Evaluating the Effect of Maternal Time on Child Development Using the Generalized Propensity Score

Pedro Carneiro
UCL, IFS and CEMMAP

Margarida Rodrigues*
Universidade Nova de Lisboa

January, 2009
(preliminary version)

Abstract

Mainly due to data unavailability, time with the mother is usually not included as an input in empirical papers trying to find the determinants of child's achievement. We attempt to do so in this paper by using unique data that collects a child's time-use diary, cognitive and non cognitive test scores. We implement the methodology developed recently on the treatment evaluation for continuous treatments. We find that more time with mothers leads both young and old children to perform better in cognitive tests, but the effect for the former group is more pronounced. Once we divide our sample according to race we find that young black children tend to perform worse if they spend more than 5 hours with the mother in a day. This negative effect is not present for white children, who tend to perform better if the time spent with mothers increases. Also white children benefit from this effect until later ages.

Keywords: Maternal time, Cognitive and non-cognitive outcomes, Treatment evaluation.
JEL codes: C21, I29.

*Presenter; m.margarida.r@fe.unl.pt; Campus de Campolide, Colégio de Campolide, 1099-032 Lisboa

1 Introduction

Several papers in the economic literature attempt to determine the determinants of child's scores from cognitive and non-cognitive tests. There are not many papers that include the time mother and child spend together as an explanatory variable mainly due to data unavailability. The best some papers do is to relate mother employment (in intensive and extensive margin) to child outcomes, i.e. they assume that the mother's non-working time is entirely spent with the child (e.g. Bernal and Keane (2005), James-Burdumy (2005), Blau and Grossberg (1992), Ruhm (2004)). The above methodology, despite being very interesting, allows uniquely to assess the impact of working (or working hours) on test scores while it does not allow to make any considerations about the impact of the actual time they spend together. We can think of two cases in which the former analysis would lead to wrong conclusions: i) mothers that don't work or do work very few hours but spend very few of their free time with children; and ii) mothers that do work or work for longer hours but that spend almost their entire free time with their children. So, if one attempts to relate mother-child time with child outcomes it is important to have a direct measure of the actual time they spend together. The type of data needed to do such an analysis is very hard to find because we must have simultaneously child outcomes and an accurate measure of time. The usual time diary surveys do not have the former while other databases do not have the later. In this paper we use a unique data set from the 1997 PSID - Child Development Supplement made with the purpose of providing researchers with a comprehensive database of children and families that would enable to study the dynamic process of early human capital formation. Up to 2 children per PSID family aged 0-12 years during the calendar year of 1997 were selected for interview. The interview included a time use diary with 24-hour detailed accounting of time use for one randomly selected weekday and weekend, with the type, duration and location of activities and also the social context of activities, i.e. detailed information about whom participated in the activity and who else was there but not directly engaged. It was also collected aptitude and achievement test scores (reading and math and memory) and psychological, emotional and social well being (behavior problem index, positive behavior) as well as a bunch of family context variables. So this seems the perfect data set to assess the impact of mother-child time on child's test scores, as we have the two necessary variables and a bunch of variables to control for. We focus on three outcomes: letter word identification, applied problems and behavior problem index. Our sample has 1497 children, which we then divide into two six sub-samples according to their age and race.

In order to assess the causal effect of the time spent with mothers in the cognitive and non cognitive achievement we use the recent development in program evaluation made by Hirano and Imbens (2004). The latter extends the usual binary treatment case to a continuous treatment, which suits particularly well our needs since the treatment variable we consider, time, is continuous. We are able then to use the generalized propensity scores and estimate dose response functions, i.e. the response function of each outcome to each level of the treatment variable.

The remainder of the paper is organized as follows. Next section describes the use of the generalized propensity score and methodology of estimating a dose response function to evaluate a continuous treatment. Section 3 presents the data used in this paper. The fourth section presents our empirical results and section 5 concludes.

2 The Generalized Propensity Score

In recent years the research in program evaluation has made comprehensive use of matching methods. The standard case considers a binary treatment. Rosenbaum and Rubin (1983) provided the key result that has made the matching such an attractive method: rather than conditioning on the full set of covariates, conditioning on the propensity score, i.e. on the probability of receiving the treatment given the covariates, is sufficient to balance treatment and comparison groups.

More recently, the literature has extended propensity score methods to the cases of multi-valued treatments (Imbens(2000) and Lechner (1999)) and continuous treatments (Hirano and Imbens (2004). The approach of the latter paper is particularly suitable for our purpose because it enables to estimate the entire dose response function of our continuous treatment- time. Therefore we follow closely Hirano and Imbens's (2004) methodology and implementation, that are summarized below.

2.1 Methodology

We have a random sample of units $i = 1, ..., N$ and for each we observe a set of potential outputs $Y_i(t)$ from a treatment t. In the usual binary case the treatment set is $\check{T} = \{0, 1\}$ whereas in the continuos case $\check{T}$ is an interval $[t_0, t_1]$. For each sample unit i we observe a vector of covariates X_i, the level of treatment actually received $T_i \in [t_0, t_1]$ and the outcome $Y_i(T_i)$. Our objective is to estimate the average dose response function $\mu(t) = E[Y(t)]$. In the remainder of the section we ignore subscript i.

The key assumption of Hirano and Imbens (2004) generalizes the unconfoundedness assumption for binary treatments to the continuos case. The weak unconfoundedness assumption is the following:

$$Y(t) \perp T \mid X \ , \forall t \in \check{T} \tag{1}$$

and it is named so because it only requires independence to hold for each level of treatment t rather than the joint independence of all potential outcomes.

Call $r(t, x) = f_{T|X}(t|x)$, i.e. $r(t, x)$ is the conditional density of the treatment given the covariates. The generalized propensity score (GPS) is defined as:

$$R = r(T, X). \tag{2}$$

The GPS has a balancing property similar to that of the standard propensity score, as within the strata with the same value of $r(t, x)$ the probability of $T = t$ does not depend on the value of X. In other words, GPS has the following property:

$$X \ \perp \ 1\{T = t\} \mid r(t, x). \tag{3}$$

Hirano and Imbens (2004) highlight that this property does not require unconfoundedness. However, when combined with unconfoundedness, it implies that assignment to treatment is unconfounded given the GPS:

$$Y(t) \perp T \mid X \ , \forall t \in \check{T} \quad \Rightarrow \quad Y(t) \perp T \mid r(T, X) \ , \forall t \in \check{T}. \tag{4}$$

Given this result it is possible to use the GPS to remove the bias associated with differences in covariates in two steps:

1. Estimate the conditional expectation of the outcome as a function of two variables- treatment T and GPS R:
$$\beta(t, r) = E[Y \mid T = t, R = r] \tag{5}$$

2. Estimate the dose response function (DRF) at each particular level of the treatment. This is implemented by averaging the conditional expectation over the GPS at that particular level of treatment:
$$\mu(t) = E[\beta(t, r(t, X))] \tag{6}$$
Notice that we do not average over the GPS $R = r(T, X)$ but instead we average over the score evaluated at the treatment level of interest $r(t, X)$.

It should be stressed that the regression function $\beta(t,r)$ does not a causal interpretation but $\mu(t)$ corresponds to the value of the DRF for treatment level t, which compared to another treatment level t' does have a causal interpretation.

2.2 Implementation

In the practical implementation of the methodology outlined in the previous section we use a normal distribution for the treatment given the covariates:

$$T_i \mid X_i \tilde{} N(\beta_0 + \beta_1' X_i, \sigma^2). \tag{7}$$

1. In the first stage we estimate the parameters β_0, β_1 and σ^2 by OLS. Then the estimated GPS is:
$$\widehat{R}_i = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(T_i - \widehat{\beta}_0 - \widehat{\beta}_1' X_i)^2}{2\widehat{\sigma}^2}\right] \tag{8}$$

2. In the second stage we model the conditional expectation of Y_i given T_i and R_i as a flexible function of its two arguments. In the application used by Hirano and Imbens (2004) a quadratic approximation is used:
$$E[Y_i|T_i, R_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 R_i + \alpha_4 R_i^2 + \alpha_5 T_i R_i \tag{9}$$
The parameters are estimated by OLS using the estimated GPS $\widehat{R}_i$.

3. In the third stage, and given the estimated parameters in the second stage, the average potential outcome at treatment level t is estimated:
$$\widehat{E[Y(t)]} = \frac{1}{N}\sum_{i=1}^{N}\left[\widehat{\alpha}_0 + \widehat{\alpha}_1.t + \widehat{\alpha}_2.t^2 + \widehat{\alpha}_3\widehat{r}(t,X_i) + \widehat{\alpha}_4\widehat{r}(t,X_i)^2 + \widehat{\alpha}_5.t.\widehat{r}(t,X_i)\right] \tag{10}$$
We should do this for every treatment level we are interested in order to obtain the entire dose-response function. It is convenient to use bootstrap methods to form standard errors and confidence intervals.

3 Data

In this paper we use the Child Development Supplement (CDS) of the Panel of Income Dynamics (PSID). The PSID is a nationally representative sample of U.S. families that collects data on income, work and consumption. PSID families have been interviewed since 1968 and sample members are followed as they split off into new households. In 1997, with the purpose of providing researchers with a comprehensive database of children and families that would enable to study the dynamic process of early human capital formation, it was developed the PSID Child Development Supplement. Up to 2 children per PSID family aged 0-12 years during the calendar year of 1997 were selected. Then these children were followed and re-interviewed in 2002/03 with ages 5-18 years. The interview included a time use diary with 24-hour detailed accounting of time use for one randomly selected weekday and weekend, with the type, duration and location of activities and social context of activities (detailed information about whom participated in the activity and who else was there but not directly engaged). It was also collected aptitude and achievement test scores (Reading and math and memory) and psychological, emotional and social well being (behavior problem index, positive behavior) as well as a bunch of family context variables. So this seems the perfect data set to assess the impact of mother-child time on child's test scores.

For each individual in our sample we compute the total time they spend with mothers in a week day [1]. This will be our treatment variable. This variable englobes the time that the mother is participating in the activity with the child or not, i.e. the mother might just be around while the child does an activity by herself. We are particular interested in three outcomes, Letter-Word Identification (LW), Applied Problems (AP) and Behavior Problem Index (BPI), such that we have information both on cognitive and non-cognitive skills. Table 1 in the appendix describes all the variables we use in our analysis.

In 1997, 3563 children were interviewed. We will only consider the 2478 children that completed the time diary questionnaire (both in week and weekend days), that live with the mother and for whom their mother was reported to be the primary caregiver. Our final sample consists of 1497 children that have available at least one of the three outcome variables and for whom none of the several covariates we consider (child's, mother's and family's characteristics) is missing. In the forthcoming analysis we always divide children in two age groups: the first group is composed by younger children (from 3 to 6 years old) and is named *Young* and the second group is composed by older children (from 7 to 12 years old) and is named *Old*. We believe it is important to make this distinction because the time needs for one and other group seem to be very different and can therefore affect determinantely the impact of the treatment. Furthermore we distinguish between white and black children because the majority of their covariates is significantly different from each other. We see this by computing the difference in means and checking whether these are significantly different from zero. We conclude that, both for young and old children, the means of the following covariates are significantly different between races: birth weight, age breast feeding stopped, mother's age at birth, marital status at birth, mother's education, dummy for both parents living with the child, dummy of other adults living in house, dummy of living at own house, father's education, 5 year average of total income, mother's fixed effect and state's average weekly wage.

So, summarizing we have 6 different samples: young children with 561 observations, old children with 936 observations, young white children with 341 observations, young black children with 220 observations, old white children with 540 observations and old black children with 396 observations. We implement the methodology presented in the previous section separately for each of them. Table 2 presents the summary statistics for every variable we use and for each of the six samples.

On average mothers with young children spend more hours in a week day with children than mothers with old children. In this table we present the standard outcomes but in all our analysis we normalize them to have mean 0 and standard deviation 1 within each sample. Young children preform worse than old children in LW, AP and better in BPI. Analyzing the columns corresponding to the division of children according to age and race (columns (3) to (12)) we conclude that white children preform much better in all outcomes considered. Also, white mothers spend more than 1 hour per week day with their children irrespectively of being young or old. This could lead us to think that indeed time with mothers determines the outcomes as we observe an association between higher treatments and higher outcomes. However, notice that there exists selection bias, i.e. we observe that the characteristics of black and white samples are very different and significant as explained above. Furthermore the black samples seem to be disadvantaged in comparison with their white counterparts in almost every characteristics we can observe: white children weighted more at birth, were breast fed longer, had an older mother at birth, have more educated mothers and fathers, are more likely to be part of a biparental family, have less siblings, are more likely to own their house, belong to richer families, and have abler mothers. So indeed white children seem to have a privileged background.

[1] For now we focus on a week day as these are the days in which a time constraint is more relevant. Later on we intend to use also the weekend time diary and see whether there is any time compensation during the weekend.

Table 2- Summary Statistics

	Young		Old		Young Black		Young White		Old Black		Old White	
Treatment	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Time	5.99	3.23	4.29	2.50	5.28	2.91	6.46	3.45	4.11	2.46	4.42	2.52
Outcomes (standard)												
LW	99.2	15.6	106.7	19.1	95.4	15.7	101.6	15.1	99.3	16.5	112.5	18.9
AP	101.9	18.2	109.2	16.8	94.1	18.6	107.1	15.9	101.1	14.1	115.5	16.0
BPI	7.49	4.83	7.96	5.95	7.20	4.68	7.67	4.93	8.04	6.36	7.91	5.63
Covariates												
Ageatpcg	60.6	13.3	121.8	22.1	61.2	12.9	60.2	13.5	123.1	22.5	120.8	21.7
RaceWhite	0.61	0.49	0.58	0.49	0	0	1	0	0	0	1	0
RaceBlack	0.39	0.49	0.42	0.49	1	0	0	0	1	0	0	0
Chgender	0.56	0.49	0.49	0.50	0.56	0.49	0.55	0.49	0.53	0.49	0.47	0.49
Birthorder	1.89	1.12	2.08	1.12	1.98	1.31	1.83	0.97	2.29	1.27	1.93	0.96
Birthweight	6.88	1.39	6.94	1.38	6.51	1.47	7.13	1.28	6.59	1.44	7.19	1.27
Ageatbreast	3.23	5.97	3.10	5.31	1.30	3.79	4.47	6.74	1.11	3.25	4.56	6.00
AgeBirthM	27.2	5.95	27.3	5.18	25.1	5.96	28.5	5.56	26.3	5.32	28.1	4.94
MaritalBirth	0.68	0.46	0.73	0.44	0.39	0.49	0.87	0.33	0.45	0.49	0.93	0.25
EducationM	13.2	2.02	13.1	2.14	12.5	1.77	13.6	2.04	12.4	1.87	13.5	2.19
ParentLive	0.66	0.47	0.66	0.47	0.38	0.48	0.85	0.36	0.42	0.49	0.84	0.37
ChildrenFU	2.22	1.03	2.42	1.02	2.26	1.21	2.19	0.90	2.52	1.21	2.35	0.83
AgeYoungest	3.41	1.58	7.27	3.14	3.49	1.52	3.354	1.62	7.19	3.41	7.32	2.93
OtherAdults	0.77	0.49	0.86	0.59	0.53	0.58	0.93	0.34	0.71	0.73	0.97	0.44
OwnHouse	0.56	0.49	0.66	0.47	0.31	0.46	0.72	0.45	0.47	0.49	0.80	0.4n
EducationF	9.06	6.68	8.86	6.67	4.93	6.45	11.7	5.34	5.21	6.30	11.5	5.57
Avg(Total Income)	10.6	0.97	10.8	0.91	10.1	0.96	11.0	0.76	10.3	0.93	11.2	0.69
Mother's FE	-0.77	2.10	0.47	1.99	-1.36	2.02	-0.39	2.06	0.15	1.96	0.71	1.98
AvgWeeklyWage	561	76.9	567	86.3	543	64.7	573	81.8	554	86.9	576	84.5
Observations	516		936		220		341		396		540	

4 Results

4.1 GPS, Covariate Balance and Common Support

We first estimate the conditional distribution of the treatment by estimating equation (7) by OLS. Table 3, in the appendix, presents the results of this estimation for all the samples. We then compute the GPS according to equation (8).

As in the binary treatment case it is important to impose common support. This guarantees that for the same propensity score there are observations from the different treatment groups and therefore they can compared. Given the GPS estimate we divide the sample in 3 groups according to the 33th (p33) and 66th (p66) percentile of the treatment variable T_i. So:

$$\begin{aligned} i &\in G1 \text{ if } T_i \in [0, p33] \\ i &\in G2 \text{ if } T_i \in (p33, p66] \\ i &\in G3 \text{ if } T_i \in (p66, p33] \end{aligned}$$

Then we evaluate the GPS for each observation in each group, which yields the variables:

$$\begin{aligned} \text{if } i &\in G1 \Rightarrow R_i^1 = \widehat{R}_i; \text{ if } i \notin G1 \Rightarrow R_i^1 = NaN \\ \text{if } i &\in G2 \Rightarrow R_i^2 = \widehat{R}_i; \text{ if } i \notin G2 \Rightarrow R_i^2 = NaN \\ \text{if } i &\in G3 \Rightarrow R_i^3 = \widehat{R}_i; \text{ if } i \notin G3 \Rightarrow R_i^3 = NaN. \end{aligned}$$

We assume that the control group for each group is formed by the other two groups combined. For instance the control group of G1 is formed by individuals in the other two groups combined (G2 and G3). Then we eliminate observations in each group that have a GPS higher than the maximum or lower than the minimum GPS observed in the control group, i.e.:

$$\begin{aligned} \text{eliminate } i &\in G1 \text{ iff } \left[R_i^1 < \min\left\{\min(R^2), \min(R^3)\right\} \text{ or } R_i^1 > \max\left\{\max(R^2), \max(R^3)\right\}\right] \\ \text{eliminate } j &\in G2 \text{ iff } \left[R_j^2 < \min\left\{\min(R^1), \min(R^3)\right\} \text{ or } R_j^2 > \max\left\{\max(R^1), \max(R^3)\right\}\right] \\ \text{eliminate } k &\in G3 \text{ iff } \left[R_k^3 < \min\left\{\min(R^1), \min(R^2)\right\} \text{ or } R_k^3 > \max\left\{\max(R^1), \max(R^2)\right\}\right]. \end{aligned}$$

Part A of table 4 present the percentage of observations eliminated for each sample as well as the maximum percentage of observations eliminated in our 1000 bootstrap samples.

It is also important to evaluate how well the adjustment for the GPS works in balancing the covariates, i.e. if the specification of equation (7) is adequate. Whereas in the binary case the typical approach is to compare the covariate means for the treated and control groups, before and after matching, testing for covariate balance is more difficult in the continuos case.

Here we implement exactly what is proposed by Hirano and Imbens (2004). We use the 3 groups defined in point 2 above: G1, G2 and G3.

1. **Unadjusted**: For each of the covariates we investigate the balance by testing whether the mean in one of the three treatment groups is different from the mean in the other two treatment groups combined. Take for instance variable X and suppose we want to test the balance of this covariate in G1 vs the other two groups. We compute the mean and variance of X for observations belonging to G1, $\overline{X}_1$ and $Var(X_1)$. Next we compute the mean and variance of X for the observations belonging to G2 *and* G3, $.\overline{X_{2\&3}}$ and $Var(X_{2\&3})$. Then we compute the T-test statistic corresponding to the null that these means are indeed the same and compare them to the 10% and 5% critical values. We reject the null, i.e. the covariates are unbalanced, if the t-statistic is higher than the critical value.

2. **Adjusted for the GPS**: As an example consider we want to test the balance of the covariate age in the sample of young children[2]. To implement the GPS-adjusted version of this statistic we discretize both the level of treatment and the GPS. The steps are the following:

   - Given the groups formed in point 2, we find the median treatment in each group and evaluate the GPS at this median treatment. For instance the median treatment in group 1 is 3.17 and we get $r(3.17, X_i)$. Basically what we will test is whether:

$$X_i \perp 1\{0 \leq T_i < 4.17\} \mid r(3.17, X_i).$$

   - Next we block on the score $r(3.17, X_i)$. We make five blocks defined by the quintiles of $r(3.17, X_i)$ in the group with $1\{0 \leq T_i < 4.17\}$: $[0, 0.0066]$, $(0.0066, 0.0380]$, $(0.0380, 0.0819]$, $(0.0819, 0.1912]$ and $(0.1912, 0.3242]$.
   - In each of these five blocks we compute the difference in means of age with respect to individuals that have a GPS such that they belong to that block, but have a treatment level different from the one being evaluated. I.e., if we are testing for group 1, we will compute the following difference in means:

$$\begin{aligned} mean(age_i \mid i \in & \text{ G1 \& } i \in \text{block 1 of G1}) - \\ -mean(age_j \mid j \notin & \text{ G1 \& } j \in \text{block 1 of G1}) \end{aligned}$$

   - For example:
     - the first of these five blocks, with $r(3.17, X_i) \in [0, 0.0066]$, has a total of 174 observations: 42 with $T \in [0, 4.17]$ and 132 with $T \notin [0, 4.17]$. Testing for equality of the mean of age in the first versus the other treatment groups in this GPS group give a mean difference of -0.057 and a standard deviation of 1.72.
     - the second of these five blocks, with $r(3.17, X_i) \in (0.0066, 0.0380]$, has a total of 147 observations: 42 with $T \in [0, 4.17]$ and 105 with $T \notin [0, 4.17]$. Testing for equality of the mean of age in the first versus the other treatment groups in this GPS group give a mean difference of 2.86 and a standard deviation of 2.32.
     - the third of these five blocks, with $r(3.17, X_i) \in (0.0380, 0.0819]$, has a total of 84 observations: 41 with $T \in [0, 4.17]$ and 43 with $T \notin [0, 4.17]$. Testing for equality of the mean of age in the first versus the other treatment groups in this GPS group give a mean difference of 0.95 and a standard deviation of 2.48.
     - the fourth of these five blocks, with $r(3.17, X_i) \in (0., 0819.1912]$, has a total of 89 observations: 42 with $T \in [0, 4.17]$ and 47 with $T \notin [0, 4.17]$. Testing for equality of the mean of age in the first versus the other treatment groups in this GPS group give a mean difference of -2.69 and a standard deviation of 2.13.
     - the fifth of these five blocks, with $r(3.3, X_i) \in (0.1912, 0.3242]$, has a total of 58 observations: 42 with $T \in [0, 4.17]$ and 16 with $T \notin [0, 4.17]$. Testing for equality of the mean of age in the first versus the other treatment groups in this GPS group give a mean difference of -2.47 and a standard deviation of 2.15.
   - Combining these five difference in means, weighted by the total number of observations in each block leads to a mean difference of 0.70 and a standard error of 3.91 and thus to a t-statistic of 0.18. So in this example we do not reject the null of equality of means between individuals with the same GPS but in different treatment groups. This means that, when adjusted for the GPS, the covariate age is balanced between groups.

[2]For this sample the 33^{th} and 66^{th} percentiles of treatment variable are 4.17 and 6.75 hours per week day, respectively.

- This procedure needs to be repeated for each treatment group and for each covariate, therefore we will have 57 t-statistics for the young and old samples and 51 t-statistics for the samples in which we divide both for age and race.

Part B of table 4 shows the number of t-statistics explained before that are higher than 1.645 or 1.96, i.e. that are significantly different from zero at the significant levels of 95% and 90%.

Table 4- Results on common support and covariate balance

	Part A		Part B			
			t.st.$\geq$ 1.645		t.st.$\geq$ 1.96	
	(1)	(2)	Unadj	Adj.	Unadj.	Adj.
Young	0.016	0.060	20	0	19	0
Old	0.038	0.081	11	0	2	0
Young Black	0.045	0.100	18	0	14	0
Young White	0.009	0.102	9	0	6	0
Old Black	0.035	0.119	8	0	3	0
Old White	0.046	0.144	8	0	6	0

If adjustment for the GPS properly balances the covariates, we expect that the differences in means to not be statistically different from zero. The values in the table show that the adjustment for the GPS improves the balance as, before the adjustment, and depending on the sample considered, there is an important number of covariates that present t-statictis higher than 1.645 which indicates a clear unbalanced covariates distribution. After the adjustment, the number of t-statistics higher than 1.645 or 1.96 is zero for all samples meaning that adjusting for the GPS is successful in balancing the covariates.

4.2 Estimating and Plotting the Dose Response Functions

The final step is to estimate the GPS-adjusted dose response function. Table 5 contains the estimation results for the dose response function, i.e. it presents the estimated coefficients of equation (9). Again notice that these coefficients do not have any interpretation but we use these to estimate equation (10). For young children we evaluate this equation for levels of treatment between 0 and 12 hours per week day and for old children we evaluate that equation for levels of treatment between 0 and 8 hours per week day.

4.2.1 Age Divided Samples

Figures 1 and 2 present the dose response functions of LW, AP and BPI for young and old children, respectively. Notice that these figures show the results of equation (10), i.e. the expected average outcome for each level of treatment we have considered and the 95% confidence intervals. For young children, the LW and AP d.r.f. are clearly increasing in the treatment variable, meaning that spending more hours per week day with the mothers lead young children to perform better in these tests. For old children, it seems that the LW and AP d.r.f. slightly increase in the treatment variable but much less pronoucedly than those of young children. Furthermore, for higher levels of treatment, the d.r.f. seems to decrease. The BPI d.r.f. is more flat both for young and old children, suggesting that spending more time with the mother does not have an impact on the BPI score.

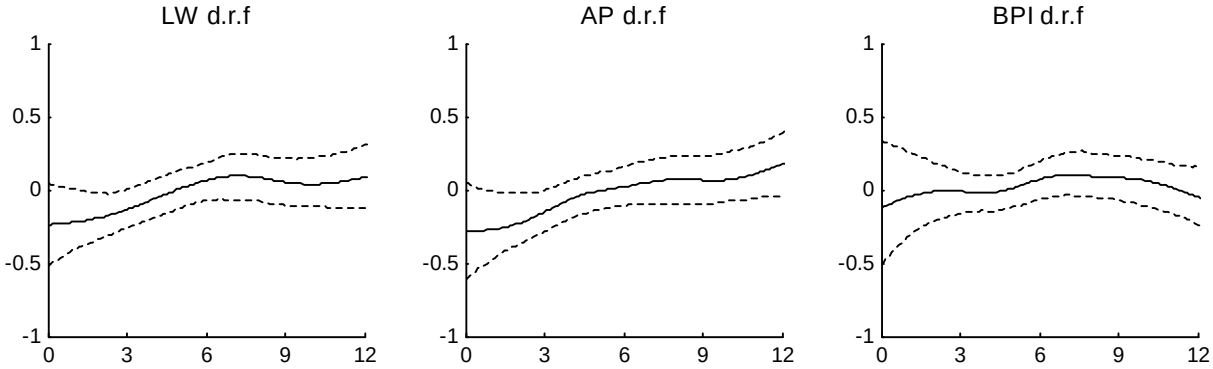

Figure 1- Dose response functions for young children

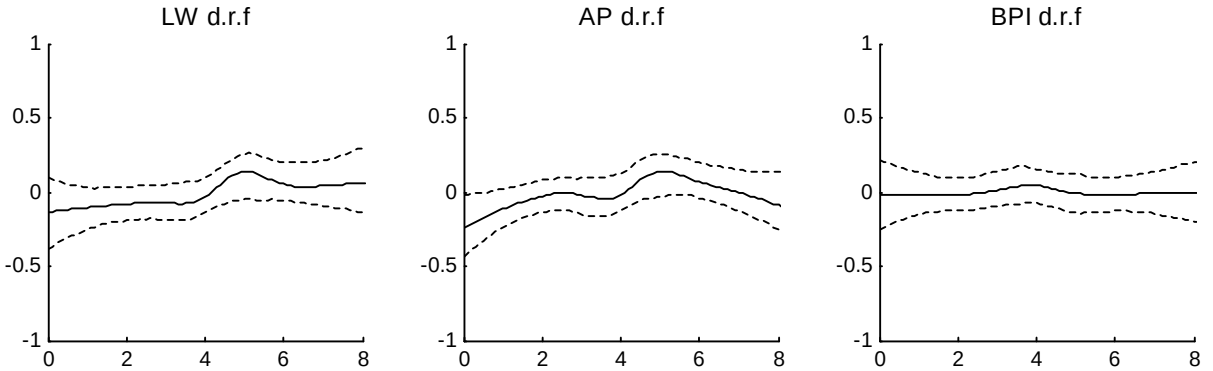

Figure 2- Dose response functions for old children

Notice that these figures allow us to compare the estimated d.r.f. for the different levels of treatment but do not give any information on whether the differences are significant. For instance, for young children, we can tell that spending 12 hours per day with the mother is certainly better for LW score than spending any time at all. However we can not tell whether it is significantly better. To give some insight about the significance of these differences we evaluate the d.r.f. in 5 treatment levels, compute the difference between them and test whether these differences are indeed significant. This information is presented in table 6.

Table 6 - Dose response differences

	Young children					Old children				
	T	0	3	6	9	T	0	2	4	6
LW	3	0.10	-	-	-	2	0.05	-	-	-
	6	0.30*	0.20***	-	-	4	0.10	0.05	-	-
	9	0.29	0.19**	-0.01	-	6	0.18	0.13*	0.08	-
	12	0.32*	0.22	0.02	0.03	8	0.19	0.14	0.09	0.01
AP	3	0.13	-	-	-	2	0.21**	-	-	-
	6	0.30	0.17**	-	-	4	0.22*	0.01	-	-
	9	0.35*	0.22**	0.05	-	6	0.30**	0.09	0.08	-
	12	0.46***	0.33***	0.16	0.11	8	0.15	-0.06	-0.07	-0.16*
BPI	3	0.11	-	-	-	2	0.01	-	-	-
	6	0.19	0.08	-	-	4	0.07	0.06	-	-
	9	0.21	0.09	0.01	-	6	0.01	0.00	-0.06	-
	12	0.08	-0.04	-0.12	-0.13	8	0.01	0.01	-0.05	0.00

For young children we evaluate treatments 0, 3, 6, 9 and 12 and for old children treatments 0, 2, 4, 6 and 8. In each cell of the table we present the difference between the d.r.f. of the treatment

in the row and in the column. For instance in the first cell (top left cell) the 0.10 means that we expect to have a difference of 0.10 in the LW scores of a child that spends 3 hours with the mother and a child that spends no time at all with the mother. However this difference is not significantly different from zero.

Almost all the values for LW and AP are positive for young children, which indeed confirms that the d.r.f. is increasing in the treatment variable. Several values are indeed significant and it is worthwhile mentioning the magnitude of these differences as they reach almost half a standard deviation. Notice that the figures are much smaller for old children than those of young children, in particular for LW. For AP it is significantly better to spend more time with the mother up until the 6 hours per day. In both samples the BPI values are very small and none is significant, which again suggests that the treatment does not have an impact on children behavior.

4.2.2 Age and Race Divided Samples

Figures 3 and 4 present the d.r.f. for young black and white children, respectively.

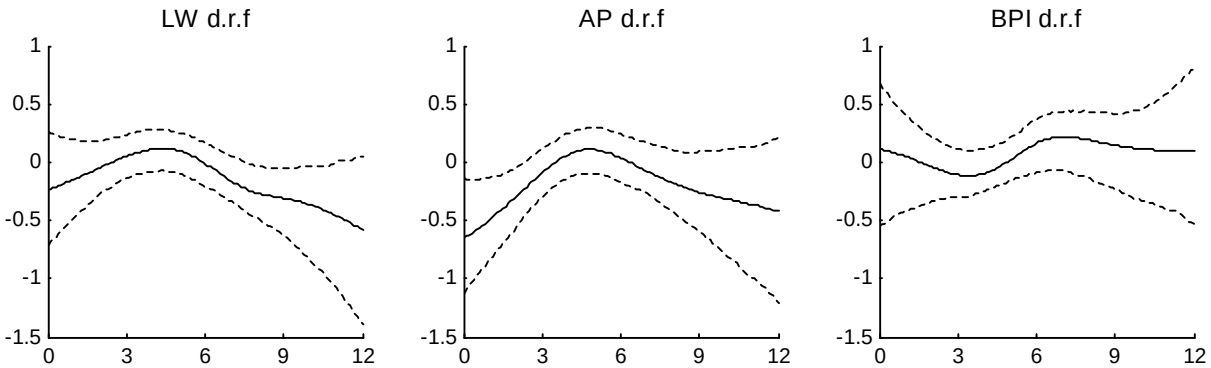

Figure 3- Dose response functions for young black children

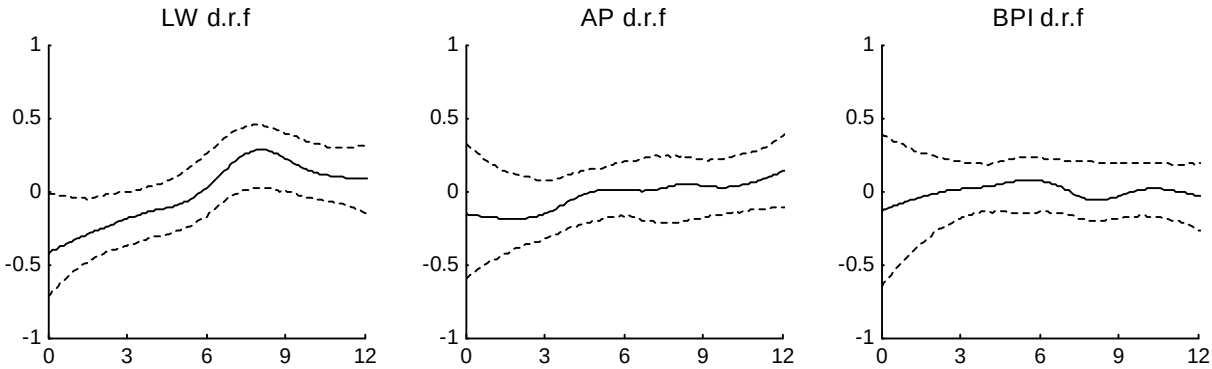

Figure 4- Dose response functions for young white children

Interestingly the d.r.f. present very differently patterns when we take each race separately. For young white children the LW and AP d.r.f. seem to be increasing in the treatment. Even though the LW d.r.f. seems to decrease for higher levels of treatment, overall there is no doubt that it presents an upward trend. This is confirmed by table 7: almost all the values for the white sample are positive and are particularly significant for LW. The black sample present very different d.r.f.. The LW and AP d.r.f. increase for smaller levels of treatment but then present a sharp decrease for high levels of treatment, presenting a peak around the 5 hours per week day. Notice that we expect black children to perform better in LW, and almost the same in AP, if they spend no time with their mothers or if they spend 12 hours per day with them. These findings are supported by the values in table 7: all the significant values in LW are negative and for AP the peak seems

indeed significant. The BPI d.r.f. is rather flat for young white children and for black children there is no clear pattern.

Table 7 - Dose response differences for young children

		Black Young				Young White			
	T	0	3	6	9	0	3	6	9
LW	3	0.29	-	-	-	0.23	-	-	-
	6	0.23	-0.06	-	-	0.44*	0.21*	-	-
	9	-0.07	-0.36**	-0.30*	-	0.64***	0.41***	0.20	-
	12	-0.34	-0.66*	-0.57	-0.27	0.51***	0.27*	0.07	-0.14
AP	3	0.56**	-	-	-	-0.00	-	-	-
	6	0.69***	0.12	-	-	0.16	0.06	-	-
	9	0.39	-0.17	-0.29*	-	0.19	0.19	0.03	-
	12	0.23	-0.33	-0.45	-0.16	0.29	0.29*	0.12	0.09
BPI	3	-0.22	-	-	-	0.15	-	-	-
	6	0.05	0.27*	-	-	0.21	0.06	-	-
	9	0.03	0.25	-0.02	-	0.09	-0.05	-0.11	-
	12	-0.01	0.21	-0.06	-0.04	0.10	-0.05	-0.10	0.00

Figures 5 and 6 present the d.r.f. for black and white old children, respectively.

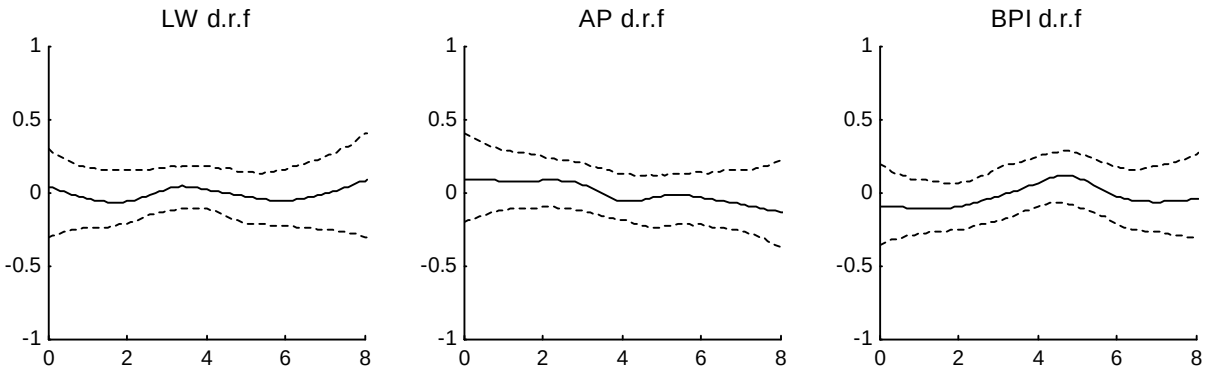

Figure 5- Dose response functions for old black children

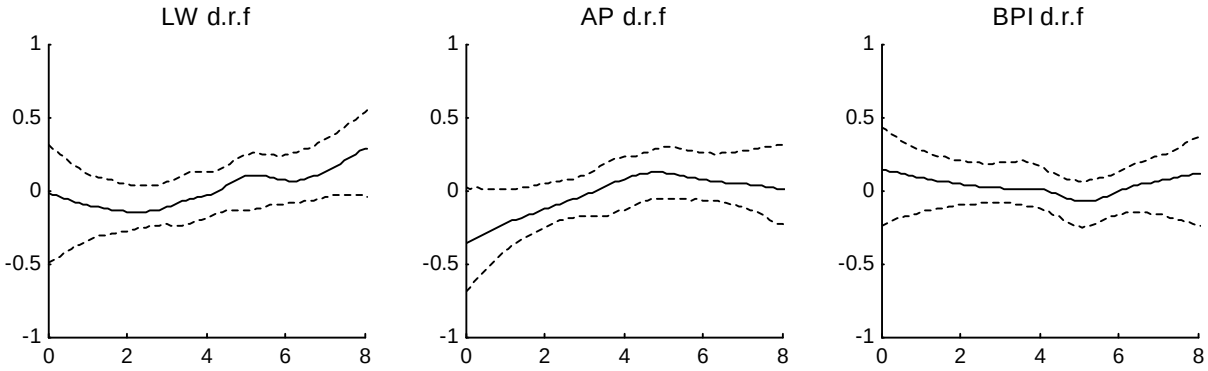

Figure 6- Dose response functions for old white children

For black children the d.r.f. seem flat suggesting that spending more time with the mother may not have a significant effect on child's performance in LW, AP and BPI. Indeed, we can see in table 8 that all the values are insignificant. For white children the LW and AP d.r.f. are overall positively

inclined, even though for LW there is a slight decrease for low levels of the treatment and for AP there is a slight decrease for high levels of treatment. However in table 8 all the significant values for the white sample are positive suggesting that spending more time with the mother increases significantly the performance in LW and AP tests. There are no significant values for the white sample when the BPI d.r.f. is considered.

Table 8 - Dose response differences for old children

		Old Black				Old White			
	T	0	2	4	6	0	2	4	6
LW	2	-0.10	-	-	-	-0.12	-	-	-
	4	-0.02	0.08	-	-	-0.02	0.11	-	-
	6	-0.09	0.01	0.08	-	0.09	0.21*	0.10	-
	8	0.04	0.14	0.06	0.14	0.30	0.43***	0.32*	0.21
AP	2	-0.00	-	-	-	0.23	-	-	-
	4	-0.14	-0.14	-	-	0.43**	0.20*	-	-
	6	-0.12	-0.12	0.02	-	0.43*	0.20**	-0.00	-
	8	-0.22	-0.22	-0.08	-0.09	0.37*	0.14	-0.07	-0.06
BPI	2	-0.00	-	-	-	-0.10	-	-	-
	4	0.16	0.16	-	-	-0.14	-0.04	-	-
	6	0.07	0.07	-0.09	-	-0.15	-0.05	-0.01	-
	8	0.05	0.05	-0.11	-0.02	-0.03	0.07	0.11	0.12

4.3 Sensitivity Analysis

We perform a sensitivity analysis of our estimation by considering higher orders of approximation when estimating (9) and (10). In particular we use a cubic approximation and a 4^{th} degree approximation. Equation (10) for these cases is respectively:

$$\widehat{E[Y(t)]} = \frac{1}{N}\sum_{i=1}^{N}\left[\begin{array}{c}\widehat{\alpha}_0+\widehat{\alpha}_1.t+\widehat{\alpha}_2.t^2+\widehat{\alpha}_3\widehat{r}(t,X_i)+\widehat{\alpha}_4\widehat{r}(t,X_i)^2+\widehat{\alpha}_5.t.\widehat{r}(t,X_i)+\\ +\widehat{\alpha}_6t^3+\widehat{\alpha}_7\widehat{r}(t,X_i)^3+\widehat{\alpha}_8t^2\widehat{r}(t,X_i)+\widehat{\alpha}_9\widehat{r}(t,X_i)^2.t\end{array}\right]$$

$$\widehat{E[Y(t)]} = \frac{1}{N}\sum_{i=1}^{N}\left[\begin{array}{c}\widehat{\alpha}_0+\widehat{\alpha}_1.t+\widehat{\alpha}_2.t^2+\widehat{\alpha}_3\widehat{r}(t,X_i)+\widehat{\alpha}_4\widehat{r}(t,X_i)^2+\widehat{\alpha}_5.t.\widehat{r}(t,X_i)+\\ +\widehat{\alpha}_6t^3+\widehat{\alpha}_7\widehat{r}(t,X_i)^3+\widehat{\alpha}_8t^2\widehat{r}(t,X_i)+\widehat{\alpha}_9\widehat{r}(t,X_i)^2.t+\alpha_{10}t^4\\ +\alpha_{11}\widehat{r}(t,X_i)^4+\alpha_{12}t^3\widehat{r}(t,X_i)+\alpha_{13}\widehat{r}(t,X_i)^3t+\alpha_{14}t^2\widehat{r}(t,X_i)^2\end{array}\right]$$

In figures 7 to 12 (in the appendix) we plot the d.r.f. for the three approximations considered for each of the six samples. Even though there are some differences in detail, the six figures show that in general the d.r.f present the same trends and shapes under the different specifications.

4.4 Falsification exercise

In the appendix.

5 Conclusion

In this paper we use unique data from the PSID that collects a diary on children's time use and test scores. This data allows to measure the time that children spend with mothers during a week day and to assess the importance that this time has on child's cognitive and non-cognitive test scores. We use the treatment evaluation methodology applied for continuous treatments, as our

treatment variable, time, is continuous. This methodology eliminates the selection bias problem and ensure that we find a causal effect between time with mothers and test scores.

We conclude that more time with mothers leads both young and old children to perform better in cognitive tests, and that this effect is more pronounced for younger children. Once we divide our sample according to race and age we find that young black children tend to perform worse in those tests if they spend more than 5 hours in a day with mothers. This negative effect is not present for young white children who perform significantly better as time with mothers increases. For old black children we find no relation between time with mothers and test scores, while for old white children the positive effect of time with mothers still persists. So we conclude that white children always benefit from time spent with mothers in a week day and benefit until later ages. For all the age and race samples we consider, we find no effect between time spent with mothers and non-cognitive test scores.

References

[1] Bernal, R. and Keane, M.

[2] Blau, F. and Grossberg, A. (1992), Maternal Labor Supply and Children's Cognitive Development, *The Review of Economic and Statistics*, vol. 74, n^{o}3, pp. 474-481

[3] Heckman, J., J. Stixrud and S. Urzua (2006), The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior, *Journal of Labor Economics*, vol.24(3), pp. 411-482

[4] Hirano, K. and Imbens, G. (2994), The Propensity Score with Continuous Treatments, in A. Gelman and X. Meng (eds) *Applied Bayesian Modeling and Causal Inference from Incomplete Data Perspectives*

[5] Imbens, G. (2000), The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, vol.87, pp. 706-710

[6] James-Burdumy, S. (2005), The Effect of Maternal Labor Force Participation on Child Development, *Journal of Labor Economics*, vol.23, n^{o}1

[7] Lechner, M. (1999), Identification and estimation of causal effects of multiple treatments under the conditional independence assumption, IZA, discussion paper n^{o}91

[8] Ruhm, C. (2994), Parental Employment and Child Cognitive Development, *The Journal of Human Resources*, vol. 39, n^{o}1, pp. 155-192

[9] Rosenbaum, P. and Rubin, D. (1983), The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, vol. 70, pp- 41-55

A Tables

Table 1- Variable names and definitions

Variable Name	Definition
Outcomes	
LW	Letter- Word Identification
AP	Applied Problems
BPI	Behavior Problem Index
Treatment	
Time	Hours child spend with mother in a week day
Covariates	
Ageatpcg	Child's age at caregiver interview (months)
RaceWhite	Race - White (=1 if yes)
RaceBlack	Race - Black (=1 if yes)
Chgender	Child's gender (=1 if male)
Birthorder	Birth order from mother
Birthweight	Birth weight
Ageatbreast	Age breastfeeding stopped (months)
AgeBirthM	Mother's age at birth
MaritalBirth	Mother's marital status at birth (=1 if married)
EducationM	Mother's years of education
ParentLive	Both parents living (=1 if yes)
ChildrenFU	Number children in household
AgeYoungest	Age youngest child
OtherAdults	Number of other adults (excluding mother)
OwnHouse	Family own house (=1 if yes)
EducationF	Spouse years of education
Avg(Total Income)	5 year average of total income in previous year (log, real terms)
Mother's FE[1]	Mother's fixed effect
AvgWeeklyWage	State's average weekly wage

[1]- We compute the Mother's FE as follows. We have information on the wage the mother receives every 2 yeras from 1990 until 2003, so approximately 7 years before after the CDS data collection point. Furthermore we can compute the age of the mother at each of these data points. We build a panel with the wage, the age and the age squared and run a fixed effect model ($xtreg\ wage\ age\ age^2,\ fe$).

Table 3- Coefficients of the regression (7)

	Young		Old		Young Black		Young White		Old Black		Old White	
	β	S.D.	β	S.D.	β	S.D.	β	S.D.	β	S.D.	β	S.D.
Ageatpcg	−0.05	0.01***	−0.00	0.00	−0.06	0.02***	−0.04	0.02**	−0.01	0.01	−0.00	0.01
RaceWhite	8.60	2.91***	5.09	1.89***	–	–	–	–	–	–	–	–
RaceBlack	7.62	2.82***	5.01	1.85***	–	–	–	–	–	–	–	–
Chgender	−0.07	0.26	−0.14	0.16	−0.18	0.38	0.03	0.37	−0.19	0.25	−0.14	0.22
Birthorder	−0.18	0.26	−0.13	0.12	−0.20	0.35	−0.04	0.39	−0.20	0.16	−0.01	0.18
Birthweight	−0.01	0.09	−0.04	0.06	0.05	0.13	−0.02	0.14	−0.01	0.09	−0.03	0.09
Ageatbreast	0.05	0.02***	0.02	0.02	0.05	0.05	0.06	0.03**	0.03	0.04	0.01	0.02
AgeBirthM	0.07	0.04	0.05	0.03**	0.01	0.06	0.09	0.06	0.00	0.04	0.10	0.04***
MaritalBirth	0.11	0.42	0.21	0.25	0.75	0.57	0.02	0.64	0.49	0.31	−0.13	0.46
EducationM	0.02	0.09	0.04	0.05	0.10	0.13	−0.01	0.13	−0.01	0.08	0.06	0.07
ParentLive	−0.34	1.19	−0.24	0.70	−0.44	2.15	0.48	1.57	−0.71	1.34	0.69	0.92
ChildrenFU	−0.03	0.27	−0.08	0.12	−0.01	0.37	−0.20	0.39	−0.10	0.15	−0.11	0.19
AgeYoungest	0.02	0.13	0.00	0.04	0.02	0.19	−0.06	0.19	0.00	0.05	−0.01	0.06
OtherAdults	−0.23	0.43	0.17	0.20	−0.08	0.51	0.18	0.73	0.59	0.25***	−0.46	0.33
OwnHouse	0.29	0.35	−0.14	0.22	0.30	0.50	0.25	0.48	0.01	0.31	−0.26	0.31
EducationF	0.06	0.08	0.04	0.05	0.01	0.16	0.04	0.11	0.04	0.10	0.02	0.06
Avg(Total Income)	−0.33	0.23	−0.29	0.15*	−0.77	0.33**	−0.07	0.34	−0.46	0.24*	−0.19	0.21
Mother's FE	−0.43	0.12***	−0.19	0.07***	−0.35	0.20*	−0.43	0.15***	−0.11	0.12	−0.24	0.08***
AvgWeeklyWage	0.00	0.00	0.00	0.00***	0.00	0.00*	0.00	0.00	0.00	0.00***	0.00	0.00
Intercept	0.00	0.00	0.00	0.00	11.6	4.55***	5.53	3.98	7.89	3.17***	3.24	2.52
Observations	516		936		220		341		396		540	

Table 5a- Estimated dose response function

	Young			Old		
Outcome	LW	AP	BPI	LW	AP	BPI
$Intercept$	-0.24	-0.28	-0.12	-0.13	-0.23	-0.02
$(s.e.)$	0.18	0.18	0.17	0.12	0.12^{**}	0.11
$Time$	0.01	-0.00	0.09	0.03	0.13	0.00
$(s.e.)$	0.07	0.07	0.07	0.07	0.07^{**}	0.07
$Time^2$	0.00	0.00	-0.01	-0.00	-0.01	-0.00
$(s.e.)$	0.01	0.01	0.01	0.01	0.01	0.01
GPS	1.57	3.87	-3.75	-0.98	-1.55	0.58
$(s.e.)$	2.02	2.03	1.95	0.95	0.95	0.94
GPS^2	-4.83	-10.6	8.38	-0.77	-0.06	-0.07
$(s.e.)$	5.49	5.51^{*}	5.33	0.81	0.82	0.80
$Time * GPS$	0.08	-0.03	0.19	0.36	0.35	-0.11
$(s.e.)$	0.19	0.19^{*}	0.18	0.18^{**}	0.18^{*}	0.18
N	491	487	529	827	825	879

Table 5b- Estimated dose response function

	Young Black			Old Black		
Outcome	LW	AP	BPI	LW	AP	BPI
$Intercept$	-0.25	-0.68	0.14	0.04	0.08	-0.09
$(s.e.)$	0.13	0.33^{**}	0.31	0.16	0.16	0.15
$Time$	0.08	0.11	-0.01	-0.09	0.00	-0.02
$(s.e.)$	0.14	0.14	0.14	0.01	0.01	0.01
$Time^2$	-0.01	-0.01	0.00	0.01	-0.00	0.00
$(s.e.)$	0.01	0.01	0.01	0.01	0.01	0.01
GPS	0.66	5.66	-5.60	1.82	0.66	0.66
$(s.e.)$	3.52	3.55	3.50	1.47	1.48	1.45
GPS^2	9.97	-6.23	6.74	-1.69	-1.39	-2.18
$(s.e.)$	10.2	10.2	10.1	1.93	1.94	1.91
$Time * GPS$	-0.44	-0.41	0.73	-0.15	-0.04	0.22
$(s.e.)$	0.34	0.34	0.33^{**}	0.25	0.25	0.25
N	188	188	200	363	361	370

Table 5c- Estimated dose response function (White)

	Young White			Old White		
Outcome	LW	AP	BPI	LW	AP	BPI
$Intercept$	-0.42	-0.15	-0.13	-0.02	-0.36	0.14
$(s.e.)$	0.23^{*}	0.23	0.22	0.19	0.19^{*}	0.18
$Time$	0.09	-0.03	0.07	-0.09	0.14	0.14
$(s.e.)$	0.09	0.09	0.09	0.11	0.11	0.10
$Time^2$	-0.00	0.00	0.00	0.02	-0.01	0.01
$(s.e.)$	0.00	0.01	0.01	0.01	0.01	0.01
GPS	-0.99	4.43	-0.98	0.53	0.12	0.29
$(s.e.)$	2.47	2.54	2.41	1.36	1.36	1.27
GPS^2	-6.61	-10.2	8.25	-1.71	-0.23	0.81
$(s.e.)$	5.66	5.79^{*}	5.53	1.14	1.15	1.10
$Time * GPS$	0.53	-0.11	-0.27	0.17	0.04	-0.19
$(s.e.)$	0.25^{**}	0.26^{*}	0.24	0.25	0.25	0.24
N	299	295	325	461	461	505

B Sensitivity Analysis Figures

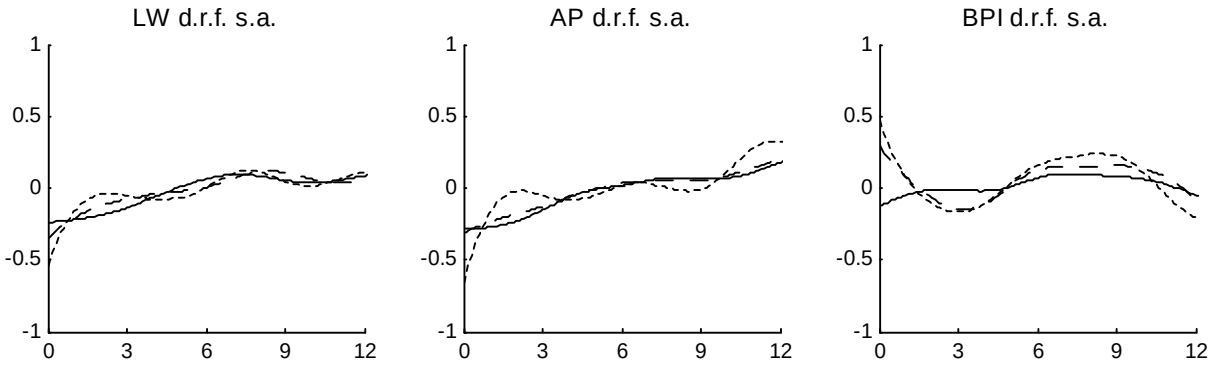

Figure 7- Dose response functions for young children: quadratic approximation (solid line), cubic approximation (dashed line) and 4^{th} degree approximation (dotted line)

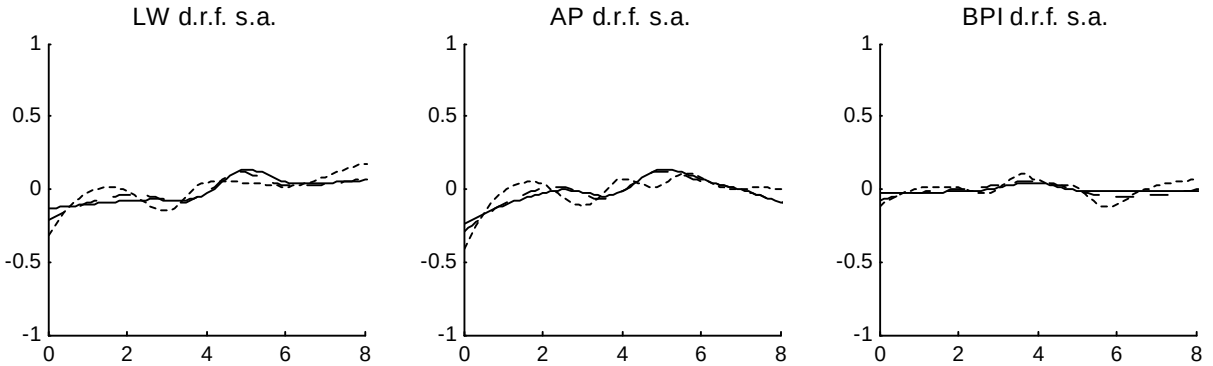

Figure 8- Dose response functions for old children: quadratic approximation (solid line), cubic approximation (dashed line) and 4^{th} degree approximation (dotted line)

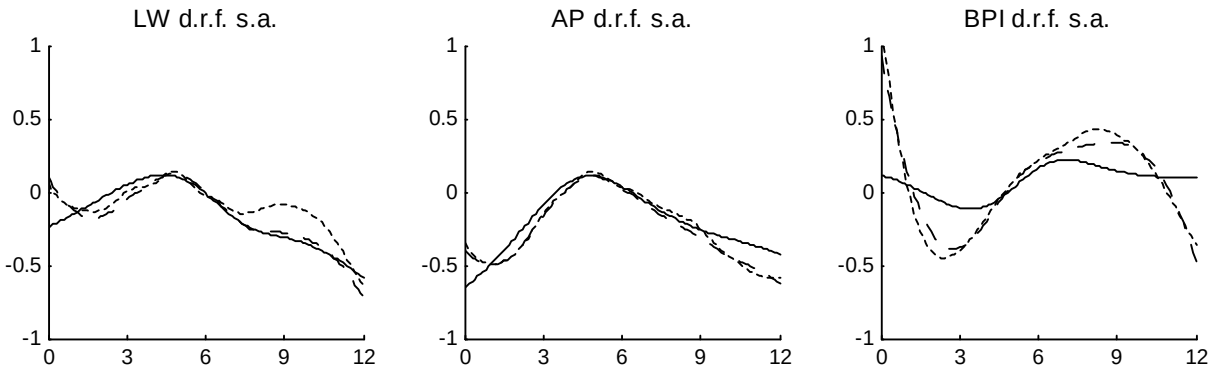

Figure 9- Dose response functions for young black children: quadratic approximation (solid line), cubic approximation (dashed line) and 4^{th} degree approximation (dotted line)

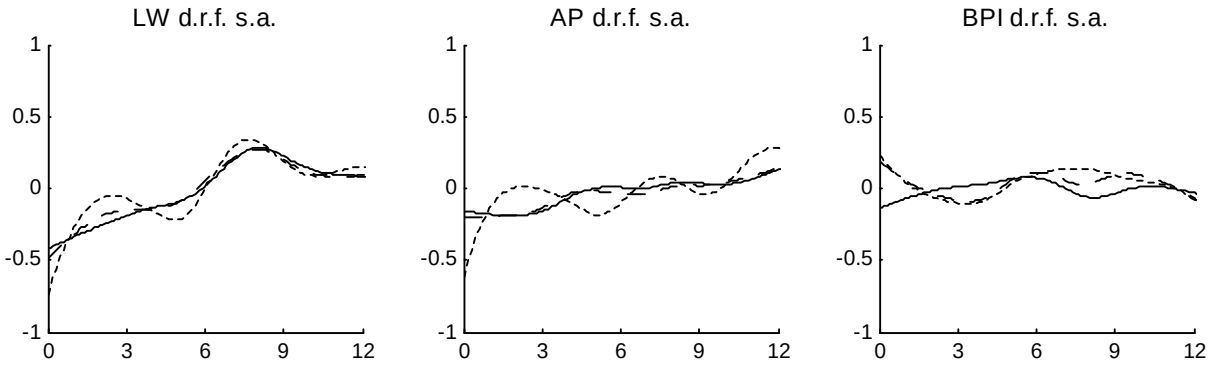

Figure 10- Dose response functions for young white children: quadratic approximation (solid line), cubic approximation (dashed line) and 4^{th} degree approximation (dotted line)

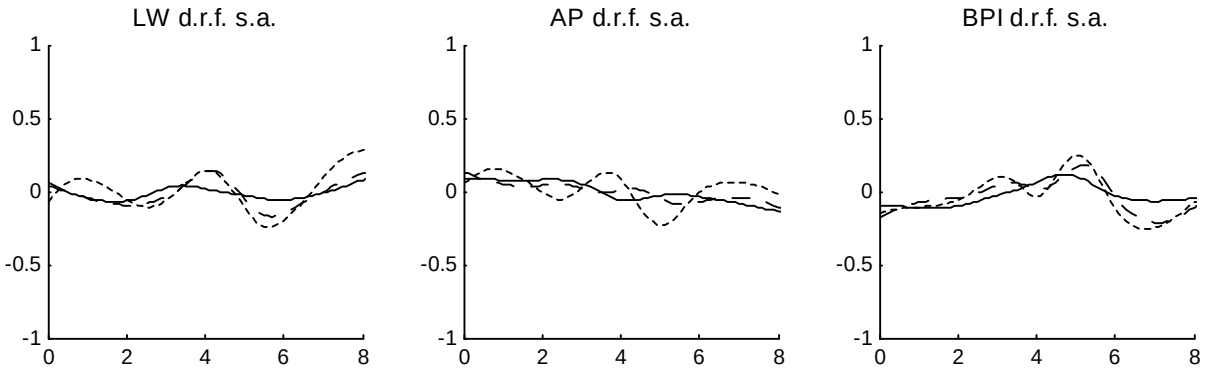

Figure 11- Dose response functions for old black children: quadratic approximation (solid line), cubic approximation (dashed line) and 4^{th} degree approximation (dotted line)

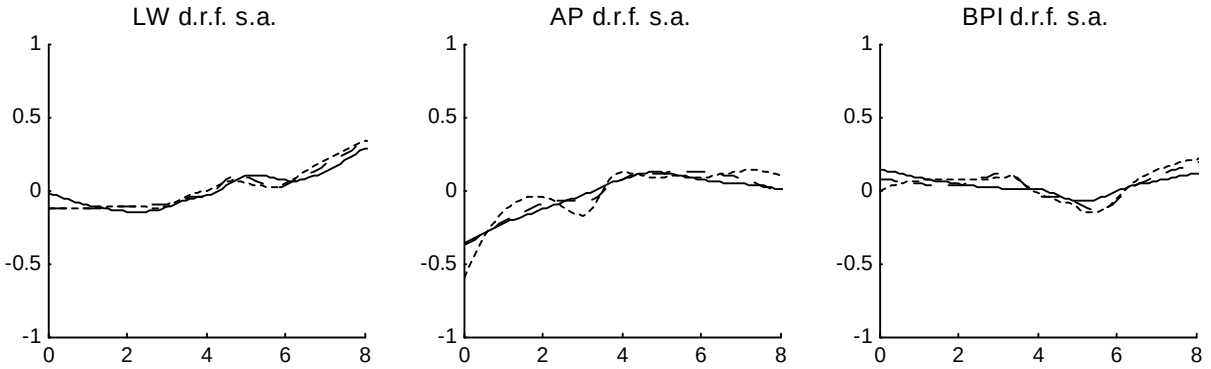

Figure 12- Dose response functions for old white children: quadratic approximation (solid line), cubic approximation (dashed line) and 4^{th} degree approximation (dotted line)

C Falsification Exercise

To validate our identification assumptions we do a falsification exercise. We take a variable that intuitively should not be affected by the treatment and use it as the outcome variable in steps 2 and 3 presented in the section implementation. We use the years of education of the child's grandfather (GrandF) and grandmother (GrandM) for this exercise. Our estimated dose response functions should be rather flat for our procedure to be valid as this would mean that child's grandfather and grandmother education is not affected by the time the mother spends with the child as one would expect.

Similar to the text we present the dose response functions and the tables.

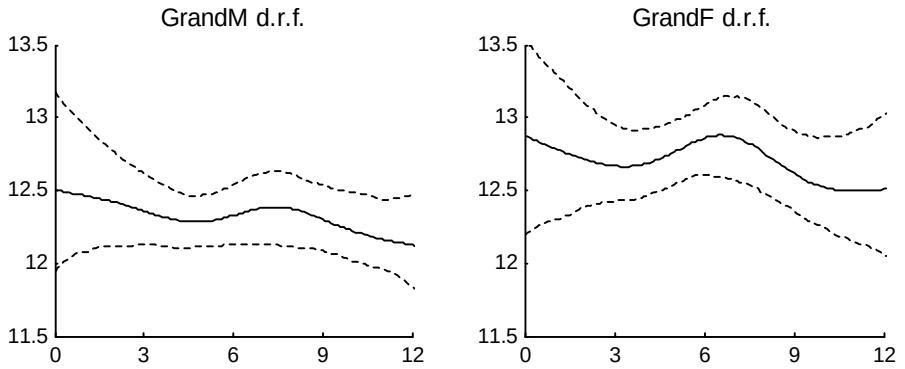

Figure 13- Grandmother's and grandfather's dose response functions for young children

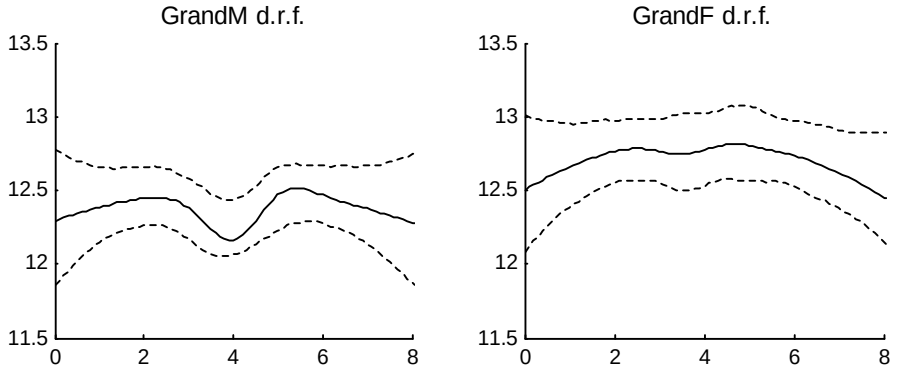

Figure 14- Grandmother's and grandfather's dose response functions for old children

Table 9 - Dose response differences for falsification exercise

		Young children					Old children			
	T	0	3	6	9	T	0	2	4	6
GrandM	3	-0.14	-	-	-	2	0.15	-	-	-
	6	-0.17	-0.03	-	-	4	-0.13	-0.28**	-	-
	9	-0.20	-0.06	-0.03	-	6	0.17	0.02	0.30**	-
	12	-0.37	-0.23	-0.20	-0.17	8	-0.01	-0.17	0.11	-0.19
GrandF	3	-0.21	-	-	-	2	0.26	-	-	-
	6	-0.02	0.19	-	-	4	0.27	0.01	-	-
	9	-0.26	-0.06	-0.25*	-	6	0.23	-0.03	-0.04	-
	12	-0.37	-0.16	-0.35	-0.10	8	-0.05	-0.31	-0.32	-0.28*

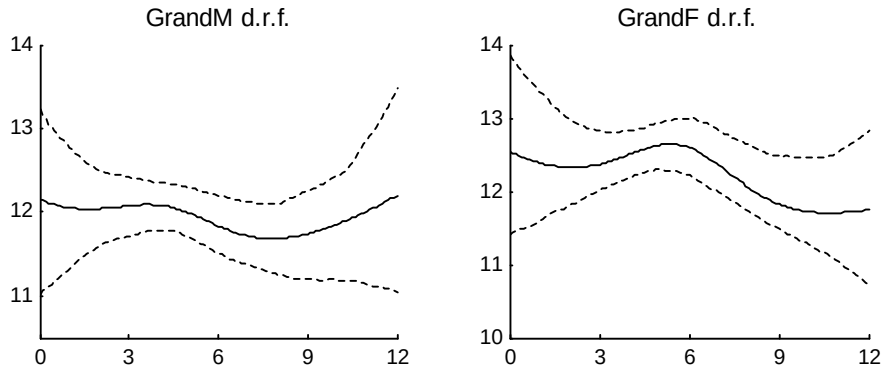

Figure 15- Grandmother's and grandfather's dose response functions for young black children

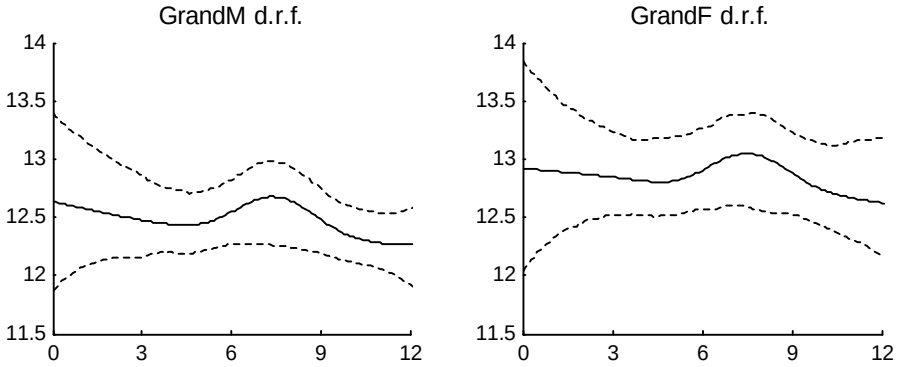

Figure 16- Grandmother's and grandfather's dose response functions for young white children

Table10 - Dose response differences for falsification exercise

		Black Young				Young White			
	T	0	3	6	9	0	3	6	9
LW	3	-0.07	-	-	-	-0.16	-	-	-
	6	-0.31	-0.24	-	-	-0.09	0.07	-	-
	9	-0.40	-0.33	-0.09	-	-0.16	0.00	-0.07	-
	12	0.05	0.12	0.36	0.45	-0.37	-0.21	-0.28	-0.21
AP	3	-0.17	-	-	-	-0.07	-	-	-
	6	0.06	0.23	-	-	-0.02	0.06	-	-
	9	-0.71	-0.55*	-0.77***	-	-0.04	0.03	-0.02	-
	12	-0.79	-0.62	-0.85	-0.07	-0.29	-0.22	-0.28	-0.25

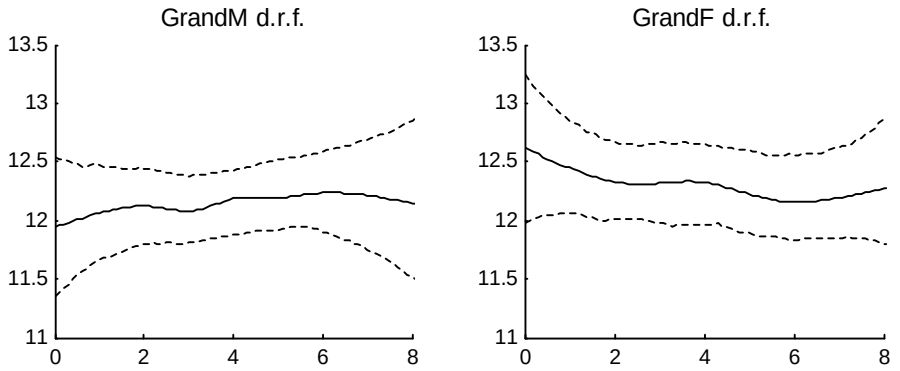

Figure 17- Grandmother's and grandfather's dose response functions for old black children

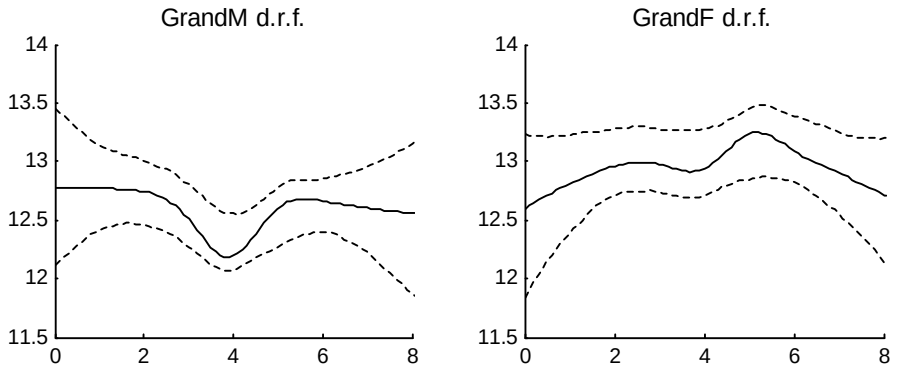

Figure 18- Grandmother's and grandfather's dose response functions for old white children

Table 8 - Dose response differences for falsification exercise

		Old Black				Old White			
	T	0	2	4	6	0	2	4	6
LW	2	0.18	-	-	-	-0.04	-	-	-
	4	0.24	0.06	-	-	-0.59	-0.55***	-	-
	6	0.29	0.11	0.05	-	-0.12	-0.08	0.47***	-
	8	0.20	0.23	-0.04	-0.09	-0.23	-0.19	0.36	-0.11
AP	2	-0.29	-	-	-	0.36	-	-	-
	4	-0.29	-0.01	-	-	0.33	-0.02	-	-
	6	-0.46	-0.17	-0.16	-	0.49	0.13	0.16	-
	8	-0.34	-0.05	-0.05	0.11	0.11	-0.24	-0.22	-0.39