_{j'}(\text{Input}_i^*, \theta_i) - nc_{j'}(\text{Input}_i^*, \theta_i, \omega_i).$$

That is, if the optimal choice for input j is zero, then the marginal net return of input j should be lower than the marginal net return of input j', for $\text{Input}_i^{*,j'} > 0$.

Given the discussion above, it is difficult to predict *ex ante* the expected distribution of $f_j(\text{Input}_i^*, \theta_i)$. The effects depend implicitly on the distribution across children of the marginal net costs of each activity, $nc_j(\text{Input}_i^*, \theta_i, \omega_i)$, which are in turn functions of the joint distribution of (θ_i, ω_i).[49]

This framework is useful to understand the role of heterogeneity in shaping our estimates

[48] For instance, if children and their parents want to maximize the true skill but perceive the production function to be $\tilde{f}(\text{Input}_i, \theta_i, \omega_i)$ instead of $f(\text{Input}_i, \theta_i)$, u can be written as $u := \tilde{f}(\text{Input}_i, \theta_i, \omega_i) - f(\text{Input}_i, \theta_i)$, where in this case ω_i is interpreted as the vector representing the heterogeneity of this misperception across children and their family. If instead they maximize just fun, then $u := u'(\text{Input}_i, \theta_i, \omega_i) - f(\text{Input}_i, \theta_i)$ where u' represents the actual component of the utility representing "fun".

[49] Moreover, non-linearities in the production function can complicate the interpretation even further. If $f(\text{Input}_i^*, \theta_i)$ is non-separable between Input_i^* and θ_i, or if $f(\cdot, \theta)$ is non-linear in inputs, as it appears to be according to our results in Section 4, then children with different values of (θ_i, ω_i) should choose different levels of $\text{Input}^*(\theta, \omega)$, leading them to have potentially different values of $f_j(\text{Input}^*(\theta, \omega), \theta)$. Remark 1 and Remark 2 in the appendix discuss this topic in more detail.

of the effect of time allocation on skills. As discussed in Remark 1, the estimates of our surviving models should represent an unbiased average of the distribution of $f_j(\text{Input}_i^*, \theta_i)$ across all children. The fact that we find that active time with grandparents has a positive return on cognitive skills suggests that on average, if all children increased the time they spend with grandparents by one hour we would observe an increase in cognitive skills. However, it may be that the cognitive skill of some children would decline with such a reallocation. Our specification of inputs is not detailed enough to capture such heterogeneous effects. To compensate for a lack of data, we ensure the test of exogeneity has power to detect endogeneity from heterogeneous effects that are not captured by our specification of inputs. Thus, we can reasonably conclude that the unobserved heterogeneity not incorporated in our specification of inputs does not generate endogeneity. However, we cannot conclude that this unobserved heterogeneity is small or unimportant for policy. Future investigation of heterogeneous effects of time allocation on skills along dimensions other than the ones we have studied is warranted.

5.3 Relationship with the Previous Literature

It is widely believed that child outcomes might improve if more of their time is spent in active activities.[50] However, evaluating this conventional wisdom is difficult because it is not clear which activities are actually productive and what these activities might substitute for. This paper adds to the literature by examining how child cognitive and noncognitive skills are impacted by time use, where time is categorized into comprehensive and precisely defined activities. We find that active time with parents or other activity partners helps children but only in developing math skills. Additional passive time does not hurt and sometimes helps with skill development. Further, schooling helps develop cognitive skills.

Although there is an extensive literature in economics on child skill development, there are only three studies, Del Boca et al. (2013), Del Boca et al. (2016) and Fiorini and Keane (2014), that estimate the effect of children's time allocation on skill formation. Del Boca et al. (2013) and Del Boca et al. (2016) also use the PSID-CDS, but do not incorporate all child activities, making it difficult to compare our results to theirs even if all three papers provided unbiased estimates. In contrast, Fiorini and Keane (2014) incorporate a comprehensive list of activities as we do, but use data from Australia rather than the US, and focus on earlier ages. Thus, it is difficult to make comparisons between our estimates and theirs even if both

[50] According to the American Academy of Pediatrics (AAP), children today spend seven hours a day on entertainment media (a passive activity). The AAP, however, recommends that children and teens should engage with entertainment media for no more than an hour or two a day. It is recommended that more time be allocated to outdoor play, reading, hobbies and free-play, all of which are active activities. See https://www.aap.org for additional details.

papers provided unbiased estimates. Indeed, one can think of institutional differences across countries that may lead to different estimates of $E[f_j(\text{Input}_i^*, \theta_i)]$ because w_i is distributed differently for children with the same value of θ_i (e.g., child care costs, female labor supply elasticity, social norm about how children should be raised, etc).[51]

Nevertheless, for completeness we compare our main findings with those from Fiorini and Keane (2014). While our findings regarding the production of noncognitive skills is similar to Fiorini and Keane (2014),[52] our results relating to the production of cognitive skills are quite different. In particular, we find that active time with parents or others in the US has little to no effect on cognitive skill formation, while Fiorini and Keane (2014) find that educational time with parents or others in Australia is quite productive. The source of this difference is difficult to pin down. In addition to the issues cited above, our aggregation scheme for time inputs and the set of controls included in our models are different from theirs.[53]

To truly understand the differences in findings, and ultimately the role of time allocation in skill development more broadly, much richer data and models of skill production and time allocation are needed. It is not enough to simply estimate more flexible production functions, since as noted above it is difficult to interpret the results without a formal model of time allocation.[54] Such a model would require specifying a utility function, determining the costs associated with each time input, and assessing the information available to children and their parents as they consider these input choices. While such a model is beyond the scope of this paper, we believe our approach and estimates of skill production are an important step towards the creation of this broader framework.

[51]The data confirms that the joint distribution of (θ_i, ω_i) in the Australian data is completely different from that in the American data. This can be inferred by the difference in the distribution of $\text{Input}_i^*(\theta, \omega)$ across these two countries as seen in the summary statistics in both papers. For instance, on average American children spend more passive time and less active time with their mother than Australian children do. As discussed in Section 5.2, differences in the joint distribution of (θ_i, ω_i) should lead to different estimates of $E[f_j(\text{Input}_i^*, \theta_i)]$ purely due to the presence of heterogeneous effects.

[52]We report the estimates regarding the noncognitive skill production function only in the appendix, since we cannot reasonably argue that they can be interpreted as causal. A common finding across the two studies is that noncognitive skills are relatively unresponsive to parental time inputs. Additionally, the fit of the noncognitive skill regressions in both papers tends to be poor, suggesting that much of the variation in child noncognitive skills remains unexplained. Both studies also find that sleeping is one of the more important activities for noncognitive skill production.

[53]We perform the same diagnostic procedure as in Section 3.4 on the time inputs Fiorini and Keane (2014) classify, using the same set of potential confounders. Results (Table D.3 in the appendix) show that compared to our time inputs, proportion of type (a1) confounders is on average much lower (i.e. 58% (71%) at 5% (10%) significance level) when using Fiorini and Keane (2014)'s time inputs. This suggests that our aggregation of time inputs leads to a more powerful test of exogeneity.

[54]The non-linearity we incorporate in our models is of a relatively modest form, a limitation imposed by the size of our sample.

6 Conclusion

Cognitive and noncognitive skills are critical for a host of economic and social outcomes as an adult. While there appears to be a consensus view that a significant amount of skill acquisition and development occurs early in life, the precise activities and investments that drive this process are not well understood. In this paper we examine how children's time allocation affects the accumulation of skill.

To do this, we apply a recently developed test of exogeneity to search for models that yield causal estimates of the impact time inputs have on child skills. The test exploits bunching in time inputs induced by a non-negativity time constraint. We provide evidence that the test is able to detect endogeneity arising from omitted variables, simultaneity, measurement error, and a host of misspecification errors. There are potential sources of endogeneity that the test is unable to detect. However, our robustness exercises, which are designed to detect them, suggest that our rich set of controls, together with a comprehensive list of time inputs, are able to absorb them. The test indicates that with a sufficient set of controls, already available in the most detailed datasets, we are unable to reject exogeneity of time inputs for cognitive skill formation.

We find that active time with adults, particularly grandparents, is valuable in developing cognitive skills. Relative to parents, grandparents tend to spend more time socializing, playing games, and pursuing artistic activities and are more engaged when they do so. The effects of time inputs are likely to be heterogeneous across families, children within families, and activities within our time input categories. As better data become available, a similar approach to the one implemented here can be used to uncover causal estimates at a more disaggregated level.

Finally, as time diaries become more ubiquitous, the methodology employed here provides researchers with a potential tool to study causality without an ex-ante source of exogenous variation. This is particular important when many IVs are required at the same time, as in studies involving resource allocations. For instance, a similar logic and testing strategy can be employed to address endogeneity concerns in other time use applications, such as the estimation of health production functions (e.g., exercising once a month should have a similar impact on physical well-being as exercising zero times a month in the absence of endogeneity) or of the impact of media consumption choices (DellaVigna and Ferrara (2015)).

References

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

Bernal, R. and Keane, M. P. (2010). Quasi-structural estimation of a model of childcare choices and child cognitive ability production. *Journal of Econometrics*, 156(1):pp.164–189.

Caetano, C. (2015). A test of exogeneity without instrumental variables in models with bunching. *Econometrica*, 83(4):pp.1581–1600.

Caetano, G. and Maheshri, V. (2016). Identifying dynamic spillovers of crime with a causal approach to model selection. *Quantitative Economics*. Forthcoming.

Cameron, S. V. and Heckman, J. J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of american males. *Journal of Political Economy*, 106(2):pp. 262–333.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *The Journal of Human Resources*, 43(4):pp. 738–782.

Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:pp. 697–812.

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

Currie, J. and Thomas, D. (1999). Early test scores, socioeconomic status and future outcomes. Technical report, National bureau of economic research.

Del Boca, D., Flinn, C., and Wiswall, M. (2013). Household choice and child development. *The Review of Economic Studies*, page rdt026.

Del Boca, D., Monfardini, C., and Nicoletti, C. (2016). Parental and child time investments and the cognitive development of adolescents. *Journal of Labor Economics*, 0(ja):null.

DellaVigna, S. and Ferrara, E. L. (2015). Economic and social impacts of the media. Technical report, National Bureau of Economic Research.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from head start. *American Economic Journal: Applied Economics*, pages 111–134.

Dustmann, C. and Schönberg, U. (2012). Expansions in maternity leave coverage and children's long-term outcomes. *American Economic Journal: Applied Economics*, 4(3):pp. 190–224.

Fiorini, M. and Keane, M. P. (2014). How the allocation of children's time affects cognitive and noncognitive development. *Journal of Labor Economics*, 32(4):pp. 787–836.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *The American Economic Review*, 103(6):pp. 2052–2086.

Keane, M. P. (2010). A structural perspective on the experimentalist school. *The Journal of Economic Perspectives*, 24(2):pp. 47–58.

Keane, M. P. and Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, 105(3):pp. 473–522.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, 142(2):pp. 698–714.

McLeod, J. D. and Kaiser, K. (2004). Childhood emotional and behavioral problems and educational attainment. *American Sociological Review*, 69(5):pp. 636–658.

Smith, P. K. (2003). *The psychology of grandparenthood: An international perspective*. Routledge.

Timmer, S. G., Eccles, J., and O'Brien, K. (1985). How children use time. *Time, goods, and well-being*, pages pp. 353–382.

Todd, P. E. and Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal*, 113(485):pp. F3–F33.

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

Woodcock, R. W. and Johnson, M. B. (1989). *Woodcock-Johnson tests of achievement*. DLM Teaching Resources.

Wooldridge, J. M. (2002). *Econometric Analysis Cross Section Panel*. MIT press.

Table 1: Summary of Ages

	Age Range	Average Age
CDS I: 1997	0-12 years old	6 years and 9 months
CDS II: 2002	5-17 years old	11 years and 9 months
CDS III: 2007	10-22 years old	16 years and 9 months
Our Sample	10-18 years old	14 years and 4 months

Table 2: Weekly Time in Each Activity (in Hours)

	Mean	SD	Proportion of Zero
Active time with mother	7.87	8.33	0.23
Passive time with mother	21.75	14.31	0.07
Active time with father	1.23	3.79	0.80
Passive time with father	2.61	5.71	0.61
Active time with grandparents	0.47	2.22	0.92
Passive time with grandparents	1.31	5.27	0.86
Active time with siblings	1.61	4.15	0.75
Passive time with siblings	3.50	6.68	0.54
Active time with friends	4.44	7.39	0.55
Passive time with friends	5.33	8.23	0.37
Active time with others	1.84	5.07	0.82
Passive time with others	2.29	7.38	0.60
Self active time	34.81	13.95	0.06
Self passive time	10.91	8.16	0.00
Sleeping or napping	64.93	10.02	0.00
Refused to answer or do not know	3.10	6.65	0.60

Note: The third column shows the proportion of children who spend zero minutes in a week on the corresponding time category.

Table 3: Demographics and Parental Background

	Mean	SD
Child's age (months)	172.41	24.83
Child's gender	0.49	0.50
Birth order to mother	1.97	1.08
Born in US	0.98	0.14
Mother's age	41.33	6.10
Father's age	44.21	6.73
Mother's age at child birth	28.00	5.71
Father's age at child birth	31.05	6.29
Mother has only high school degree	0.30	0.46
Mother has college degree	0.22	0.41
Father has only high school degree	0.23	0.42
Father has college degree	0.28	0.45
Number of siblings child lives with	2.24	2.68
Lives with two biological parents	0.55	0.50
Lives with grandparent	0.07	0.26
Household annual income (in $10,000s)	11.90	13.01

Table 4: Type of Controls: Our Time Inputs

Skills	Significance	Number of Variables	Number of Confounders	Type (a1)	Type (b)	Proportion of Confounders	Proportion of Type (a1)
Math	0.05	98	24	22	2	0.24	0.92
Vocabulary	0.05	98	23	22	1	0.23	0.96
Comprehension	0.05	98	23	22	1	0.23	0.96
Noncognitive	0.05	98	2	2	0	0.02	1.00
Math	0.10	98	26	24	2	0.27	0.92
Vocabulary	0.10	98	25	24	1	0.26	0.96
Comprehension	0.10	98	26	24	2	0.27	0.92
Noncognitive	0.10	98	3	3	0	0.03	1.00

Note: Column 3 shows the total number of variables in our initial pool of potential confounders, which includes lagged test scores, lagged time inputs, child characteristics, parental characteristics, family environmental characteristics, school environmental characteristics, school experience as well as variables related to misreporting of time diaries. Column 4 shows the number of confounders, which are identified if adding a variable significantly change the estimates of time inputs coefficients in a model with only time inputs as regressors (i.e. no controls). Column 5 shows number of type (a1) confounders, which are identified through a regression of a confounder on time inputs and their zero dummy variables: the confounder is of type (a1) if the coefficients of 15 time input dummies are jointly significantly different from zero. Column 6 shows number of type (b) confounders, which are confounders that do not belong to type (a1). Column 7 shows the ratio of number of confounders (i.e. column 4) over number of variables (i.e. column 3). Column 8 shows the ratio of number of type (a1) confounders (i.e. column 5) over number of confounders (i.e. column 4).

Table 5: Exogeneity Test Results

	Controls	Math		Vocabulary		Comprehension		Noncognitive	
		F-stat	p-Value	F-stat	p-Value	F-stat	p-Value	F-stat	p-Value
(1)	Lagged Score	5.237	0.000	3.559	0.000	3.332	0.000	**0.913**	**0.549**
(2)	Child Chrs.	1.801	0.030	**1.219**	**0.249**	1.548	0.081	**0.971**	**0.484**
(3)	Mother Demog. Chrs.	1.611	0.064	**1.039**	**0.411**	**1.268**	**0.214**	**0.999**	**0.453**
(4)	Family Demog. Chrs.	**1.328**	**0.177**	**0.899**	**0.564**	**1.111**	**0.340**	**0.994**	**0.459**
(5)	Family Environ. Chrs.	**1.334**	**0.173**	**0.881**	**0.585**	**1.013**	**0.438**	**1.005**	**0.447**
(6)	School Experience	**1.254**	**0.224**	**0.878**	**0.589**	**1.020**	**0.431**	**1.090**	**0.360**

Note: Entries in bold are "surviving specifications" for which we cannot reject exogeneity at 10% of significance. Each specification contains different control variables: (1) no controls, except for the lagged corresponding input; (2) child characteristics; (3) mother demographic characteristics; (4) family demographic characteristics; (5) Family environmental characteristics; (6) Child's school experience. See footnote 40 for a full description of the control variables. All standard errors are corrected for heteroskedasticity.

Table 6: Effects of Children's Time Allocation

	Math	Vocabulary	Comprehension
Active time with mother	0.005*	0.001	0.001
	(0.003)	(0.003)	(0.003)
Passive time with mother	0.004**	0.001	-0.001
	(0.002)	(0.002)	(0.003)
Active time with father	0.015**	0.007	0.000
	(0.005)	(0.005)	(0.007)
Passive time with father	-0.000	-0.001	0.008*
	(0.004)	(0.004)	(0.004)
Active time with grandparents	0.020**	0.020*	0.032**
	(0.010)	(0.012)	(0.010)
Passive time with grandparents	-0.004	-0.003	-0.006
	(0.004)	(0.005)	(0.006)
Active time with siblings	-0.003	-0.009	-0.015**
	(0.005)	(0.006)	(0.007)
Passive time with siblings	0.006*	0.003	0.003
	(0.003)	(0.003)	(0.004)
Active time with friends	0.005*	0.001	-0.001
	(0.003)	(0.003)	(0.003)
Passive time with friends	0.001	-0.005*	-0.001
	(0.003)	(0.003)	(0.003)
Self active time	0.005**	-0.001	0.000
	(0.002)	(0.002)	(0.002)
Self passive time	0.004*	-0.001	0.000
	(0.002)	(0.002)	(0.003)
Active time with others	0.001	-0.008*	-0.005
	(0.005)	(0.004)	(0.005)
Passive time with others	0.001	-0.005	-0.008**
	(0.003)	(0.004)	(0.004)
Don't know or refuse to answer	0.004	-0.001	0.004
	(0.003)	(0.003)	(0.003)
R-Squared	0.660	0.635	0.565
Observations	1698	1698	1698
Exogeneity test F-statistic	1.254	0.878	1.020
Exogeneity test p-value	0.224	0.589	0.431

Note: All estimates are for specification (6). See footnote 40 for a full description of the control variables. Standard errors corrected for heteroskedasticity are in parentheses. * Significant at the 10% level. ** Significant at the 5% level.

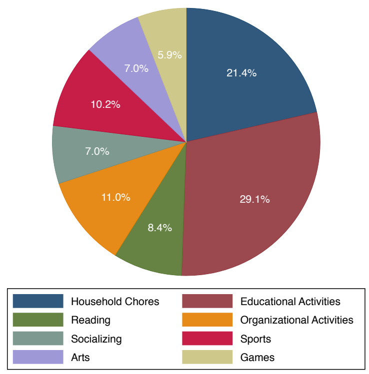
Table 7: P-Values for Comparing Surviving and Non-surviving Specifications

	Controls	Math	Vocabulary	Comprehension	Noncognitive
(1)	Lagged Score	0.000	0.000	0.000	**0.805**
(2)	Child Chrs.	0.004	**0.326**	0.001	**0.634**
(3)	Mother Demog. Chrs.	0.040	**0.693**	**0.029**	**0.634**
(4)	Family Demog. Chrs.	**0.109**	**0.745**	**0.140**	**0.612**
(5)	Family Environ. Chrs.	**0.383**	**0.912**	**0.512**	**0.620**

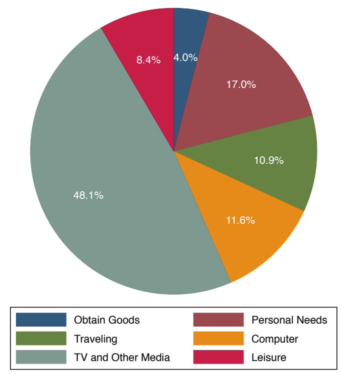
Note: This table shows the p-values of a joint test for whether the 15 coefficients of Input_i for each specification are the same as the corresponding ones from specification (6) in Table 5. Entries in bold are "surviving specifications" with respect to the exogeneity test, i.e., those for which we cannot reject exogeneity at 10% of significance. Each specification contains different control variables: (1) no controls, except for the lagged corresponding input; (2) child characteristics; (3) mother demographic characteristics; (4) family demographic characteristics; (5) family environmental characteristics. See footnote 40 for a full description of the control variables. All standard errors are corrected for heteroskedasticity.

Figure 1: Activity Composition

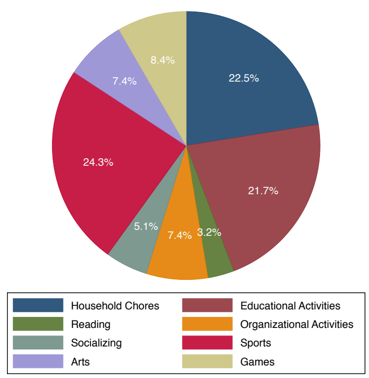
(a) Active Time with Mother



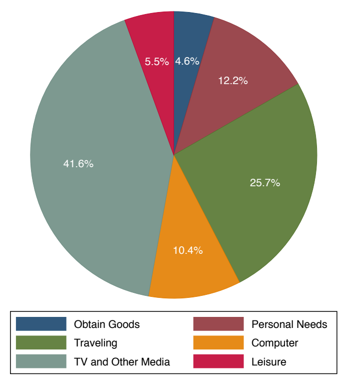
(b) Passive Time with Mother



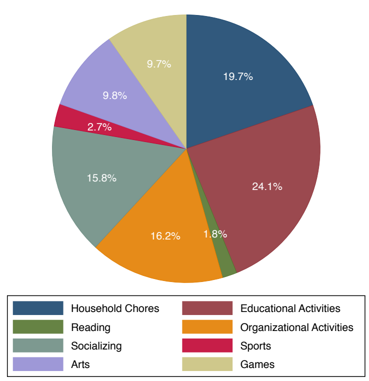
(c) Active Time with Father



(d) Passive Time with Father



(e) Active Time with Grandparents



(f) Passive Time with Grandparents

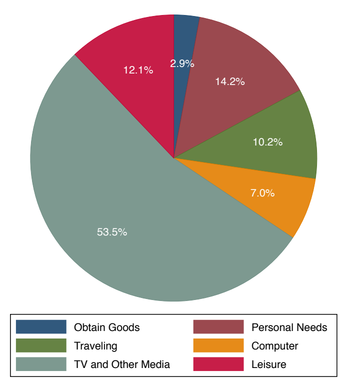
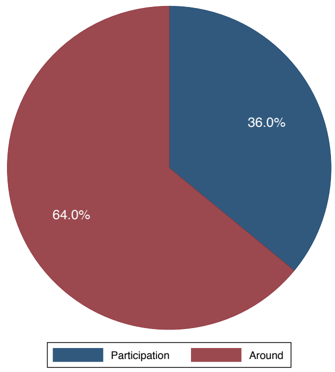
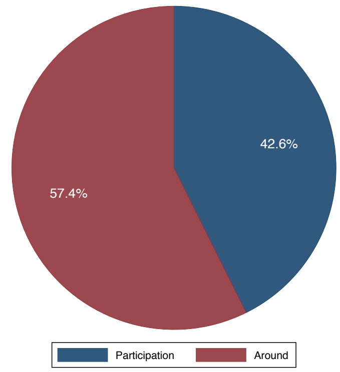


Figure 2: Participation Time

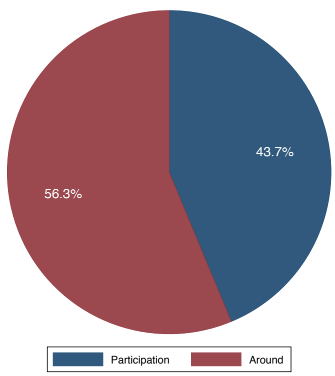
(a) Active Time with Mother



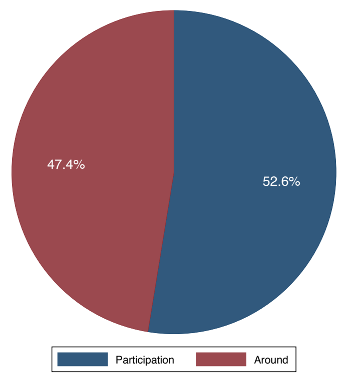
(b) Passive Time with Mother



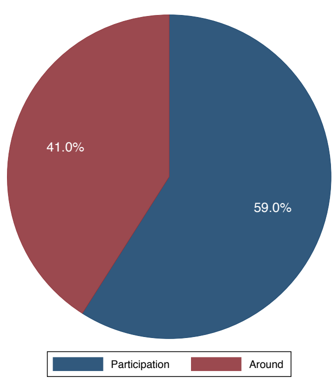
(c) Active Time with Father



(d) Passive Time with Father



(e) Active Time with Grandparents



(f) Passive Time with Grandparents

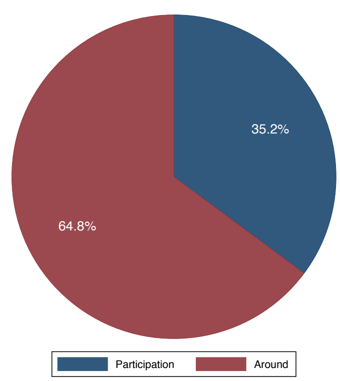
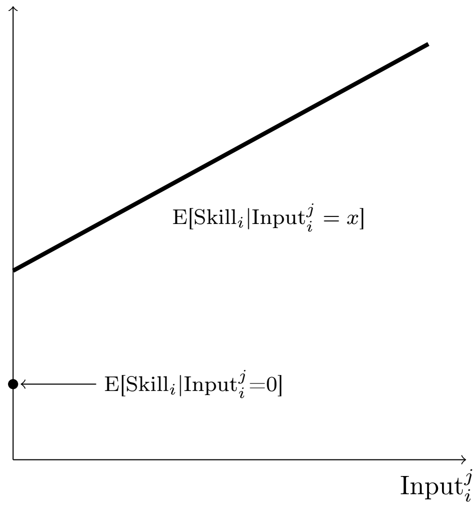
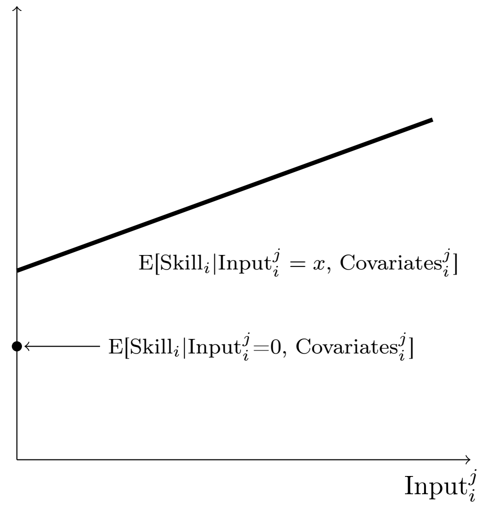


Figure 3: Intuition for the Test of Exogeneity



(a) Correlation between Time Input and Child Skill, Unconditional



(b) Correlation between Time Input and Child Skill, Conditional on Covariates

Figure 4: Why are Unobservables Discontinuous at Input$_i^j = 0$?

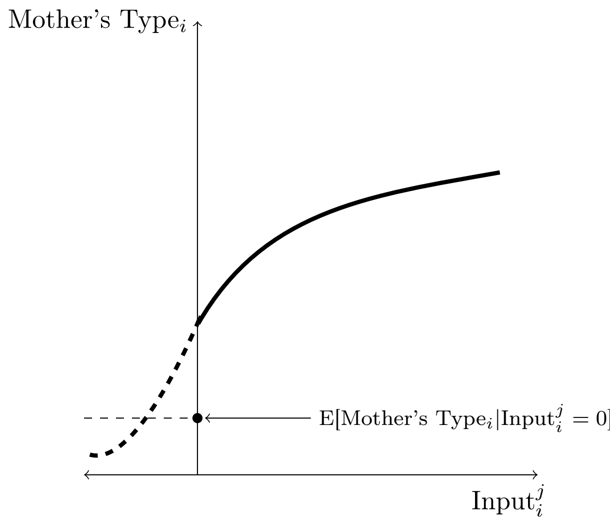
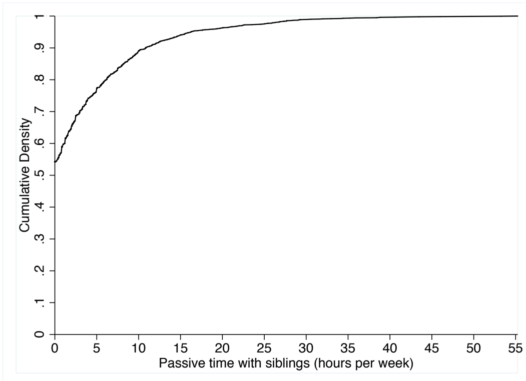
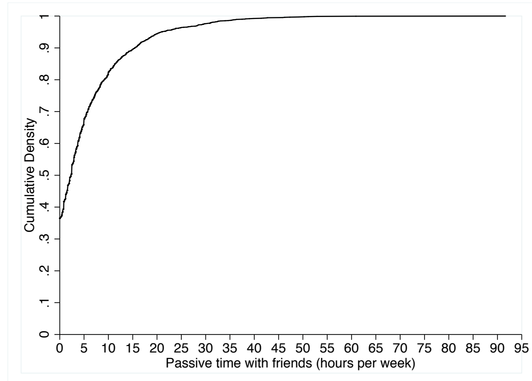


Figure 5: Evidence of Bunching

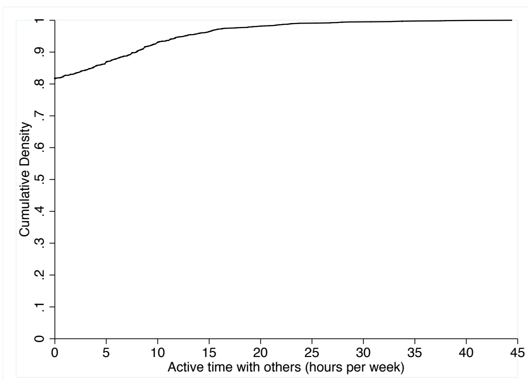
(a) Passive Time with Siblings



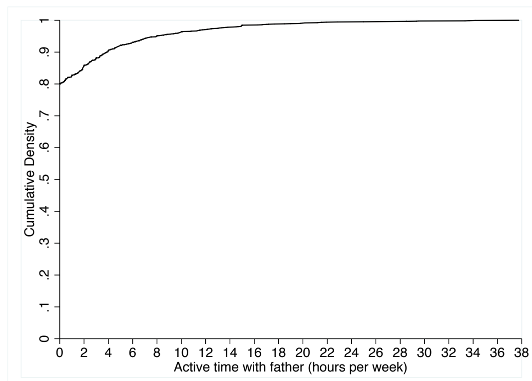
(b) Passive Time with Friends



(c) Active Time with Others

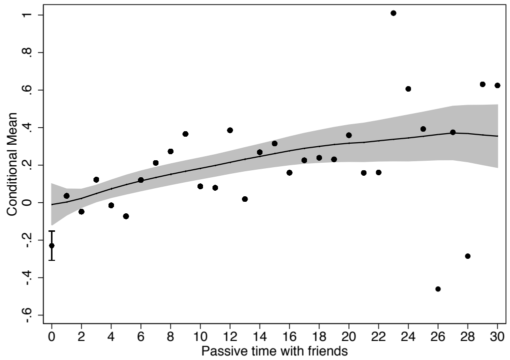


(d) Active Time with Father

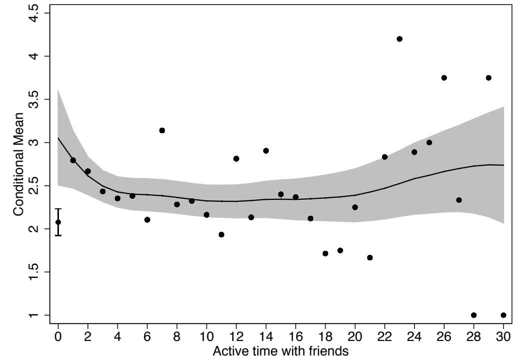


Note: Each plot shows the cumulative density function of the time spent in the corresponding activity for the corresponding cohort. The fact that these plots cross the vertical axis not at the origin is direct evidence of bunching, as it implies the probability density function is discontinuously larger at zero. Time described in the horizontal axis is reported in hours per week, but continuously (in minutes per week).

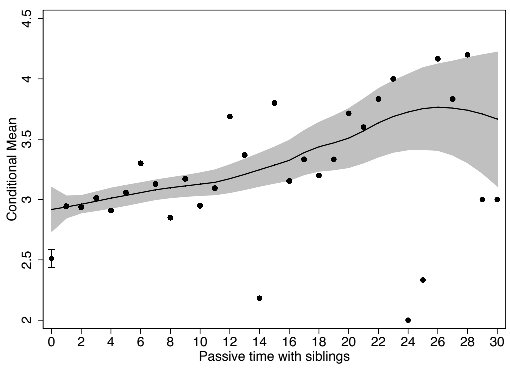
Figure 6: Evidence of Power to Detect Endogeneity from Omitted Variables



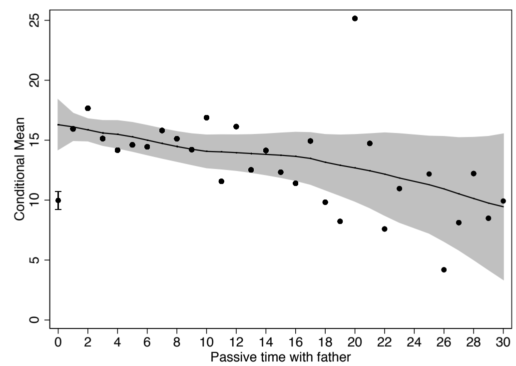
(a) Lagged Math Score



(b) Number of Siblings



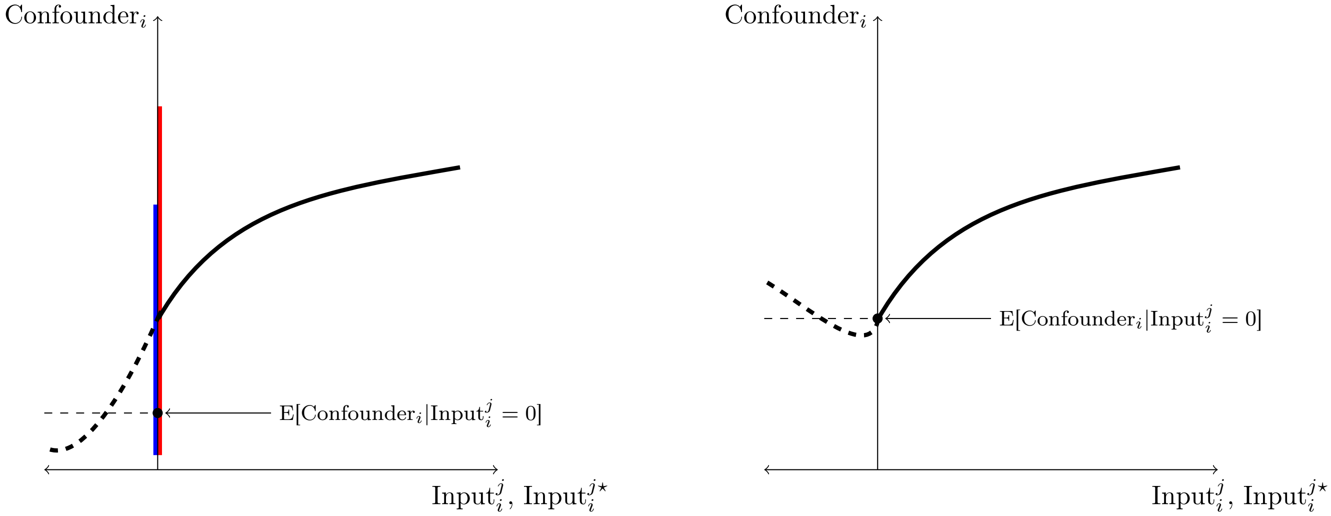
(c) Number of Children Born to Mother



(d) Household Income ($10,000s Per Year)

Note: In each plot, the vertical axis shows the mean of a potential confounder conditional on a given level of time input (i.e. horizontal axis variable). The scatter plot represents the observed conditional mean of the confounder (aggregated to the next hour of the time input). At zero time input, we show the 95% confidence interval. The solid curve represents a third order local polynomial regression of the confounder on the time input, using time input data at the minute per week level. The shaded region represents the 95% confidence interval for this regression with an out-of-sample prediction at zero minutes. See footnote 33 for more details on the regression and confidence interval.

Figure 7: Types of Confounders



(a) Type (a1) vs. Type (a2)

(b) Type (b)

Note: Input$_i^{j\star}$ represents the optimal choice of input j by individual i. Red range: Support of confounder among all observations of sample. Blue range: Support of confounder among all observations of sample for which Input$_i^j = 0$. The confounder is of type (a1) if some of its correlation with Skill$_i$ happens for values of the confounder in the blue range, otherwise it is of type (a2).