9-2007

Dynamic Treatment Effect Analysis of TV Effects on Child Cognitive Development

Fali HUANG
Singapore Management University, flhuang@smu.edu.sg

SMU ECONOMICS & STATISTICS WORKING PAPER SERIES

Dynamic Treatment Effect Analysis of TV Effects on Child Cognitive Development

Fali Huang

September 2007

Paper No. 10-2007

Dynamic Treatment Effect Analysis of TV Effects on Child Cognitive Development

(September 21, 2007)

Fali Huang
School of Economics and Social Sciences
Singapore Management University
90 Stamford Road, Singapore 178903
flhuang@smu.edu.sg;
fax: 65-6828-0833

Myoung-jae Lee*
Department of Economics
Korea University
Anam-dong, Sungbuk-ku
Seoul 136-701, South Korea
myoungjae@korea.ac.kr
phone: 82-2-3290-2229

We investigate whether TV watching at ages 6-7 and 8-9 affects cognitive development measured by math and reading scores at ages 8-9 using a rich childhood longitudinal sample from NLSY79. Dynamic panel data models are estimated to handle the unobserved child-specific factor, endogeneity of TV watching, and *dynamic nature of the causal relation*. A special emphasis is put on the last aspect where TV watching affects cognitive development which in turn affects the future TV watching. When this feedback occurs, it is not straightforward to identify and estimate the TV effect. We adopt estimation methods available in the biostatistics literature which can deal with the feedback feature; we also apply the "standard" econometric panel data IV approaches. Overall, for math score at ages 8-9, we find that watching TV for more than two hours per day during ages 6-9 has a negative total effect mostly due to a large negative effect of TV watching at the younger ages 6-7. For reading score, there are evidences that TV watching between 2-4 hours per day has a positive effect whereas the effect is negative outside this range. In both cases, however, the effect magnitudes are economically small.

Key Words: TV watching, treatment effect, panel data, dynamic model, Granger causality.

JEL Classification Numbers: C33, I20, J13, E60.

* Corresponding Author.

** The authors are grateful to the anonymous referees for their detailed comments.

1 Introduction

The U.S. children spend the second largest chunk of their waking time on watching TV (Juster and Stafford, 1991). That is, the most time-consuming activity after attending school is TV watching. For example, an average eight-year old in NLSY79 (National Longitudinal Survey of Youth 1979) child sample spends about 25 hours per week in front of the television. Not surprisingly, the public and parents have been concerned about potentially bad effects of child TV watching. The goal of this paper is to find *the effects of TV watching on child cognitive development measured by standardized mathematics and reading scores.*

The cognitive development in early childhood may be crucial to human capital formation in later years, since "success or failure at this stage feeds into success or failure in school which in turn leads to success or failure in post-school learning" (Heckman, 1999). This relation is just one of many such relations where something happened far back has long-term lingering effects. For instance, the skill heterogeneity at age 16 is shown to account for as much as 90% of the total variation of one's lifetime earnings by Keane and Wolpin (1997).

There are, however, a number of difficulties in establishing the causal link between TV watching and cognitive development. First, inappropriate home and school inputs (for instance, economically and intellectually deficient environments) may induce both more TV watching and lower test scores. This is an omitted variable (or unobserved confounder) problem, which can be resolved by detailed data with sufficient "environmental" control variables. Second, children and the parents may share predispositions for certain habits and behaviors, which cannot be measured. This is an 'unit-specific effect' problem, which may be overcome with panel data. Third, TV watching can affect cognitive development, which can in turn affect the future TV watching. This is an issue of *dynamic treatment effects with feedback*, calling for a proper dynamic model and estimation method. With a rich childhood longitudinal sample from NLSY79 child data that have not only child characteristics and family background variables but also detailed home and school inputs in the current and earlier periods, we will estimate dynamic models to overcome these problems.

Possibly due to the above difficulties, there has been hardly any research in the economics literature for the effects of TV watching on cognitive development. An exception is Zavodny (2006) who studies the effects of TV watching for high school students and reviews related studies in other social sciences. Her main finding is that, although there exist significantly negative effects in cross-sectional results, TV watching has no effect on test scores once individual-specific effects are taken into account. Differently from Zavodny (2006) who deals with individual-specific effects with panel data and family-specific effects with twins and siblings data, we tackle the dynamic (feedback) nature of the causal relation; also we focus on young children, not on high school students. Another exception is Gentzkow and Shapiro (2007) who use the well known 'Coleman study' data and difference-in-differences methods, taking advantage of different TV-introduction timings across regions. Their preferred estimate shows a small (about 2% of one standard deviation) but positive effect of TV watching on test scores at ages 11-17. As in Zavodny (2006), this study does not address the dynamic nature of the causal relation.

The impact of TV viewing on child development has been studied also in the literature of communication, child development, and other psychological domains. These non-economic literature tends to focus more on content-specific studies through experiments and detailed routes for how TV viewing may affect child development. As far as the effects of TV viewing time are concerned, the statistical analysis is mostly on the correlation rather than the causal relationship between TV viewing and child development, and most studies found modest and negative associations of TV viewing with cognitive development of young children (Van Evra, 2004, Zimmerman and Christakis 2005, and Anderson 2005).

Putting our main findings in advance, the TV watching effects are statistically significant, but economically small in terms of their magnitude. The directions of the effects vary across ages, and differ for math and reading scores—two different measures of cognitive development used in this paper. Our preferred estimates indicate that the total effect of watching more than two hours TV per day during ages 6-7 and 8-9 is negative for math score at age 8-9, while a coherent pattern emerging from various specifications is that between two and four hours TV time per day seems to bring the best reading score than too much or too little TV watching.

This paper differs from the TV effect literature in economics on several grounds. First, we introduce and modify a method in the biostatistics literature to deal with the dynamic feedback feature of TV watching on test scores. Second, we focus on the effects of TV watching for children at ages 8-9, who are younger than the groups in the two studies mentioned above. Third, with a large set of control variables including detailed home and school inputs as well as family backgrounds, we still find significant concave/convex effects of TV watching on both scores based on conventional methods of least squares estimator (LSE) and fixed effects models, which differs from Zavodny (2006). Additionally, when the dynamic feedback effects of earlier TV watching are taken into account, the estimated overall effects of TV watching at ages 6-7 and 8-9 are significantly negative for math score, mostly due to a large harmful effects of TV viewing at early ages 6-7. This contrasts with the positive effects found by Gentzkow and Shapiro (2007). On the other hand, our results are consistent with these former studies on the small magnitudes of TV effects, and on the possible positive effects of TV viewing for reading score at relatively older ages.

Section 2 reviews 'G-algorithm' used in biostatistics that identifies the desired full TV effects taking the dynamic feedback feature into account, and then shows how to implement G-algorithm in practice using linear models. As a comparison to G-algorithm, Section 3 presents typical econometric dynamic panel data approaches and points out that these approaches misses an important part of the TV effect, although they relax the critical 'selection-on-observable' assumption in G-algorithm. Section 4 describes our data. Section 5 presents the empirical findings. Finally, Section 6 concludes.

2 G-algorithm and Its implementation

This section explains G-algorithm and how it can be implemented in practice. First, we explain the basics of G-algorithm, which needs regression functions and conditional densities as well as carrying out integration with them. Second, we present a simple model to illustrate how G-algorithm works and when it may fail. Third, as G-algorithm is difficult to implement non-parametrically in general, we examine how to simplify the G-algorithm with linear models. This modified version of G-algorithm is much easier to implement as it requires essentially only LSE. Fourth, G-algorithm for binary responses is presented, which could be the easiest to apply in practice

2.1 G-algorithm Basics

Suppose

$$(x_{i0}, y_{i0},\ x'_{i1}, d_{i1}, y_{i1},\ x'_{i2}, d_{i2}, y_{i2}),\ i = 1, ..., N,\ \text{are observed and iid across } i = 1, ..., N.$$

We will often omit the subscript i in the rest of this paper in view of the iid assumption. In each period, x_{it} is the "period-t baseline" covariate which can affect the treatment d_{it} and response y_{it}, and d_{it} then may affect y_{it}. The dynamic framework with feedback to be dealt with is, omitting x_{it}'s,

$$\begin{array}{cccccc}
d_2 & \longrightarrow & \longrightarrow & \longrightarrow & y_2 \\
 & \nwarrow & & \nearrow & \uparrow \\
 & & y_1 & & \uparrow \\
 & \nearrow & & & \uparrow \\
d_1 & \longrightarrow & \longrightarrow & \longrightarrow & \uparrow
\end{array}$$

The feedback feature is $y_1 \longrightarrow d_2$: d_2 gets adjusted after the interim response y_1 has been observed. This sounds natural: parents adjust their children's TV watching hours having seen their test scores, but this would make d_2 non-randomized even if d_1 were so.

Define the 'potential responses' for the observed responses y_1 and y_2:

y_1^j : potential response when treatment j is given exogenously at time 1,

y_2^{jk} : potential response when treatments j, k are given exogenously at time $1, 2$, $\quad j, k \in [0, \infty)$.

With $d_1 = j$ and $d_2 = k$ observed, we have $y_1 = y_1^j$ and $y_2 = y_2^{jk}$; i.e., only the potential responses corresponding to the realized treatment levels are observed, and all the other potential responses—'counter-factuals'—are not. Also define the 'potential treatment' for d_2:

d_2^j : potential treatment when treatment j is given exogenously at time 1 (thus y_1^j realized);

with $d_1 = j$ observed, we have $d_2 = d_2^j$.

Our goal is to find the mean effect $E(y_2^{j_o k_o} - y_2^{00})$ for the treatment 'profile' (j_o, k_o) versus no treatment at all. Although the mean effect on some treated group may be also of interest, there are problems in identifying the effect on the treated unless essentially the treated is only for the first period; this is discussed in detail by Lechner and Miquel (2001). In this paper, we take the position that TV effect is of interest to the entire population, not just to some subpopulation. Nevertheless, if desired, our models allow interaction terms between the treatments and some elements of the covariates/lagged-responses, which can account for the effect on the subpopulation characterized by those elements.

Before proceeding further, one word on conditioning notations: when a random vector appears in a conditioning set, it means that the condition holds for each support point of the random vector. In relation to this, we will assume that any 'intervention' value (j,k) on (d_1, d_2) falls in the support of the random vector (d_1, d_2).

Let

$$X_y \equiv (x_0', y_0, x_1', x_2')'$$

and denote the conditional independence of a and b given c as '$a \amalg b|c$'. Assume '*no unobserved confounder' (NUC)*:

$$\begin{aligned} \text{Nuc1} &: \ y_2^{jk} \amalg d_1|X_y, \quad \text{for } j = j_o, 0 \text{ and } k = k_o, 0 \\ \text{Nuc2} &: \ y_2^{jk} \amalg d_2^j|(d_1 = j, y_1^j, X_y), \qquad \text{for } j = j_o, 0 \text{ and } k = k_o, 0. \end{aligned}$$

Nuc1 holds if d_1 is determined by X_y and some error term independent of y_2^{jk} given X_y. Saying "d_1 determined by X_y" may sound preposterous because parts of X_y are realized after d_1, but the dependence between d_1 and those components of X_y in Nuc1 should be construed as d_1 being allowed to affect y_2^{jk} through those components in X_y. If one finds Nuc2 somewhat confusing, it may help to rephrase it with densities: with $f(y_{2o}^{jk}|d_{2o}^j, d_{1o}, y_{1o}^j, X_{yo})$ denoting the conditional density/probability of $y_2^{jk}|(d_2^j, d_1, y_1^j, X_y)$ evaluated at $(y_{2o}^{jk}, d_{2o}^j, d_{1o}, y_{1o}^j, X_{yo})$, Nuc2 states

$$f(y_{2o}^{jk}|d_{2o}^j, j, y_{1o}^j, X_{yo}) \text{ does not change as } d_{2o}^j \text{ changes.}$$

Both Nuc1 and Nuc2 are 'selection-on-observables', because Nuc1 states that the d_1-selection is independent of y_2^{jk} given the observable X_y, and Nuc2 states that the d_2^j selection is independent of y_2^{jk} given the observable $(d_1 = j, y_1^j, X_y)$.

G-algorithm (*or G-estimation*) under NUC is (see Robins, 1986 (with errata and appendum 1987), 1998, 1999 and the references therein)

$$E(y_2^{jk}|X_y) = \int E(y_2|d_1 = j, d_2 = k, y_1, X_y) f(y_1|d_1 = j, X_y) \partial y_1 \tag{2.1}$$

where $f(y_1|d_1 = j, X_y)$ denotes the conditional density of $y_1|(d_1 = j, X_y)$, '∂' is used for integration/differentiation to avoid the confusion with treatment d. In (2.1), y_1 is an integration dummy, not a random variable. To distinguish random variables from constants, it may be better to use notations such as y_{1o}, instead of y_1, to write (2.1) as

$$E(y_2^{jk}|X_y) = \int E(y_2|d_1 = j, d_2 = k, y_1 = y_{1o}, X_y) f(y_{1o}|d_1 = j, X_y) \partial y_{1o}.$$

But to save/simplify notations, we will keep writing as in (2.1).

The important point is that the right-hand side of (2.1) is identified, and so is the conditional mean $E(y_2^{jk}|X_y)$. The equality holds because the right-hand side is

$$\begin{aligned}
&\int E(y_2^{jk}|d_1 = j, d_2^j = k,\ y_1^j, X_y) f(y_1^j|d_1 = j, X_y)\partial y_1^j \\
&= \int E(y_2^{jk}|d_1 = j,\ y_1^j, X_y) f(y_1^j|d_1 = j, X_y)\partial y_1^j \qquad \text{(due to Nuc2)} \qquad (2.2) \\
&= E(y_2^{jk}|d_1 = j,\ X_y) \qquad \text{(for } y_1^j \text{ is integrated out)} \\
&= E(y_2^{jk}|X_y) \qquad \text{(due to Nuc1).}
\end{aligned}$$

The role of Nuc1 in G-algorithm is relatively minor, as it is to remove $d_1 = j$ in the conditioning set at the last stage. When Nuc1 does not hold, one may go for $E(y_2^{jk}|d_1 = j,\ X_y) - E(y_2^{j0}|d_1 = j,\ X_y)$, which is the mean treatment effect on the "treated ($d_1 = j$)"—a kind of treatment effects examined in Lechner and Miquel (2001) as mentioned ahead.

In essence, G-algorithm starts with the mean of y_2^{jk} for the subpopulation ($d_1 = j, d_2 = k, y_1$). The condition $d_1 = j, d_2 = k$ is needed because y_2^{jk} is observed only for those with $d_1 = j, d_2 = k$, and y_1 is needed to account for the dynamic feedback nature. Then the subpopulation is generalized to the whole population (i.e., the selection problem is ruled out) as $d_1 = j$ and $d_2 = k$ are removed by Nuc1 and Nuc2, respectively, and y_1 is removed by integration.

Getting $E(y_2^{j_o k_o}|X_y)$ and $E(y_2^{00}|X_y)$ and then integrating out X_y, we obtain the desired (marginal) effect:

$$E(y_2^{j_o k_o} - y_2^{00}) = E_{X_y}[\{\text{rhs of (2.1) for } j = j_o, k = k_o\} - \{\text{rhs of (2.1) for } j = 0, k = 0\}]$$

where $E_{X_y}[\cdot]$ means integrating out X_y using its population distribution and 'rhs' stands for 'right-hand side'. Even for two periods, implementing G-algorithm in (2.1) requires finding $E(y_2|d_1 = j, d_2 = k, y_1, X_y)$ and $f(y_1|d_1 = j, X_y)$ first, and then carrying out the one-dimensional integration, which could be daunting. Part of this problem can be avoided by adopting linear models.

2.2 A Simple Linear Model for G-algorithm

It is illuminating to see G-algorithm with a simple linear model:

$$\begin{aligned}
d_1 &= \zeta_1 + \zeta_x x_1 + \varepsilon_1, \\
y_1 &= \beta_1 + \beta_x x_1 + \beta_d d_1 + u_1 \quad (d_1 \text{ affects } y_1), \\
d_2 &= \zeta_1 + \zeta_x x_2 + \zeta_y y_1 + \varepsilon_2 \quad (y_1 \text{ affects } d_2), \\
y_2 &= \beta_1 + \beta_x x_2 + \beta_{dlag} d_1 + \beta_d d_2 + \beta_y y_1 + u_2 \quad (d_1, d_2, y_1 \text{ affect } y_2), \\
& \quad \varepsilon_1, \varepsilon_2, x_1, x_2, u_1, u_2 \text{ are iid } N(0,1)
\end{aligned} \tag{2.3}$$

where ζ and β are parameters, $(\varepsilon_1, \varepsilon_2, u_1, u_2)$ are mean-zero errors, and the $N(0,1)$ assumption for the random variables is only for ease of exposition. Corresponding to the y_1 and y_2 equations, we get

$$\begin{aligned}
y_1^j &= \beta_1 + \beta_x x_1 + \beta_d j + u_1, \\
y_2^{jk} &= \beta_1 + \beta_x x_2 + \beta_{dlag} j + \beta_d k + \beta_y y_1^j + u_2, \\
d_2^j &= \zeta_1 + \zeta_x x_2 + \zeta_y y_1^j + \varepsilon_2.
\end{aligned} \tag{2.4}$$

Hence the desired effect is

$$E(y_2^{j_o k_o} - y_2^{00}) = \beta_{dlag} j_o + \beta_d k_o + \beta_y E(y_1^{j_o} - y_1^0) = \beta_{dlag} j_o + \beta_d k_o + \beta_y \beta_d j_o. \tag{2.5}$$

To see that G-algorithm identifies this effect, examine (2.2) for the linear model:

$$\begin{aligned}
& E(y_2^{jk}|X_y) = \int (\beta_1 + \beta_x x_2 + \beta_{dlag} j + \beta_d k + \beta_y y_1^j) f(y_1^j | d_1 = j, X_y) \partial y_1^j \\
= \; & \beta_1 + \beta_x x_2 + \beta_{dlag} j + \beta_d k + \beta_y E(y_1^j | d_1 = j, X_y) \quad \text{(integration becomes averaging)} \\
= \; & \beta_1 + \beta_x x_2 + \beta_{dlag} j + \beta_d k + \beta_y \{\beta_1 + \beta_x x_1 + \beta_d j + E(u_1 | d_1 = j, X_y)\} \\
= \; & \beta_1 + \beta_x x_2 + \beta_{dlag} j + \beta_d k + \beta_y (\beta_1 + \beta_x x_1 + \beta_d j)
\end{aligned}$$

because $E(u_1|d_1 = j, X_y) = E(u_1 | \varepsilon_1 = j - \zeta_1 - \zeta_x x_1, X_y) = E(u_1|X_y) = 0$. Thus

$$E(y_2^{jk} - y_0^{00} | X_y) = \beta_{dlag} j + \beta_d k + \beta_y \beta_d j = E(y_2^{jk} - y_2^{00}).$$

This can be obtained also directly by using the y_2^{jk} 'reduced form' equation with y_1^j substituted out:

$$\begin{aligned}
y_2^{jk} &= \beta_1 + \beta_x x_2 + \beta_{dlag} j + \beta_d k + \beta_y (\beta_1 + \beta_x x_1 + \beta_d j + u_1) + u_2 \\
&= \beta_1 + \beta_y \beta_1 \; + \beta_x x_2 + \beta_y \beta_x x_1 \; + \beta_{dlag} j + \beta_d k + \beta_y \beta_d j \; + (\beta_y u_1 + u_2);
\end{aligned} \tag{2.6}$$

the terms with j and k constitute $E(y_2^{jk} - y_0^{00})$.

We can verify Nuc1 for the linear model. From (2.4), given $(X_y, d_1 = j, y_1^j)$, y_2^{jk} depends only on u_2 whereas d_2^j depends only on ε_2. Since

$$\varepsilon_2 \amalg u_2 | (X_y, d_1 = j, y_1^j) \text{ (as this is implied by } \varepsilon_2 \amalg u_2 | (X_y, \varepsilon_1, u_1))$$

in the linear model, Nuc2 holds. As for Nuc1, given X_y, d_1 is determined only by ε_1, whereas y_2^{jk} is determined only by (y_1^j, u_2) and y_1^j is determined only by u_1. Because

$$\varepsilon_1 \amalg (u_1, u_2) | X_y,$$

Nuc1 holds. These two displayed conditions show that we can allow heteroskedasticity and serial correlations within $(\varepsilon_1, \varepsilon_2)$ as well as within (u_1, u_2), but not between $(\varepsilon_1, \varepsilon_2)$ and (u_1, u_2). That is, Nuc1 and Nuc2 hold in the linear model if

$$(\varepsilon_1, \varepsilon_2) \amalg (u_1, u_2) | X_y;$$

the above condition '$\varepsilon_1, \varepsilon_2, x_1, x_2, u_1, u_2$ are iid $N(0,1)$' was overly sufficient. The last display is at the heart of the selection on observables: the selection equation (i.e., the treatment equation) error terms should be unrelated to the outcome equation errors conditional on the covariates.

In the above linear model, y_1 affects d_2 and y_2. Other than d_2 and y_2, x_2 is also a period-2 variable: what happens if y_1 affects x_2? For instance, if the period 1 test score is poor (good), the parents may take some disciplinary (rewarding) measure, such as grounding the child (taking the child out to a ballpark). To examine this possibility, augment the above linear model with an x_2 equation—x_2 is no longer $N(0,1)$ independently of all the other random variables:

$$x_2 = \theta_1 + \theta_2 y_1 + \xi_2 \ (\Longrightarrow x_2^j = \theta_1 + \theta_2 y_1^j + \xi_2), \quad \theta_1, \theta_2 \text{ are parameters and } \xi_2 \sim N(0,1). \quad (2.7)$$

This hardly disturbs Nuc2, because y_2^{jk} still depends only on u_2 given $(X_y, d_1 = j, y_1^j)$, and d_2^j still depends only on ε_2, and what we need for Nuc2 is only

$$\varepsilon_2 \amalg u_2 | (x_1, x_2^j, d_1 = j, y_1^j) \quad \text{(which is implied by } \varepsilon_2 \amalg u_2 | (x_1, \xi_2, \varepsilon_1, u_1)).$$

What goes wrong when y_1 affects x_2 is Nuc1. To see this easily, set $x_1 = \xi_2 = 0$ and all parameters at 1, which implies $x_2 = y_1$. From the d_1 equation in (2.3), d_1 is ε_1 (plus a constant). Substituting the d_1 equation into the y_1 equation in (2.3), y_1 becomes $\varepsilon_1 + u_1$ (plus a constant). From the y_2^{jk} reduced form with $x_2 = y_1 = \varepsilon_1 + u_1$ plugged in, y_2^{jk} is $\varepsilon_1 + 2u_1 + u_2$ (plus a constant). Now, using f to denote densities, observe

$$\begin{aligned} f(y_2^{jk}|d_1, x_2) &= f(\varepsilon_1 + 2u_1 + u_2|\varepsilon_1, \varepsilon_1 + u_1) = f(\varepsilon_1 + 2u_1 + u_2|\varepsilon_1, u_1) \\ &\neq f(\varepsilon_1 + 2u_1 + u_2|\varepsilon_1 + u_1) = f(y_2^{jk}|x_2). \end{aligned}$$

From the first and last terms, we can see that Nuc1 can fail when y_1 affects x_2. The next subsection will provide a simple solution to this failure of G-algorithm.

2.3 G-algorithm with More General Linear Models

Carrying out G-algorithm non-parametrically using a local approximation method such as kernel methods is difficult when the dimension of X_y is large as in our data. But linear models can lead to a much simpler and practical procedure; formally, the linear models may be taken as a nonparametric series approximation. This is explored in this subsection.

Generalizing (2.3) for nonlinear treatment effects and non-stationary parameters by allowing the parameters of the y_1 equation to differ from those of the y_2 equation, suppose (now x_1 and x_2 are multi-dimensional)

$$\begin{aligned} y_1 &= \alpha_1 + \alpha_x' x_1 + \alpha_{d1} d_1 + \alpha_{d1q} d_1^2 + u_1, \quad E(u_1|d_1, X_y) = 0 \ \{\Longrightarrow\ E(u_1|d_1, x_1) = 0\} \\ &\Longrightarrow\ y_1^j = \alpha_1 + \alpha_x' x_1 + \alpha_{d1} j + \alpha_{d1q} j^2 + u_1 \qquad (2.8) \\ y_2 &= \beta_1 + \beta_x' x_2 + \beta_{d1} d_1 + \beta_{d1q} d_1^2 + \beta_{d2} d_2 + \beta_{d2q} d_2^2 + \beta_y y_1 + u_2, \quad E(u_2|d_1, d_2, y_1, X_y) = 0 \\ &\Longrightarrow\ y_2^{jk} = \beta_1 + \beta_x' x_2 + \beta_{d1} j + \beta_{d1q} j^2 + \beta_{d2} k + \beta_{d2q} k^2 + \beta_y y_1^j + u_2 \end{aligned}$$

where the error terms u_1 and u_2 are also more general than in (2.3) because only certain conditional means are specified to be zero. The quadratic terms d_1^2 and d_2^2 are to account for the potential nonlinear effect of TV watching hours: even if TV watching is beneficial, too much TV watching should be harmful. If necessary, various interaction terms can be included in these equations to capture the effect on the subpopulations characterized by the variables interacting with the treatments. Also y_1^2 can be included as well in the y_2 equation. But, for our data with $N \simeq 1800$ and a high dimensional x_{it}, adding high-order terms can quickly go out of hand.

Take $E(\cdot|d_1 = j, X_y)$ on the y_2^{jk} equation in (2.8)—this is the integration step in G-algorithm—to obtain

$$E(y_2^{jk}|d_1 = j, X_y) = \beta_1 + \beta_x' x_2 + \beta_{d1} j + \beta_{d1q} j^2 + \beta_{d2} k + \beta_{d2q} k^2 + \beta_y E(y_1^j|d_1 = j, X_y).$$

Under Nuc1, this becomes

$$E(y_2^{jk}|X_y) = \beta_1 + \beta_x' x_2 + \beta_{d1} j + \beta_{d1q} j^2 + \beta_{d2} k + \beta_{d2q} k^2 + \beta_y E(y_1^j|X_y).$$

The next section will show that typical econometric panel dynamic models will miss the effect conveyed by y_1, i.e., the part due to $E(y_1^j|d_1 = j, X_y)$, which is the key component in dynamic treatment effects with feedbacks.

Substitute $E(y_1^j|X_y) = \alpha_1 + \alpha_x' x_1 + \alpha_{d1} j + \alpha_{d1q} j^2$ from the y_1^j equation into $E(y_1^j|X_y)$ to get

$$E(y_2^{jk}|X_y) = \beta_1 + \beta_x' x_2 + \beta_{d1} j + \beta_{d1q} j^2 + \beta_{d2} k + \beta_{d2q} k^2 + \beta_y(\alpha_1 + \alpha_x' x_1 + \alpha_{d1} j + \alpha_{d1q} j^2).$$

From this with $j = j_o, 0$ and $k = k_o, 0$, we obtain

$$E(y_2^{j_o k_o} - y_2^{00}|X_y) = \beta_{d1} j_o + \beta_{d1q} j_o^2 + \beta_{d2} k_o + \beta_{d2q} k_o^2 + \beta_y(\alpha_{d1} j_o + \alpha_{d1q} j_o^2) = E(y_2^{j_o k_o} - y_2^{00}). \quad (2.9)$$

Turning to estimation, all α and β parameters can be estimated by LSE to the y_1 and y_2 equations in (2.8) separately. But if x_2 is affected by y_1 as in (2.7), then Nuc1 (thus G-algorithm) can fail as noted already. To see this in (2.8), recall the linear model (2.3) with the true effect $\beta_{dlag} j_o + \beta_d k_o + \beta_y \beta_d j_o$ in (2.5), which is a special case of (2.9) with the quadratic terms removed and $\beta_{d1} = \beta_{dlag}$, $\beta_{d2} = \beta_d$, and $\alpha_{d1} = \beta_d$. Although we can still assume $E(u_1|d_1, x_1) = 0$, $E(u_1|d_1, X_y) = 0$ in (2.8) no longer holds. In the LSE of y_1 on $(1, d_1, X_y')$, x_2 in X_y is an endogenous regressor. But there is a simple solution to this endogeneity problem: drop x_2 in the LSE for the y_1 equation in (2.8); this is a *simple modification of G-algorithm when it fails due to x_2 affected by y_1*. In fact, even when x_2 is not endogenous, there is no need to include x_2 in the LSE for the y_1 equation, because the slope estimator for x_2 is consistent for zero. That is, endogenous or not, drop x_2 from the LSE to the y_1 equation in (2.8).

Although we adopted G-algorithm and proposed its modification using linear models, other approaches are certainly possible for dynamic treatment effects. For instance, Lechner (2004,2007) propose matching approaches using the 'propensity score' idea in Rosenbaum and Rubin (1983), which is, however, applicable only when the treatment is binary. The following subsection presents a simplified version of the G-algorithm when the response is binary.

2.4 G-algorithm with Binary Responses

We proposed a series-approximation-based linear model approach above to alleviate the implementation problem of the G-algorithm. If many high-order terms should appear (particularly y_1^2,y_1^3,... in the y_2 equation), however, then even the linear model approach becomes cumbersome. The G-algorithm can be implemented with much ease if y is binary—no need for y_1^2,y_1^3,...any more—in which case the G-algorithm becomes

$$\begin{aligned} E(y_2^{jk}|y_0, X_2) &= P(y_2 = 1|d_1 = j, d_2 = k, y_1 = 0, y_0, X_2) \cdot P(y_1 = 0|d_1 = j, y_0, X_2) \\ +P(y_2 &= 1|d_1 = j, d_2 = k, y_1 = 1, y_0, X_2) \cdot P(y_1 = 1|d_1 = j, y_0, X_2). \end{aligned} \quad (2.10)$$

For instance, apply probit (or logit) to y_2 on d_1, d_2, y_1, y_0, X_2 to obtain the two probit probabilities for $y_2 = 1$ in (2.10):

$$\Phi(\psi_1 + \psi_{d1}d_1 + \psi_{d2}d_2 + \psi_{y1}y_1 + \psi_{y0}y_0 + \psi_x' X_2)$$

where the ψ-parameters are to be estimated. Also apply probit (or logit) to y_1 on d_1, y_0, X_2 to get the probit probabilities for $y_1 = 1$ (and $y_1 = 0$):

$$\Phi(\eta_1 + \eta_{d1}d_1 + \eta_{y0}y_0 + \eta_x' X_2)$$

where the η-parameters are to be estimated. Substituting these into (2.10) will do. This version will be applied to our data as well.

When x_2 is affected by y_1, (2.10) may run into a problem as the preceding linear model does because Nuc1 can fail. When this happens, the left-hand side of (2.10) becomes $E(y_2^{jk}|d_1 = j, y_0, X_2)$ because Nuc1 is no longer available to get rid of $d_1 = j_o$ in the conditioning set. In this case, however, we may still obtain the effect on the treated $d_1 = j_o$ by getting (2.10) for (j_o, k_o) and $(j_o, 0)$ to obtain the difference

$$E(y_2^{j_o k_o}|d_1 = j_o, y_0, X_2) - E(y_2^{j_o 0}|d_1 = j_o, y_0, X_2).$$

Integrate out (y_0, X_2) conditional on $d_1 = j_o$ to get the effect on the treated $d_1 = j_o$:

$$E(y_2^{j_o k_o} - y_2^{j_o 0}|d_1 = j_o) = E[\ \{E(y_2^{j_o k_o}|d_1 = j_o, y_0, X_2) - E(y_2^{j_o 0}|d_1 = j_o, y_0, X_2)\}\ |d_1 = j_o].$$

If we still desire $E(y_2^{j_o k_o} - y_2^{00})$, then following the lead of the linear model case where x_2 being affected by y_1 matters only for the y_1 equation estimation, we may drop x_2 for the y_1 probit estimation to estimate $P(y_1 = 0|d_1 = j, y_0, x_0, x_1)$ instead of $P(y_1 = 0|d_1 = j, y_0, X_2)$. Although we could not work out a formal proof that using this instead of $P(y_1 = 0|d_1 = j, y_0, X_2)$ in (2.10) indeed solves the problem, our experience with (2.10) worked well with this modification. In fact, (2.10) worked just as well even without this modification, because although using x_2 affected by y_1 causes biases in the probit estimator per se, the resulting predicted probabilities needed for (2.10) tend to be hardly biased. In our empirical part later, the results for (2.10) without x_2 removed in the y_1 equation will be presented, which are virtually identical to the results for (2.10) with x_2 removed.

3 Dynamic Panel Data Models

Consider a dynamic panel data model

$$y_2 = \beta_1 + \beta_{d1} d_1 + \beta_{d2} d_2 + \beta_{y0} y_0 + \beta_{y1} y_1 + \beta'_{x1} x_1 + \beta'_{x2} x_2 + v_2 \tag{3.1}$$

where v_2 is an error term. This kind of models are popular in econometrics for a number of reasons, which are laid out in the following. Then we will show the pros and cons of G-algorithm relative to typical panel data approaches using (3.1) or special cases of (3.1).

First, model (3.1) allows for *Granger-causality* test. A panel data version of the Granger-causality test (Granger, 1969,1980) of $\{d_t\}$ on $\{y_t\}$ includes all lagged d_{it}'s and y_{it}'s (as well as the current d_{it} if d_{it} precedes y_{it}) on the right-hand side of the y_{it} equation to test for the coefficients of the lagged d_{it}'s (and the current d_{it} if d_{it} precedes y_{it}) being all zero. For the test, confounding covariates are also controlled by including them on the right-hand side. Hence model (3.1) can be used for Granger-causality. Although we adopt the 'counter-factual causality' framework as explained in the preceding section, dissenting views are also strong as can be seen in Holland (1986) and Dawid (2000). Because Granger-causality is widely used in time-series econometrics with its panel version in Holtz-Eakin et al. (1988,1989), it seems sensible to consider a model that can test for Granger-causality. Although Granger-causality does not imply nor is implied by the counter-factual causality in general, Robins et al. (1999) show that the two concepts do agree in some cases; see also Lechner (2006) further on the comparison of the counter-factual and Granger-type causalities.

Second, model (3.1) allows violations of NUC—*endogeneity problem of treatments—due to the relation between the treatment and 'unobserved unit-specific effect* δ_i*'* when v_{i2} is augmented by δ_i so that the the error term in (3.1) becomes $\delta + v_2$. For this, consider a dynamic panel data model

$$y_{it} = \alpha_t + \alpha_{yt} y_{i,t-1} + \alpha_{dt} d_{it} + \alpha'_{xt} x_{it} + \alpha_{\delta t} \delta_i + v_{it}, \quad t = 1, 2 \tag{3.2}$$

where α's are parameters and δ_i is a time-constant error possibly related with d_{it} (and some regressors). Model (3.2) is more general than typical dynamic panel data models in use because all parameters are indexed by t to allow for the data generating process to be non-stationary in early childhood. Particularly notable is $\alpha_{\delta t}$ for δ_i: the effect of δ_i on y_{it} can vary across t. Take the 'quasi-difference' $y_2 - (\alpha_{\delta 2}/\alpha_{\delta 1}) y_1$ in (3.2) to get rid of δ_i and then put $(\alpha_{\delta 2}/\alpha_{\delta 1}) y_1$ on the right-hand side to get

$$\begin{aligned} y_2 \quad = \quad & \alpha_2 - \frac{\alpha_{\delta 2}}{\alpha_{\delta 1}} \alpha_1 + (\alpha_{y2} + \frac{\alpha_{\delta 2}}{\alpha_{\delta 1}}) y_1 - \frac{\alpha_{\delta 2}}{\alpha_{\delta 1}} \alpha_{y1} y_0 \\ & + \alpha_{d2} d_2 - \frac{\alpha_{\delta 2}}{\alpha_{\delta 1}} \alpha_{d1} d_1 + \alpha'_{x2} x_2 - \frac{\alpha_{\delta 2}}{\alpha_{\delta 1}} \alpha'_{x1} x_1 + v_2 - \frac{\alpha_{\delta 2}}{\alpha_{\delta 1}} v_1. \end{aligned} \tag{3.3}$$

Model (3.1) includes this model devoid of δ_i as a special case. That is, using (3.1), we may not have to be concerned about the endogeneity problem due to δ as we may have to in using (3.2).

Third, model (3.1) allows for violations of NUC *due to the relation between the treatment and the time-variant error* v_{it}. Although the endogeneity due to δ_i is taken care of with (quasi-) differencing, the endogeneity problem due to the relation with v_{it} in (3.2) requires instrumental variable estimator (IVE). Angrist and Krueger (2001) show an ingenious list of instruments in various studies, but having that type of instruments is not always possible. Rather, in typical panel data, it is unavoidable that one finds instruments within the data—namely, lagged regressors. Because (3.3) derived from (3.2) includes already y_1, y_0, d_2, d_1, x_2, and x_1, the only source left for instruments for endogenous regressors is x_0. Hence, when x_0 is excluded from the model as in (3.1), IVE can be done for endogenous treatments; IVE can be done also for endogenous covariates.

To see what type of orthogonality conditions are invoked when x_0 is used as instruments for (3.3), observe that

$$COR(v_2 - \frac{\alpha_{\delta 2}}{\alpha_{\delta 1}} v_1,\ x_0) = 0 \text{ is implied by } COR(v_2, x_0) = COR(v_1, x_0) = 0.$$

This is a '*predeterminedness*' type of assumption (see, e.g., Lee (2002) for several types of orthogonality conditions in panel data IVE). This assumption allows for a simultaneous relation between (d_t, x_t') and v_t.

In practice, the so-called 'fixed-effect estimator' for a simple stationary model is popular, which is equivalent to first-differencing the non-dynamic version of (3.2) with all parameters time-constant. For our empirical analysis, we will also apply first-differencing to a slightly more general model: defining $\Delta y_{i2} \equiv y_{i2} - y_{i1}$, $\Delta d_{i2} \equiv d_{i2} - d_{i1}$, $\Delta d_{i2}^2 \equiv d_{i2}^2 - d_{i1}^2$, and $\Delta v_{i2} \equiv v_{i2} - v_{i1}$,

$$\begin{aligned} y_{it} &= \alpha_t + \alpha_d d_{it} + \alpha_{dq} d_{it}^2 + \alpha_{xt}' x_{it} + \alpha_\delta \delta_i + v_{it} \\ &\Longrightarrow \quad \Delta y_{i2} = \alpha_2 - \alpha_1 + \alpha_d \Delta d_{i2} + \alpha_{dq} \Delta d_{it}^2 + \alpha_{x2}' x_{i2} - \alpha_{x1}' x_{i1} + \Delta v_{i2}. \end{aligned} \tag{3.4}$$

In short, as a comparison to G-algorithm, we will apply three panel data approaches: (i) LSE to (3.1) for Granger-causality, (ii) IVE to (3.1) to allow for endogeneity of the treatment or covariates, and (iii) LSE to (3.4) which is a 'fixed-effect' estimator.

Turning to the comparison of G-algorithm and the panel data approaches, since we already pointed out the main advantage of the dynamic panel data model relative to G-algorithm—namely, relaxations of NUC—here we show only the main disadvantage of the dynamic panel data approach. Recall the figure showing the feedback feature where d_2 has only a direct effect on y_2 while d_1 has both direct and indirect effects (through y_1) on y_2. Because y_1 *is controlled in the dynamic panel data model, the indirect effect of* d_1 *on* y_2 *is not identified.* Formally, the treatment effect that (3.1) can deliver is not $E(y_2^{jk} - y_2^{00})$ but

$$\beta_{d1} j + \beta_{d2} k = E(y_2^{jk} - y_2^{00} | y_1) \tag{3.5}$$

If we do not control y_1 to avoid this problem, then the effect of d_2 on y_2 can be distorted because y_1 becomes a 'common factor' for d_2 and y_2. That is, even if there is no true effect of d_2 on y_2, we may find a spurious effect of d_2 due to not controlling y_1. *This dilemma is fundamental to dynamic treatment effect analysis under feedback.* Ruling out the feedback from y_1 to d_2 a priori would be ill-advised, because when it comes to TV watching, no parents would sit idle after having seen low test scores of their children.

In summary, model (3.1) has a couple of nice features. First, we can test for Granger non-causality of $\{d_t\}$ on $\{y_t\}$. Second, using IVE allows endogeneity (i.e., violation of NUC) of d_{it} and covariates from two major sources: the relations to δ and v_t. Despite these nice features, however, model (3.1) has the critical weakness resulting from missing the 'feedback' from the interim response y_1 to d_2. In the "best of times"—that is, if the NUC holds—G-algorithm identifies the full dynamic effects. But the panel data approach with (3.1) will always miss part of the dynamic effect, even when all the assumptions for it hold.

4 Data Description

The NLSY79 child sample contains rich information on children born to the women respondents of the NLSY79. Starting from 1986, a separate set of questionnaires was developed to collect information about the cognitive, social, and behavioral development of children. The set of child development results and inputs from birth up to age 10 were grouped in three periods: 0-2 years, 3-5 years, and 6-9 years. The variables include detailed home inputs as well as family backgrounds and school inputs.

Based on children surveyed from 1986 to 1998, we constructed a longitudinal sample of about 2600 children with no missing values in Peabody Individual Achievement Test (PIAT) math and reading scores at ages 8-9 and TV watching hours at ages 6-9. The relevant questions on TV watching ask a mother how many hours her child watches TV on a typical weekday and weekend day. While examining the data, we found that some answers do not seem valid: the reported hours sometimes go well beyond 24 hours. This may be due to a confusion between a daily measure and a weekly measure of TV watching hours. Thus we excluded children reportedly spending more than ten hours watching TV on a typical day at any age. This left us with a sample of 2180 children, based on which all of our empirical analyses were conducted. The summary statistics of some variables in this sample are listed in Table 1.

For children five years old and above, PIAT Math score offers a wide-range measure of achievement in mathematics, and PIAT Reading Recognition score assesses their attained reading knowledge. Both are among the most widely used assessments of academic achievements. The norming sample has mean 100 and standard deviation (SD) 15 for both math and reading scores; these were normed against the standards based on a national sample of children in the United States in 1968. The PIAT math in our sample has mean 102.2 and SD 13.9 around age nine, and the PIAT reading score has mean 105 and SD 14.6.

In the sample, the average child spends 3.5 hours per day watching TV at ages 8-9 and

3.2 hours at ages 6-7. Specifically, around 60% of children aged 8-9 watch more than 2 hours TV per weekday, and 21% of them more than 4 hours; on weekends, they watch TV for longer hours: around 72% exceeding 2 hours and 35% exceeding 4 hours. These patterns are quite similar to those at ages 6-7, though younger children usually watch less TV. We choose to use a measure of daily TV viewing hours in the form of

$$\frac{1}{7}\{5 \times (\text{average weekday watching hours}) + 2 \times (\text{average weekend-day watching hours})\}.$$

White children on average watch about one hour less TV per day at both ages 6-7 and 8-9 than the others, and their PIAT scores are 8.8 points higher in math and 6.4 points higher in reading. There is virtually no difference between boys and girls in TV watching hours and PIAT scores, while firstborns watch about half an hour less than the others and get higher math and reading scores. A child with ten or more child-books at home watches around 1.2 hours less TV at age 8-9 and have much higher scores (11.2 points higher in math and 12.1 in reading) than those with fewer books. Similarly, children whose mothers read to them frequently (at least three times per week) and whose parents discuss TV programs with them spend less time in watching TV than others and get higher grades. In general, TV watching times are significantly and negatively correlated with other activities such as going to museums and theaters.

Public school children watch about one hour more TV on a daily basis at ages 6-9 than those in private schools, and their math and reading scores are lower. The perceived quality of detailed school inputs (including the skills of the principal and teachers, how much teachers care about the students, whether parents are given enough information and opportunity to participate in school affairs, the safety and order of the school, and whether moral teaching is offered), however, do not seem to affect much TV watching time, although they are positively associated with both math and reading scores. The correlation between TV watching hours and time spent on math homework or reading and writing assignments is quite weak and sometimes positive.

Children with mothers having 16 or more years of schooling watch about one hour less TV. Similarly, children whose mothers have above average AFQT scores watch 1.4 hours less TV at ages 8-9 and around 1.2 hours less at ages 6-7, and their PIAT scores are much higher (10 points higher in math and 9 in reading). In summary, children with high quality home inputs and better educated mothers watch much less TV, while school inputs have relatively less influence.

A salient feature in Table 1 is that less time spent viewing TV is almost always associated with higher math and reading scores. The potentially harmful effects of TV watching, however, may be over-estimated if detailed home inputs are not controlled, since a child watching more TV also lacks important home inputs. The strength of our data is that a rich set of home inputs from birth up to age nine as well as key family background variables are available; for some children, there are also many detailed school inputs available. This would greatly reduce potential biases due to omitted variables. Most home and school input variables were categorical with multiple levels, which were converted to dummy variables according to sample medians. The age-specific Home Observation Measurement of the Environment variable (HOME), which is a simple summation of the dichotomized individual input item scores, is often used in child development research as an aggregate quality indicator of home environment. The completion rates of HOME, however, are in general very low for children under age four, which causes many missing values. Whenever possible, HOME is included as a control in addition to the detailed home inputs.

In discussing dynamic treatment effects, we desire the time sequence $x_{it} \longrightarrow d_{it} \longrightarrow y_{it}$ in each period so that x_{it} works as the time-t baseline covariates which affect d_{it} and possibly y_{it}, and then the treatment d_{it} affects y_{it}. In our data, this temporal order is plausible for a couple of reasons. First, y_{it} is measured on the interview day, which means that x_{it} and d_{it} precede y_{it}. Second, many family characteristics of x_{it} are likely to be determined independently of d_{it}. Third, overall, TV watching hours tend to be the "residual' usage of time, and thus is unlikely to influence the other time-consuming activities, although we cannot completely rule out TV watching taking precedence over the other activities. As mentioned above, our panel data has three periods defined by child ages, and we use the two later periods while the first period serves as the baseline period providing x_0 and y_0.

5 Empirical Results

5.1 Granger Causality

Table 2 presents results for Granger causality for model (3.1) augmented by squared treatment variables, where PIAT math and reading scores at ages 8-9 are the dependent variables. The various specifications differ mainly in the control variables used. In the first column 'OLS' of both math and reading regressions, the earlier scores y_1 at ages 6-7 and y_0 at ages 4-5 are used as well as the basic group of controls which include the child's race, sex, birth order, home inputs at ages 6-7 and 8-9, and family backgrounds variables (mother's AFQT score, her age at the child birth, whether the child was breast-fed, her marriage status, her highest grade, and family income). The sample sizes drop much when the school inputs at ages 6-9 are controlled in the second column 'OLS(S)'. Home inputs at age 4-5 are further added in the third column 'OLS(SH)', which contains the most comprehensive set of controls and hence the smallest sample size.

For math scores at age 8-9, TV watching hours at ages 6-7 and 8-9 are jointly significant across these specifications with p values at least 0.07, though the coefficients of individual TV watching variables are usually not significant; TV hours at age 6-7 have positive and concave effects, while those at age 8-9 have negative and convex effects. For reading scores at age 8-9, the joint significant levels of TV watching hours are overall lower than those for math except in the most comprehensive specification 'OLS(SH)', where TV hours at ages 6-7 and 8-9 are jointly significant with p-value 0.009. The effects of TV watching hours at both age periods seem to have positive and concave effects.

To compare later with the other tables, the mean differences in PIAT scores between different TV watching hours at ages 6-9 are listed in the lower half of Table 2. More time spent on watching TV (up to 6 hours daily) is associated with higher math scores, while the TV effects on reading scores seem positive and concave with the best outcome achieved at 3 hours in 'OLS(S)' and 'OLS(SH)'. These results, however, are not significantly different from zero. As emphasized earlier, these mean differences capture only the direct effects of TV watching since the lagged PIAT scores are controlled.

In summary, there exists clear evidence of Granger causality of TV watching at ages 6-9 on PIAT math and reading scores at ages 8-9. The effects of TV watching seem to differ in both magnitude and sign for math versus reading scores. With the Granger-causality established now, the orderly thing to do is proceeding further to obtaining the full dynamic effects of TV watching. But before that, in the next subsection, we will present some estimation results for typical econometric panel data methods.

5.2 Fixed Effect with IV

In Table 3, columns 'FE(IV)' present the results based on model (3.1) where home inputs at age 4-5 are used as instruments for the current home inputs at age 8-9; as in Table 2, the squared treatment variables are used. This specification allows for the endogeneity problem of current home inputs. As noted already, (3.1) may be regarded as (3.3) that was obtained from (3.2) by removing the individual-specific effect with quasi-differencing. This proposition is supported by the different signs of d_2 and d_1 (and d_2^2 and d_1^2) in Tables 2 and 3—in (3.3), the signs of d_2 and d_1 are likely to differ as $\alpha_{\delta 2}/\alpha_{\delta 1} > 0$ is plausible.

Consistent with the OLS specification for Table 2, TV hours at age 6-7 have positive and concave effects, while those at ages 8-9 have negative and convex effects for math score at ages 8-9; the coefficients of individual TV watching variables, however, are insignificant due to the lower precision in estimation. But, in contrast to the OLS results in Table 2, spending more time watching TV at ages 6-9 is overall associated with lower math scores in 'FE(IV)', as shown at the lower half of the table in the same column. This is due to the much larger negative marginal effect of current TV hours (-7.90 in 'FE(IV)' versus -2.68 in 'OLS(SH)' in Table 2) when they are instrumented by TV hours and other home inputs at age 4-5. Similar results apply to the reading scores, where the different coefficients and effects of TV hours at age 8-9 do not amount to significant differences from the OLS specification, judging from the Hausman test.

Columns 'DF' and 'DF (S)' are based on the first-difference model (3.4); they differ only in terms of control variables, where school inputs at ages 6-9 are further controlled in 'DF (S)'. For PIAT math scores at ages 8-9, TV watching hours at ages 8-9 has a negative and convex effect on math scores under both specifications, which leads to an overall negative association between TV hours and math scores as shown in the lower half of the table. For reading scores at ages 8-9, TV watching hours at ages 8-9 has a positive and concave effect under both specifications, where 3-hour TV watching achieves the best outcome in 'DF' while 4-hour is best when school inputs are controlled under 'DF (S)'. These estimated TV effects, however, are statistically insignificant for both math and reading scores. There is a notable, but statistically insignificant, difference between the negative TV effects for math score in the first-difference results and the positive effects under various OLS specifications in Table 2 for math scores, suggesting that the unobserved child-specific factor may matter more for math scores than for reading scores.

The overall results of these fixed-effect and IV models are not significantly different from the OLS specifications for Table 2, possibly because many detailed home inputs and family backgrounds are already controlled. This finding suggests that we may proceed to G-estimation without worrying too much about the potential biases caused by the unobserved child-specific fixed effect or the endogeneity of current inputs. It is important to note that, even when these OLS and IVE are consistent, they can capture only the direct effects, not the total effects of TV watching.

5.3 G-algorithm with Linear Models

Table 4 presents results for G-estimation based on the linear models in (2.8), where PIAT math and reading scores at ages 6-7 and 8-9 are the dependent variables, respectively, for the y_1 and y_2 equations. The two specifications 'OLS' and 'OLS(S)' differ only in their control variables, where the current school inputs are included in 'OLS(S)' for both age periods. Under both specifications, the effects of TV watching at ages 6-7 on scores at ages 6-7 are negative and convex for math score and just negative for reading score. But, as shown in the lower half of Table 4, the overall effects of TV watching across ages 6-9 on math and reading scores at ages 8-9 in 'OLS(S)' is negative for math scores and concave for reading scores, where the best outcome is achieved at 2 hours in column 'G-est.(S)' (and at 6 hours in column 'G-est.'). Overall, however, the magnitudes of most TV effects are quite small and statistically insignificant.

To get a sense of the relative importance of the indirect effect of TV watching at age 6-7 on the math and reading scores two years later, we did some calculation based on the estimates in columns 'OLS(S)'. For math score, compared with not watching TV at all, the effect of one-hour TV watching at age 6-7 on the math score at ages 6-7 is $(-.74+.01) = -.73$, which leads to an indirect effect on the math score at ages 8-9 $-.73 * .64 = -.47$, while its direct effect is $(-.06 + .02) = -.04$ points. The effect of one-hour TV viewing at ages 8-9 is $(.20 + .002) = .202$ points. So the indirect effect of earlier TV watching at ages 6-7 (which is $-.47$) is almost three times the total direct effect of TV watching at both ages 6-7 and 8-9 (which is $.162 = .202 - .04$), and makes the total effect of one-hour TV watching at ages 6-9 negative ($-.30$ as reported at the first row in column 'G-est.(S)' for math), despite the positive effect of current TV viewing at ages 8-9.

Doing analogously, based on the estimates in column 'OLS(S)' for reading score, the mean difference between the effects of 6-hour and 2-hour TV viewing at ages 6-9 on reading scores at age 8-9 (which is -1.78 as reported at the last row in column 'G-est.(S)' for reading) can be decomposed into three components: the indirect effect of TV viewing time at ages 6-7 is $[(-.49*6 - .01*36) - (-.49*2 - .01*4)]*.78 = -1.78$, the direct effect of d_1 and d_1^2 is $(-.47*6 + .02*36) - (-.47*2 + .02*4) = -1.24$, while the direct effect of d_2 and d_2^2 is $(1.27*6 - .12*36) - (1.27*2 - .12*4) = 1.24$. Again, the indirect effect of TV viewing at ages 6-7 on the reading score at ages 8-9 dominates the total direct effect of TV viewing at ages 6-9, and outweighs the positive effect of the current TV watching.

The bottom half of Table 4 presents various mean treatment effects. All effects for math scores are negative. For reading scores, the effects are mixed with positive as well as negative effects. In all cases, unfortunately, the effect is either too small in magnitude or statistically insignificant.

5.4 G-algorithm with Discrete Responses

The insignificance of the total TV effects in Table 4 might be due to measurement errors, since few parents can recall with much accuracy the exact time their children watch TV every day. It is quite plausible, however, most parents know the time range their kids spend in front of TV. This prompts us to apply the simplified G-estimation with discrete responses, where we convert the average daily TV watching hours to three dummy variables: *High TV* if a child watches TV for more than 4 hours per day, *Middle TV* if between 2 and 4 hours, and *Low TV* if less than or equal to 2 hours. The three levels are done mainly to capture the possible concavity feature of TV watching effects. We also convert PIAT math and reading scores to dummy variables which take on 1 if higher than the sample mean and 0 otherwise. The basic set of controls includes detailed home inputs at ages 6-9, home environment indicators at ages 2-3 and 4-5, child demographic information and family backgrounds variables, while detailed home inputs at age 4-5 are further added in columns labeled 'Probit(H)' and 'G-est.(H)'.

The probit results are shown in the upper part of Table 5, where the entries are the estimated marginal effects calculated at the sample means of the control variables (i.e., the derivatives of $P(y_2 = 1|\cdots)$ evaluated at the variable sample averages). For math scores at ages 8-9 and 6-7, none of the TV dummies is significant. For reading score at ages 8-9, the coefficient of High TV at ages 6-7 is insignificantly negative in column 'Probit', while that at age 8-9 is significantly positive in 'Probit (H)'. There seems to be some nonlinear effects, but most individual estimates are insignificant.

The G-estimation using the above probit regressions are presented in the lower half of Table 5. Compared to the benchmark of watching TV less than or equal to 2 hours per day at both ages 6-7 and 8-9, the total effects of watching TV for more than 4 hours at both ages are negative and significant for both math and reading scores at ages 8-9 over different specifications. Specifically, a child with High TV at ages 6-9 reduces his/her probability of having a higher-than-average math score by 18% in the first column 'G-est.' and 23% in the second column 'G-est. (H)'; the corresponding probability for reading scores is lowered by 21% and 23% respectively for the two specifications. Since the SD of having a higher-than-average score is around .50, these percentage reductions amount to almost half the SD. Middle level TV hours (between 2 and 4 hours per day) at ages 6-9 also have negative and significant effects on math and reading scores, though their magnitudes are smaller than those of High TV: the probability of having a higher-than-average PIAT score is reduced by 16%-18% for math and 13%-14% for reading.

It seems that TV watching at ages 6-7 has much larger negative effects on both math and reading scores at ages 8-9 than TV watching at ages 8-9. For example, in the last two rows of Table 5, the results in the first column 'G-est.' shows that, watching more than 4 hours TV daily at ages 6-7 reduces the probability of having a higher-than-average math score by 23%, while watching between 2 and 4 hours TV daily reduces it by 19% even when the TV watching time is less than or equal to 2 hours per day at age 8-9; the corresponding numbers (unreported in Table 5) are 18% and 16% respectively when the TV watching hours at ages 8-9 are High and Middle instead. These results imply that watching more TV at ages 8-9 actually does much less harm compared to the effects of TV hours two years earlier. More or less the same statements can be made for reading scores.

6 Conclusions

When multiple treatments are given over time and there is a feedback of interim responses affecting some future treatments, finding the total effect of all treatments on the final response measured at the end is difficult. The fundamental dilemma is that, if the interim responses are controlled as in the usual dynamic models, then the indirect effect from the earlier treatments on the final response through the interim responses is missed; if not controlled, the interim responses become unobserved confounders for the direct effect of the affected treatments on the final response. Despite this difficulty, the G-estimation (or G-algorithm) originally developed in biostatistics can identify the total effect, unlike the usual OLS or IVE applied to dynamic models with lagged responses on the right-hand side.

This paper reviewed G-estimation, and applied two practical versions of G-algorithm to an important issue: the effect of watching TV on child cognitive development measured by math and reading scores. For math score, the G-estimation results indicated that watching TV for more than two hours per day during ages 6-9 has a negative total effect at ages 8-9, where the negative effects of TV watching at younger ages 6-7 are much larger, which was not expected beforehand. Furthermore, results from various estimators (G-algorithm and typical panel data econometric approaches) using continuous response variables led to a coherent evidence that between two and four hours TV watching per day seems to bring the best reading scores than too much or too little TV hours, while the effects of TV watching on math scores are usually negative.

These findings collectively explain why the effect of TV watching on child cognitive development has been controversial: the effect varies depending on the TV watching age, and it is nonlinear with changing signs. Also its magnitude is small, suggesting that TV effect may not matter that much after all. The total effect feature found by our dynamic framework provided a richer "story", taking only part of which would convey misleading messages.

REFERENCES

Anderson, D.R., 2005, Television and very young children, American Behavioral Scientist 48, 505-522.

Angrist, J.D. and A.B. Krueger, 2001, Instrumental variables and the search for identification: from supply and demand to natural experiments, Journal of Economic Perspectives 15, 69-85.

Dawid, A.P., 2000, Causal inference without counterfactuals, Journal of the American Statistical Association 95, 407-424.

Gentzkow, M. and J.M. Shapiro, 2007, Does television rot your brain? New evidence from the Coleman study, Quarterly Journal of Economics, forthcoming.

Granger, C.W.J., 1969, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37, 424-438.

Granger, C.W.J., 1980, Testing for causality: a personal viewpoint, Journal of Economic Dynamics and Control 2, 329-352.

Heckman, J.J., 1999, Policies to foster human capital, NBER Working Paper 7288.

Holland, P.W., 1986, Statistics and causal inference, Journal of the American Statistical Association 81, 945-960.

Holtz-Eakin, D., W. Newey, and H.S. Rosen, 1988, Estimating vector autoregressions with panel data, Econometrica 56, 1371-1395.

Holtz-Eakin, D., W. Newey, and H.S. Rosen, 1989, The revenue-expenditure nexus: evidence from local government data, International Economic Review 30, 415-429.

Juster, F.T., Stafford, F. P., 1991, The allocation of time: empirical findings, behavioral models, and problems of measurement, Journal of Economic Literature 29, 471–522.

Keane, M, and K. Wolpin, 1997, Career decisions of young men, Journal of Political Economy 105, 473-522.

Lechner, M., 2004, Sequential matching estimation of dynamic causal models, University of St. Gallen, Discussion paper.

Lechner, M., 2006, The relation of different concepts of causality in econometrics, University of St. Gallen, Discussion paper

Lechner, M., 2007, Matching estimating of dynamic treatment models: some practical issues, in Modelling and evaluating treatment effects in econometrics, edited by D. Millimet, J. Smith, and E. Vytlacil, Advances in Econometrics 21, forthcoming, JAI Press.

Lechner, M. and R. Miquel, 2001, A potential outcome approach to dynamic programme evaluation – Part I: identification, Discussion paper 2001-07, Department of Economics, University of St. Gallen.

Lee, M.J., 2002, Panel data econometrics: methods-of-moments and limited dependent variables, Academic Press.

Robins, J. M., 1986. A new approach to causal inference in mortality studies with sustained exposure periods - Application to control of the healthy worker survivor effect, Mathematical Modelling 7, 1393–1512; Errata, 1987, Computers and Mathematics with Applications 14, 917–921; Addendum, 1987, Computers and Mathematics with Applications 14, 923-945; Errata to Addendum, 1987, Computers and Mathematics with Applications 18, 477.

Robins, J.M., 1998, Structural nested failure time models, in Survival Analysis, Vol. 6, Encyclopedia of Biostatistics, edited by P. Armitage and T. Colton, Wiley.

Robins, J.M., 1999, Marginal structural models versus structural nested models as tools for causal inference, in Statistical models in epidemiology: the environment and clinical trials, edited by M.E. Halloran and D. Berry, Springer, 95-134.

Robins, J.M., S. Greenland, and F.C. Hu, 1999, Estimation of the causal effect of a time-varying exposure on the marginal mean of a repeated binary outcome, Journal of the American Statistical Association 94, 687-700.

Rosenbaum, P.R. and D.B. Rubin, 1983, The central role of the propensity score in observational studies for causal effects, Biometrika 70, 41-55.

Van Evra, J.P., 2004, Television and child development, Lawrence Erlbaum Associates.

Zavodny, M., 2006, Does watching television rot your mind? Estimates of the effect on test scores, Economics of Education Review 25, 565-573.

Zimmerman , F.J. and D.A. Christakis, 2005, Children's television viewing and cognitive outcomes: A longitudinal analysis of national data, Archives of Pediatrics and Adolescent Medicine 159, 619-625.

Table 1: TV Watching and Math and Reading Scores: Summary Statistics

	Daily TV watching hours		PIAT scores at age 8-9		Group Size
	at age 8-9	at age 6-7	math	reading	
Main Sample	3.5 (2.0)	3.3 (2.0)	102.2 (13.9)	105 (14.6)	2180
Race					
White	2.9 (1.7)	2.9 (1.8)	106.3 (13.0)	108.0 (13.8)	1165
Non-White	4.1 (2.2)	3.9 (2.2)	97.5 (13.4)	101.6 (14.7)	1015
Sex					
Boy	3.5 (2.0)	3.3 (1.9)	102.6 (14.6)	103.6 (15.3)	1095
Girl	3.5 (2.0)	3.4 (2.1)	101.7 (13.2)	106.4 (13.7)	1085
Birth order					
First-borns	3.2 (1.9)	3.1 (1.9)	104.3 (13.6)	108.4 (13.3)	868
Others	3.7 (2.1)	3.5 (2.1)	100.7 (13.9)	102.7 (15.0)	1312
How many children books a child has at home at age 6-7					
> = 10 at both	3.3 (1.9)	3.2 (2.0)	103.8 (13.2)	106.7 (13.7)	1862
< 10 at either	4.6 (2.2)	4.1 (2.2)	92.6 (13.2)	94.6 (15.4)	313
Mother reads to child at age 6-7					
Often	3.3 (2.0)	3.3 (2.0)	103.1 (13.5)	105.7 (14.2)	1636
Not often	4.1 (2.1)	3.6 (2.0)	99.2 (14.6)	102.7 (15.5)	540
Whether parents discuss TV programs with a child					
Discuss	3.4 (2.0)	3.3 (2.0)	103.2 (13.8)	106.0 (14.4)	1801
Not discuss	3.9 (2.2)	3.6 (1.9)	97.0 (13.2)	100.2 (14.7)	354
School type					
Public	3.5 (2.0)	3.4 (2.0)	102.6 (14.4)	104.8 (14.7)	885
private	2.6 (2.0)	2.3 (1.6)	105.7 (11.6)	111.1 (11.5)	66
Mother's highest grade					
> =16	2.7 (1.8)	2.7 (1.8)	107.0 (13.6)	109.4 (13.4)	534
< 16	3.7 (2.0)	3.6 (2.0)	100.6 (13.6)	103.5 (14.7)	1646
Mother's AFQT score in 1981					
Above mean	2.7 (1.6)	2.7 (1.7)	107.5 (12.5)	109.9 (12.9)	999
Below mean	4.1 (2.1)	3.9 (2.1)	97.6 (13.2)	100.9 (14.7)	1181

Note: The entries are group means and standard deviations (in parentheses). The main sample is composed of kids watching ten hours or less TV per day between ages 6-9.

Table 2: TV Watching on Math and Reading: Granger Causality

	PIAT Math at age 8-9			PIAT Reading at age 8-9		
	OLS	OLS(S)	OLS(SH)	OLS	OLS(S)	OLS(SH)
Daily TV hours at age 8-9	-1.05	-2.15	-2.68	-.06	2.21	2.49
	(.77)	(1.74)	(2.52)	(.61)	(1.83)	(2.60)
Daily TV hours at age 8-9 squared	.16**	.36**	.40	.03	-.23	-.25
	(.08)	(.18)	(.26)	(.07)	(.18)	(.25)
Daily TV hours at age 6-7	.89	4.79**	5.50**	.58	1.96	2.54
	(.68)	(1.89)	(2.31)	(.60)	(1.79)	(2.0)
Daily TV hours at age 6-7 squared	-.11	-.56***	-.66**	-.07	-.37*	-.57**
	(.08)	(.21)	(.26)	(.07)	(.21)	(.23)
PIAT Score at Age 6-7	.59***	.58***	.65***	.78***	.83***	.88***
	(.04)	(.12)	(.13)	(.03)	(.09)	(.12)
PIAT Score at Age 4-5	.10***	.13	.12	.09***	.06	.09
	(.03)	(.09)	(.11)	(.03)	(.08)	(.09)
Joint Significance of TV at 6-9	.07*	.03**	.04**	.60	.12	.009***
Joint Significance of TV at 8-9	.02**	.02**	.15	.50	.46	.61
Joint Significance of TV at 6-7	.38	.03**	.048**	.60	.04**	.001***
Sample Size	871	178	155	835	175	152
R-squared	.49	.71	.72	.64	.75	.79

Mean Difference in PIAT Scores (Bootstrapped SD in parentheses)

	PIAT Math at age 8-9			PIAT Reading at age 8-9		
Daily TV hours at age 6-9	OLS	OLS(S)	OLS(SH)	OLS	OLS(S)	OLS(SH)
1 hour versus 0 hour	-.12	2.43	2.56	.49	3.57	4.20
	(.72)	(2.68)	(5.82)	(.61)	(2 .44)	(4.23)
2 hours versus 0 hour	-.13	4.46	4.62	.90	5.94	6.77
	(1.27)	(4.73)	(9.95)	(1.07)	(4.30)	(7.34)
3 hours versus 0 hour	-.04	6.08	6.17	1.25	7.11	7.69
	(1.66)	(6.21)	(12.5)	(1.39)	(5.59)	(9.41)
4 hours versus 0 hour	.15	7.28	7.21	1.52	7.07	6.97
	(1.9)	(7.16)	(13.5)	(1.59)	(6.38)	(10.57)
5 hours versus 0 hour	.45	8.08	7.73	1.72	5.83	4.61
	(2.0)	(7.68)	(13.5)	(1.67)	(6.77)	(11.06)
6 hours versus 0 hour	.85	8.47	7.75	1.86	3.38	.60
	(2.0)	(7.96)	(13.1)	(1.67)	(6.95)	(11.35)
4 hours versus 2 hours	.28	2.82	2.58	.62	1.13	.20
	(.67)	(2.65)	(4.37)	(.56)	(2.32)	(3.78)
6 hours versus 2 hours	.98	4.01	3.13	.95	-2.56	-6.16
	(1.0)	(4.53)	(8.73)	(.88)	(4.04)	(7.26)

Notes: * p<.1; ** p<.05; *** p<.01. Standard deviations are in the parentheses. The sample is composed of kids watching ten hours or less TV per day at ages 6-9. The controlled inputs include a child's race, sex, birth order, home inputs at ages 6-9, and family backgrounds (mother's AFQT score, her age at child birth, whether the child was breastfed, her marriage status, highest grade, and family income). OLS(S) -- School inputs (hours a child spent after school working on math problems and writing projects, whether the child participated remedial programs in math and reading, whether he/she participated in programs for advanced work; the school is public or private, how much teachers care about the students, the skills of teachers and the principal, whether parents are given enough information and opportunity to participate in school affairs, the safety and order of the school, and the moral teaching offered in the school) are included. OLS(SH) -- Home inputs at age 4-5 are included as well as the above school inputs.

Table 3: TV Watching on Math and Reading: Fixed Effect Model

	PIAT Math			PIAT Reading		
	At age 8-9	Difference b/w ages 8-9 and 6-7		At age 8-9	Difference b/w ages 8-9 and 6-7	
	FE (IV)	DF	DF(S)	FE (IV)	DF	DF(S)
Difference in daily TV hours b/w ages 8-9 and 6-7		-.18 (.50)	-1.41 (1.26)		.084 (.39)	1.38 (.90)
Difference in squared daily TV hours b/w ages 8-9 and 6-7		.02 (.05)	.14 (.13)		-.015 (.04)	-.17* (.09)
Daily TV hours at age 8-9	-7.90 (11.3)			-1.89 (8.28)		
Daily TV hours at age 8-9 squared	.61 (1.68)			.14 (1.15)		
Daily TV hours at age 6-7	5.76 (5.41)			.92 (3.68)		
Daily TV hours at age 6-7 squared	-.55 (.46)			-.08 (.34)		
PIAT Score at age 6-7	.62*** (.12)			.72*** (.10)		
Sample Size	1050	1555	299	1043	1539	299
R-squared	--	.04	.22	--	.06	.27

Mean Difference in PIAT Scores (Bootstrapped SD in parentheses)

	PIAT Math at age 8-9			PIAT Reading at age 8-9		
Daily TV hours at age 6-9	FE (IV)	DF	DF(S)	FE (IV)	DF	DF(S)
1 hour versus 0 hour	-2.08 (13.4)	-.16 (.47)	-1.27 (1.22)	-.91 (24.1)	.07 (.36)	1.21 (.98)
2 hours versus 0 hour	-4.03 (22.8)	-.28 (.84)	-2.26 (2.19)	-1.70 (43.2)	.11 (.64)	2.10 (1.77)
3 hours versus 0 hour	-5.86 (28.1)	-.35 (1.11)	-2.96 (2.90)	-2.37 (57.3)	.12 (.85)	2.65 (2.37)
4 hours versus 0 hour	-7.57 (29.7)	-.38 (1.29)	-3.38 (3.36)	-2.91 (66.8)	.10 (1.0)	2.88 (2.80)
5 hours versus 0 hour	-9.15 (28.1)	-.37 (1.39)	-3.51 (3.61)	-3.33 (71.8)	.05 (1.08)	2. 77 (3.07)
6 hours versus 0 hour	-10.6 (25.2)	-.31 (1.41)	-3.37 (3.68)	-3.63 (72.8)	-.02 (1.12)	2.34 (3.22)
4 hours versus 2 hours	-3.53 (8.41)	-.10 (.47)	-1.12 (1.31)	-1.21 (24.3)	-.13 (.57)	.24 (1.65)
6 hours versus 2 hours	-6.57 (16.8)	-.03 (.65)	-1.11 (1.92)	-1.93 (33.5)	-.01 (.36)	.78 (1.04)

Notes: * p<.1; ** p<.05; *** p<.01. Standard deviations are in the parentheses. The sample is composed of kids watching ten hours or less TV per day between ages 6-9. The controlled inputs include a child's race, sex, birth order, home inputs at ages 6-9, and family backgrounds. FE(IV) – Current home inputs at ages 8-9 are instrumented by earlier home inputs at ages 4-5. DF(S) – School inputs at ages 6-9 are included.

Table 4: TV Watching on Math and Reading: G-Estimation with Linear Models

	PIAT Math				PIAT Reading			
	OLS		OLS (S)		OLS		OLS (S)	
	at 8-9	at 6-7	at 8-9	at 6-7	at 8-9	at 6-7	at 8-9	at 6-7
Daily TV hours at age 8-9	-.42		.20		.37		1.27*	
	(.54)		(.77)		(.48)		(.68)	
Daily TV hours at age 8-9 squared	.06		.002		-.04		-.12*	
	(.06)		(.08)		(.05)		(.07)	
Daily TV hours at age 6-7	.13	-.27	-.06	-.74	.18	-.11	-.47	-.49
	(.50)	(.44)	(.76)	(1.04)	(.44)	(.43)	(.67)	(.97)
Daily TV hours at age 6-7 squared	-.01	.02	.02	.01	-.01	-.01	.02	-.01
	(.05)	(.05)	(.08)	(.12)	(.05)	(.05)	(.08)	(.10)
PIAT score at age 6-7	.60***		.64***		.75***		.78***	
	(.03)		(.04)		(.03)		(.04)	
Sample Size	1594	1555	663	299	1578	1539	661	299
R-squared	.46	.25	.50	.42	.54	.22	.57	.35

Mean Difference in PIAT Scores
(Bootstrapped SD in parentheses)

Daily TV hours at age 6-9	PIAT Math		PIAT Reading	
	G-est.	G-est. (S)	G-est.	G-est. (S)
1 hour versus 0 hour	-.38 (.57)	-.30 (1.25)	.42 (.54)	.31 (1.23)
2 hours versus 0 hour	-.64 (1.0)	-.54 (2.20)	.76 (.96)	.40 (2.18)
3 hours versus 0 hour	-.76 (1.31)	-.73 (2.84)	1.02 (1.25)	.28 (2.85)
4 hours versus 0 hour	-.75 (1.50)	-.86 (3.23)	1.19 (1.42)	-.06 (3.29)
5 hours versus 0 hour	-.62 (1.58)	-.94 (3.41)	1.27 (1.50)	-.62 (3.54)
6 hours versus 0 hour	-.35 (1.61)	-.96 (3.51)	1.28 (1.53)	-1.39 (3.68)
4 hours versus 2 hours	-.12 (.54)	-.32 (1.17)	.43 (.51)	-.46 (1.23)
6 hours versus 2 hours	-.37 (.95)	-.56 (2.05)	.52 (.85)	-1.78 (2.15)

Notes: * p<.1; ** p<.05; *** p<.01. Standard deviations are in the parentheses. The sample is composed of kids watching ten hours or less TV per day between ages 6-9. The controlled inputs include a child's race, sex, birth order, family backgrounds, and detailed current home inputs. OLS(S) – Detailed current school inputs are included.

Table 5: TV Watching on Math and Reading: G-Estimation with Binary Responses

The Marginal Effects

	PIAT Math (higher than mean)				PIAT Reading (higher than mean)			
	Probit		Probit (H)		Probit		Probit (H)	
	at 8-9	at 6-7	at 8-9	at 6-7	at 8-9	at 6-7	at 8-9	at 6-7
High TV at age 8-9	.05		-.006		.06		.10*	
	(.06)		(.07)		(.05)		(.05)	
Middle TV at age 8-9	-.007		-.07		-.02		.01	
	(.06)		(.06)		(.05)		(.05)	
High TV at age 6-7	.05	-.05	.06	-04	-.07	.01	-.07	.01
	(.06)	(.06)	(.06)	(.06)	(.06)	(.06)	(.06)	(.06)
Middle TV at age 6-7	-.03	-.04	-.002	-.05	-.004	-.001	-.01	.001
	(.05)	(.05)	(.06)	(.05)	(.05)	(.05)	(.05)	(.05)
PIAT score higher than mean at age 6-7	.37***		.40***		.51***		.51***	
	(.04)		(.04)		(.04)		(.05)	
PIAT score higher than mean at age 4-5	.03	.31***	.04	.26***	.14***	.32***	.13**	.29***
	(.04)	(.04)	(.04)	(.04)	(.05)	(.05)	(.05)	(.05)
Sample Size	720	720	651	651	687	687	622	622
Pseudo R-squared	.29	.23	.31	.22	.40	.26	.41	.23

Mean difference in probability of PIAT score higher than mean at age 8-9
(Bootstrapped SD in parentheses)

	PIAT Math		PIAT Reading	
Daily TV hours at age 6-9	G-est.	G-est. (H)	G-est.	G-est. (H)
High versus Low TV at ages 6-9	-.18**	-.23***	-.21***	-23***
	(.08)	(.08)	(.08)	(.09)
Middle versus Low TV at ages 6-9	-.16**	-.18**	-.13*	-.14*
	(.07)	(.07)	(.07)	(.08)
High versus Low TV at age 6-7 with Low at 8-9	-.23*	-.19	-.23*	-.26*
	(.13)	(.13)	(.12)	(.14)
Middle versus Low TV at age 6-7 with Low at 8-9	-.19**	-.20**	-.11	-.13
	(.08)	(.08)	(.08)	(.08)

Notes: * p<.1; ** p<.05; *** p<.01. Standard deviations are in the parentheses.
High TV: watching TV more than 4 hours per day; Middle TV: more than 2 but less than or equal to 4 hours per day; Low TV: less than or equal to 2 hours per day. The sample is composed of kids watching ten hours or less TV per day between ages 6-9. The controlled inputs include a child's race, sex, birth order, detailed home inputs at ages 6-9, home environment indicators at ages 2-3 and 4-5, and family backgrounds. Probit(H) -- Detailed home inputs at age 4-5 are included.