

DYNAMIC TREATMENT EFFECT ANALYSIS OF TV EFFECTS ON CHILD COGNITIVE DEVELOPMENT

FALI HUANG^a AND MYOUNG-JAE LEE^{b*}

^a *School of Economics, Singapore Management University, Singapore*

^b *Department of Economics, Korea University, Seoul, South Korea*

SUMMARY

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 placed on the last aspect, where TV watching affects cognitive development, which in turn affects future TV watching. When this feedback occurs, it is not straightforward to identify and estimate the TV effect. We develop a two-stage estimation method which can deal with the feedback feature; we also apply the ‘standard’ econometric panel data approaches. Overall, for math score at ages 8–9, we find that watching TV during ages 6–7 and 8–9 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 is evidence that watching no more than 2 hours of TV per day has a positive effect, whereas the effect is negative outside this range. In both cases, however, the effect magnitudes are economically small. Copyright © 2010 John Wiley & Sons, Ltd.

Received 30 March 2006; Revised 29 November 2008

1. INTRODUCTION

US children spend the second largest chunk of their waking time watching television (Juster and Stafford, 1991). That is, the most time-consuming activity after attending school is TV watching. For example, an average 8-year-old in NLSY79 (National Longitudinal Survey of Youth 1979) child sample spends about 25 hours per week in front of the television. Despite the widespread use of computers and the Internet, television (TV) remains the dominant form of media in children’s lives. A recent nationally representative survey found that 8–18 year-olds watch an average of 3 hours of television a day compared with 1 hour a day spent on recreational computer use (Rideout *et al.*, 2005; see also Chernin and Linebarger, 2005; Pecora *et al.*, 2006). Not surprisingly, the public and parents are concerned about potentially bad effects of child TV watching. The goal of this paper is to discover *the effects of TV watching on child cognitive development measured by standardized mathematics and reading scores*.

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, 2000). This relation is just one of many such relations where something that happened far back has long-term lingering effects. For

* Correspondence to: Myoung-Jae Lee, Department of Economics, Korea University, Anam-dong, Sungbuk-ku, Seoul 136-701, South Korea. E-mail: myoungjae@korea.ac.kr

instance, skill heterogeneity at age 16 has been shown to account for as much as 90% of the total variation of one's lifetime earnings (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 parents may share predispositions for certain habits and behaviors that cannot be measured. This is an 'individual-specific effect' problem, which may be overcome with panel data. Third, TV watching can affect cognitive development, which can in turn affect 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 on the effects of TV watching on cognitive development. An exception is Zavodny (2006), who studied the effects of TV watching for high school students and reviewed 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. In contrast to Zavodny (2006), who deals with individual-specific effects with panel data and family-specific effects with twins and siblings data, we tackle the dynamic nature of the causal relation; also we focus on young children, not on high school students. Another exception is Gentzkow and Shapiro (2008), 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. This 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 hours are concerned, the statistical analysis is mostly on the correlation rather than the causal relationship, and most studies found modest and negative associations of TV viewing with cognitive development of young children (Van Evra, 2004; Zimmerman and Christakis, 2005; Anderson, 2005).

Putting our main findings in advance, TV 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 TV watching during ages 6–7 and 8–9 is negative for math score at ages 8–9, where the 'indirect effect' of TV viewing at the younger ages 6–7 transmitted through the score at ages 6–7 is much larger than the direct effects of TV viewing at ages 6–7 and 8–9. For the reading score at ages 8–9, however, watching not more than 2 hours TV per day brings a small but positive effect compared to too much TV viewing. For example, watching 2 hours TV per day during ages 6–7 and 8–9 reduces one's math score at ages 8–9 by 4% of one standard deviation, but increases one's reading score at ages 8–9 by 3% of one standard deviation, compared with not watching TV at all.

This paper differs from the TV effect literature in economics on several grounds. First, we develop a two-stage procedure 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 in the two studies mentioned above. Third, with many control variables including detailed home and school inputs, we still find significant concave/convex effects of TV watching on both scores based on conventional methods of least squares estimator (LSE) and fixed-effect models, which differs from Zavodny (2006). Additionally, when the dynamic feedback feature of earlier TV watching is 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 effect of TV viewing at the earlier ages 6–7. This contrasts with the positive effects found by Gentzkow and Shapiro (2008). 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.

The rest of this paper is organized as follows. Section 2 reviews the ‘G-algorithm’ used in biostatistics, which can identify the desired full TV effects taking the dynamic causal feature into account, which then motivates our two-stage procedure using linear models. Section 3 presents typical econometric dynamic panel data approaches. Section 4 describes our data. Section 5 presents the empirical findings. Finally, Section 6 concludes.

2. G-ALGORITHM AND LINEAR-MODEL TWO-STAGE PROCEDURE

2.1. G-Algorithm Basics

Suppose

$$(x'_{i0}, y_{i0}, x'_{i1}, d_{i1}, y_{i1}, x'_{i2}, d_{i2}, y_{i2}), \quad i = 1, \dots, N$$

are observed and i.i.d. across $i = 1, \dots, N$. We will often omit the subscript i in the rest of this paper in view of the i.i.d. 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, omitting x_{it} ’s, is shown in Figure 1.

The feedback feature is $y_1 \rightarrow d_2$: 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 the *feedback feature* $y_1 \rightarrow d_2$ makes 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 : when treatment j is given exogenously at time 1,
- y_2^{jk} : when treatments j, k are given exogenously at time 1, 2 $j, k \in [0, \infty)$

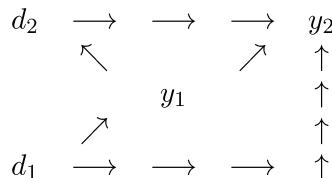


Figure 1. Dynamic Framework with Feedback from y_1 to d_2

With d_1 and d_2 observed, we have $y_1 = y_1^{d_1}$ and $y_2 = y_2^{d_1 d_2}$; i.e., only the potential responses corresponding to the realized treatment levels are observed, and all the other potential responses—‘counterfactuals’—are not. Also define the ‘potential treatment’ for d_2 :

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

with d_1 observed, we have $d_2 = d_2^{d_1}$.

Our goal is to find the mean effect $E(y_2^{j_0 k_0} - y_2^{00})$ for the intervened treatment ‘profile’ (j_0, k_0) 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 are 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.

Let

$$X_2 \equiv (x'_0, y_0, x'_1, x'_2)'$$

and denote the conditional independence of a and b given c as ‘ $a \perp\!\!\!\perp b|c$ ’. Assume ‘no unobserved confounder’ (NUC):

$$\text{NUC1: } y_2^{jk} \perp\!\!\!\perp d_1 | X_2, \quad \text{for } j = j_0, 0 \text{ and } k = k_0, 0$$

$$\text{NUC2: } y_2^{jk} \perp\!\!\!\perp d_2^j | (d_1 = j, y_1^j, X_2), \quad \text{for } j = j_0, 0 \text{ and } k = k_0, 0$$

If the interest is also in the effect of d_1 on y_1 , following the usual static counterfactual causal framework, assume additionally:

$$\text{NUC0: } y_1^j \perp\!\!\!\perp d_1 | (x_0, y_0, x_1)$$

NUC1 holds if d_1 is determined by X_2 and some error term independent of y_2^{jk} given X_2 . Saying ‘ d_1 determined by X_2 ’ may sound preposterous because parts of X_2 are realized after d_1 , but the dependence between d_1 and those components of X_2 in NUC1 should be construed as d_1 being allowed to affect y_2^{jk} through those components in X_2 . If one finds NUC2 somewhat confusing, it may help to rephrase it with densities: with $f(y_{2o}^{jk} | d_{2o}^j, d_{1o}, y_{1o}^j, X_{2o})$ denoting the conditional density/probability of $y_2^{jk} | (d_2^j, d_1, y_1^j, X_2)$ evaluated at $(y_{2o}^{jk}, d_{2o}^j, d_{1o}, y_{1o}^j, X_{2o})$, NUC2 states:

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

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_2 , and NUC2 states that the d_2^j selection is independent of y_2^{jk} given the observable $(d_1 = j, y_1^j, X_2)$.

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

$$E(y_2^{jk} | X_2) = \int E(y_2 | d_1 = j, d_2 = k, y_1, X_2) f(y_1 | d_1 = j, X_2) \partial y_1 \tag{1}$$

where $f(y_1|d_1 = j, X_2)$ denotes the conditional density of $y_1|(d_1 = j, X_2)$, ‘ ∂ ’ is used for integration/differentiation to avoid confusion with treatment d , and y_1 in equation (1) is an integration dummy, not a random variable.

The important point is that the right-hand side of (1) is identified and operational as it depends on observed variables, which means that $E(y_2^{jk}|X_2)$ is identified and can be estimated with the right-hand side. The G-algorithm holds because the right-hand side of (1) is

$$\begin{aligned} & \int E(y_2^{jk}|d_1 = j, d_2^j = k, y_1^j, X_2) f(y_1^j|d_1 = j, X_2) \partial y_1^j \\ &= \int E(y_2^{jk}|d_1 = j, y_1^j, X_2) f(y_1^j|d_1 = j, X_2) \partial y_1^j \quad (\text{due to NUC2}) \\ &= E(y_2^{jk}|d_1 = j, X_2) \quad (\text{for } y_1^j \text{ is integrated out}) \\ &= E(y_2^{jk}|X_2) \quad (\text{due to NUC1}) \end{aligned} \quad (2)$$

As NUC1 is invoked only at the last equality, if NUC1 does not hold one may go for $E(y_2^{jk}|d_1 = j, X_2) - E(y_2^{j0}|d_1 = j, X_2)$, which is the mean treatment effect on the ‘treated ($d_1 = j$)’—a kind of treatment effect in Lechner and Miquel (2001), as mentioned already.

In essence, the 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. The subpopulation is then generalized to the whole population (i.e., the selection problem 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 $E(y_2^{j_0 k_0}|X_2)$ and $E(y_2^{00}|X_2)$ and then integrating out X_2 , we obtain the desired

$$E(y_2^{j_0 k_0} - y_2^{00}) = E_{X_2}[\{\text{r.h.s. of (1) for } j = j_0, k = k_0\} - \{\text{r.h.s. of (1) for } j = 0, k = 0\}]$$

which is a ‘marginal’ effect relative to $E(y_2^{j_0 k_0} - y_2^{00}|X_2)$; $E_{X_2}[\cdot]$ means integrating out X_2 using its population distribution and ‘r.h.s.’ stands for ‘right-hand side’.

Even for two periods, implementing the G-algorithm in (1) requires finding $E(y_2|d_1 = j, d_2 = k, y_1, X_2)$ and $f(y_1|d_1 = j, X_2)$ first, and then carrying out the one-dimensional integration, which could be daunting. We will pursue two directions to avoid this computation problem. One is using a binary response, which turns the integral into a simple sum with two terms. The other is adopting linear models, which we will explain shortly.

The G-algorithm in the aforementioned papers of Robins assumes discrete treatments and covariates, and generalizing the G-algorithm to continuous treatments/covariates needs more conditions. Theorem 2 of Gill and Robins (2001) shows that the G-algorithm also works for continuous cases with zero probability of the conditioning event. Gill and Robins (2001) also show that a probability space can be constructed where the counterfactuals live with the observed variables, hence presupposing counterfactuals are ‘free’ or ‘harmless’ as they do not impose any extra restrictions on the factials. A review of dynamic treatment effects from the econometric viewpoint can be seen in Abbring and Heckman (2007).

2.2. A Simple Linear Model and G-Algorithm

To avoid the computation problem in the G-algorithm, we may adopt a linear model, which will also demonstrate well what is actually going on within the G-algorithm. Suppose that a simple linear model (omitting x_0 and y_0) holds:

$$\begin{aligned}
 d_1 &= \zeta_1 + x'_1 \zeta_x + \varepsilon_1 \\
 y_1^j &= \beta_1 + x'_1 \beta_x + \beta_d j + u_1 \quad (d_1 = j \text{ affects } y_1) \\
 d_2^j &= \zeta_1 + x'_2 \zeta_x + \zeta_y y_1^j + \varepsilon_2 \quad (y_1^j \text{ affects } d_2), \\
 y_2^{jk} &= \beta_1 + x'_2 \beta_x + \beta_{d\text{lag}} j + \beta_d k + \beta_y y_1^j + u_2 \quad (d_1 = j, d_2 = k, y_1^j \text{ affect } y_2)
 \end{aligned}
 \tag{3}$$

where ζ and β are parameters and $(\varepsilon_1, \varepsilon_2, u_1, u_2)$ are mean-zero errors. For this model, the desired effect is

$$\begin{aligned}
 &E(y_2^{j_0 k_0} - y_2^{00}) \\
 &= E\{\beta_1 + x'_2 \beta_x + \beta_{d\text{lag}} j_0 + \beta_d k_0 + \beta_y y_1^{j_0} + u_2 - (\beta_1 + x'_2 \beta_x + \beta_y y_1^0 + u_2)\} \\
 &= \beta_{d\text{lag}} j_0 + \beta_d k_0 + \beta_y E(y_1^{j_0} - y_1^0) = \beta_{d\text{lag}} j_0 + \beta_d k_0 + \beta_y \beta_d j_0
 \end{aligned}
 \tag{4}$$

This consists of two parts: the *direct effect* $\beta_{d\text{lag}} j_0 + \beta_d k_0$ of (d_1, d_2) on y_2 and the *indirect effect* $\beta_y \beta_d j_0$ of d_1 on y_2 through y_1 .

Barring some regularity conditions, the following condition is sufficient for NUC1 and NUC2, and thus for G-algorithm:

$$\begin{aligned}
 \text{(a):} & \quad (u_1, u_2) \perp\!\!\!\perp \varepsilon_1 | X_2 \quad \text{for NUC1} \\
 \text{(b):} & \quad u_2 \perp\!\!\!\perp \varepsilon_2 | (\varepsilon_1, u_1, X_2) \quad \text{for NUC2}
 \end{aligned}
 \tag{5}$$

In (3), given $(d_1 = j, y_1^j, X_2)$, y_2^{jk} and d_2^j depend only on u_2 and ε_2 , respectively. NUC2 holds thus if (b) holds. Analogously, given X_2 , d_1 depends only on ε_1 , and y_2^{jk} depends only on u_2 and y_1^j which in turn depends only on u_1 . Hence (a) implies NUC1.

To better understand (5), omit ‘ $|X_2$ ’ for a while and observe the fact:

$$(u_1, u_2) \perp\!\!\!\perp (\varepsilon_1, \varepsilon_2) \iff (u_1, u_2) \perp\!\!\!\perp \varepsilon_2 | \varepsilon_1 \text{ and } (u_1, u_2) \perp\!\!\!\perp \varepsilon_1$$

in $(u_1, u_2) \perp\!\!\!\perp \varepsilon_2 | \varepsilon_1$, it looks as if ε_2 may be related to (u_1, u_2) through ε_1 , but this is ruled out in $(u_1, u_2) \perp\!\!\!\perp \varepsilon_1$. In comparison to this display, in (5) with ‘ $|X_2$ ’ omitted, we get $(u_1, u_2) \perp\!\!\!\perp \varepsilon_1$ and $u_2 \perp\!\!\!\perp \varepsilon_2 | (\varepsilon_1, u_1)$: ε_2 may be related to u_2 through u_1 . Also heteroskedasticity and serial correlations within $(\varepsilon_1, \varepsilon_2)$ as well as within (u_1, u_2) are allowed.

A condition stronger than (5) is

$$(u_1, u_2) \perp\!\!\!\perp (\varepsilon_1, \varepsilon_2) | X_2$$

This is easier to understand: the selection equation (i.e., the treatment equation) error terms should be unrelated to the outcome equation errors conditional on the covariates—the familiar ‘static’ selection-on-observable condition. The only difference between this and (5) is that

(5)(b) has u_1 in the conditioning set, which is necessary as y_1^j gets conditioned on in G-algorithm to allow for the dynamic feedback. Because y_1 is observed at period 2, (5) is also a selection-on-observable condition.

Getting into G-algorithm details with the linear model (3), observe

$$\begin{aligned} & E(y_2^{jk} | d_1 = j, d_2^j = k, y_1^j, X_2) \\ &= \beta_1 + x_2' \beta_x + \beta_{d\text{lag}} j + \beta_d k + \beta_y y_1^j + E(u_2 | d_1 = j, d_2^j = k, y_1^j, X_2) \\ &= \beta_1 + x_2' \beta_x + \beta_{d\text{lag}} j + \beta_d k + \beta_y y_1^j + E(u_2 | d_1 = j, y_1^j, X_2) \text{ using (5)(b)} \end{aligned} \tag{6}$$

Substitute this into the G-algorithm to get

$$\begin{aligned} & E(y_2^{jk} | X_2) \\ &= \int \{ \beta_1 + x_2' \beta_x + \beta_{d\text{lag}} j + \beta_d k + \beta_y y_1^j + E(u_2 | d_1 = j, y_1^j, X_2) \} f(y_1^j | d_1 = j, X_2) \partial y_1^j \\ &= \beta_1 + x_2' \beta_x + \beta_{d\text{lag}} j + \beta_d k + \beta_y E(y_1^j | d_1 = j, X_2) + E(u_2 | d_1 = j, X_2) \end{aligned}$$

Observe

$$\begin{aligned} & E(y_1^j | d_1 = j, X_2) = \beta_1 + x_1' \beta_x + \beta_d j + E(u_1 | d_1 = j, X_2) \\ & E(u_1 | d_1 = j, X_2) = 0 \text{ and } E(u_2 | d_1 = j, X_2) = 0 \text{ using (5)(a)} \end{aligned} \tag{7}$$

Substitute these into the preceding display to obtain

$$\begin{aligned} & E(y_2^{jk} | X_2) = \beta_1 + x_2' \beta_x + \beta_{d\text{lag}} j + \beta_d k + \beta_y (\beta_1 + x_1' \beta_x + \beta_d j) \\ \implies & E(y_2^{jk} - y_2^{00} | X_2) = \beta_{d\text{lag}} j + \beta_d k + \beta_y \beta_d j = E(y_2^{jk} - y_2^{00}) \end{aligned}$$

Comparing the last display with (4), we can see that $E(y_2^{jk} - y_2^{00})$ is identified and estimable with the G-algorithm. As will be shown later, typical econometric panel dynamic models will miss the effect conveyed by y_1 , i.e., the part due to $E(y_1^j | d_1 = j, X_2)$. Note that (5) is sufficient for the G-algorithm to work with the linear model; the necessary and sufficient condition can be seen in (7) and (6):

$$\begin{aligned} \text{(a) : } & E(u_1 | d_1 = j, X_2) = 0 \text{ and } E(u_2 | d_1 = j, X_2) = 0 \text{ for NUC1} \\ \text{(b) : } & E(u_2 | d_1 = j, d_2^j = k, y_1^j, X_2) = E(u_2 | d_1 = j, y_1^j, X_2) \text{ for NUC2} \end{aligned} \tag{8}$$

Although we considered the linear model to simplify G-algorithm computation, (4) shows that the effect $E(y_2^{jk} - y_2^{00})$ can be found by estimating the y_2^{jk} equation in a single step because only $\beta_{d\text{lag}}$, β_d , and β_y appear in $E(y_2^{jk} - y_2^{00})$; finding $E(y_2 | d_1 = j, d_2 = k, y_1, X_2)$ and $f(y_1 | d_1 = j, X_2)$ and then doing the integration in the G-algorithm is all unnecessary. The linear model (3), however, looks too simple to be realistic. For one thing, it assumes linear effects of d_1 and d_2 ; for another, it assumes a stationary environment. For child TV watching, both restrictions

are implausible, and hence we consider a nonstationary model allowing for nonlinear effects of d_1 and d_2 next, which will then lead to a two-stage procedure to be used in our empirical analysis.

2.3. A More General Linear Model and Two-Stage Procedure

Instead of (3), consider

$$\begin{aligned} & (d_1 \text{ and } d_2^j \text{ equations are the same as in (3)}) \\ y_1^j &= \alpha_1 + x_1' \alpha_x + \alpha_{d1} j + \alpha_{d1q} j^2 + u_1 \\ y_2^{jk} &= \beta_1 + x_2' \beta_x + \beta_{d1} j + \beta_{d1q} j^2 + \beta_{d2} k + \beta_{d2q} k^2 + \beta_y y_1^j + u_2 \end{aligned} \quad (9)$$

This generalizes (3) for nonlinear treatment effects and non-stationary parameters by introducing quadratic terms in j and k and allowing the parameters of the y_1^j equation to differ from those of the y_2^{jk} equation. The quadratic terms j^2 and k^2 are to account for the potential nonlinear effects 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 the squared y_1^j can be included as well in the y_2^{jk} equation. But, for our data with $N \simeq 2180$ and high-dimensional x_1 and x_2 , adding high-order terms can quickly get out of hand.

The desired effect is

$$\begin{aligned} & E(y_2^{j_0 k_0} - y_2^{00}) \\ &= E\{\beta_1 + x_2' \beta_x + \beta_{d1} j_0 + \beta_{d1q} j_0^2 + \beta_{d2} k_0 + \beta_{d2q} k_0^2 + \beta_y y_1^{j_0} + u_2 - (\beta_1 + x_2' \beta_x + \beta_y y_1^0 + u_2)\} \\ &= \beta_{d1} j_0 + \beta_{d1q} j_0^2 + \beta_{d2} k_0 + \beta_{d2q} k_0^2 + \beta_y E(y_1^{j_0} - y_1^0) \\ &= \beta_{d1} j_0 + \beta_{d1q} j_0^2 + \beta_{d2} k_0 + \beta_{d2q} k_0^2 + \beta_y (\alpha_{d1} j_0 + \alpha_{d1q} j_0^2) \end{aligned} \quad (10)$$

Although we can show that the G-algorithm identifies this effect, it is not necessary, because we can just *estimate the y_1^j and y_2^{jk} equations separately for the α and β parameters first, and then combine those estimators to find the desired effect*. This two-stage procedure is our main proposal. As in (4), the parameters in the d_1 and d_2^j equations do not appear in the desired effect, which is why we did not bother with generalizing the d_1 and d_2^j equations in (9).

Estimating the α and β parameters in the linear y_1^j and y_2^{jk} equations brings up a new issue for observational data. For instance, consider the y_1 equation in the observed random variables:

$$y_1 = \alpha_1 + x_1' \alpha_x + \alpha_{d1} d_1 + \alpha_{d1q} d_1^2 + u_1 \quad (11)$$

For this to be taken as a data-generating process, we need to specify how d_1 and u_1 (and x_1) are related to each other. In the y_1^j equation, we may assume $E(u_1|x_1) = 0$, but this is moot on how d_1 is related to u_1 —no d_1 in $E(u_1|x_1)$ is natural as j is a constant in the y_1^j equation. In contrast, in order to estimate the α parameters in (11) with LSE, we need $E(u_1|x_1, d_1) = 0$ (or at least zero correlations), which in fact appeared in (8) for the G-algorithm.

Specifically, getting $E(y_1|x_1, d_1) = \alpha_1 + x'_1\alpha_x + \alpha_{d_1}d_1 + \alpha_{d_1q}d_1^2$ from

$$E(y_1^j|x_1) = \alpha_1 + x'_1\alpha_x + \alpha_{d_1}j + \alpha_{d_1q}j^2 \iff E(u_1|x_1) = 0$$

in (9) requires

$$E(y_1^j|d_1 = j, x_1) = E(y_1^j|x_1) \iff E(u_1|d_1 = j, x_1) = E(u_1|x_1) = 0 \quad (12)$$

Formally, this can be seen in

$$\begin{aligned} E(y_1|d_1, x_1) &= \int E(y_1|d_1 = j, x_1)\partial 1[j \leq d_1] = \int E(y_1^j|d_1 = j, x_1)\partial 1[j \leq d_1] \\ &= \int E(y_1^j|x_1)\partial 1[j \leq d_1] \quad (\text{using (12)}) \\ &= \int (\alpha_1 + x'_1\alpha_x + \alpha_{d_1}j + \alpha_{d_1q}j^2)\partial 1[j \leq d_1] = \alpha_1 + x'_1\alpha_x + \alpha_{d_1}d_1 + \alpha_{d_1q}d_1^2 \end{aligned}$$

where $1[j \leq d_1]$ denotes the distribution for j with unit mass at $j = d_1$. The requirement (12)—i.e., no selection problem from d_1 —was in fact noted in Lee and Kobayashi (2001) when d_1 is binary. Equation (12) is weaker than (8)(a) for the G-algorithm in that x_2 does not appear in the conditioning set.

Analogously, $E(y_2|d_1, d_2, y_1, x_2) = \beta_1 + x'_2\beta_x + \beta_{d_1}d_1 + \beta_{d_1q}d_1^2 + \beta_{d_2}d_2 + \beta_{d_2q}d_2^2 + \beta_y y_1^j$ from

$$E(y_2^{jk}|y_1^j, x_2) = \beta_1 + x'_2\beta_x + \beta_{d_1}j + \beta_{d_1q}j^2 + \beta_{d_2}k + \beta_{d_2q}k^2 + \beta_y y_1^j \iff E(u_2|y_1^j, x_2) = 0$$

which requires

$$\begin{aligned} E(y_2^{jk}|d_1 = j, d_2 = k, y_1^j, x_2) &= E(y_2^{jk}|y_1^j, x_2) \\ \iff E(u_2|d_1 = j, d_2 = k, y_1^j, x_2) &= E(u_2|y_1^j, x_2) (= 0) \end{aligned} \quad (13)$$

Equation (13) is weaker than (8)(b) for the G-algorithm in that x_1 does not appear in the conditioning set. The separate appearance of x_1 and x_2 in (12) and (13) is due to estimating the period 1 and 2 equations separately in the two-stage procedure, while the G-algorithm does the computation together and thus needs X_2 in the conditioning set. Earlier we noted that the dynamic feedback nature $y_1 \rightarrow d_2$ makes d_2 non-randomized even if d_1 were so. NUC2 and (13) show clearly how this problem gets handled: assume that d_2 is as good as randomized conditioning on y_1^j (and other variables).

The main advantage of the two-stage procedure is ease of implementation: only two LSEs are needed. Of course, this practicality comes at the cost of specifying the linear functional form for the conditional mean in the G-algorithm which is nonparametric in its original form. Other than the G-algorithm and our two-stage linear-model-based procedure, other approaches are certainly possible for dynamic treatment effects. For instance, Lechner (2004, 2008) proposes matching approaches using the ‘propensity score’ idea in Rosenbaum and Rubin (1983), which is, however, applicable only when the treatment is binary.

2.4. Covariates Affected by Early Treatment or Response

Turning back to the G-algorithm, there is a complication in the G-algorithm in relation to x_2 being affected by either y_1 or d_1 , and we address this issue in this subsection. Note that, by construction, x_1 precedes d_1 and y_1 , and thus x_1 cannot be affected by d_1 or y_1 .

First, suppose that d_1 affects x_2 , which in turn affects y_2 (i.e., $d_1 \rightarrow x_2 \rightarrow y_2$). For example, the period 1 TV hours may affect the physical fitness x_2 of the child and the fitness may influence y_2 . In this case, the role of x_2 differs little from that of the early response y_1 . This means that x_2 affected by d_1 can be handled in the same way as y_1 is handled: replace y_1 in the G-algorithm with (y_1, x_2) and integrate (y_1, x_2) out conditioning, not on X_2 , but on

$$X_1 \equiv (x'_0, y_0, x'_1)'$$

Specifically, with x_2^j denoting the potential version for x_2 when $d_1 = j$ is exogenously set, we need new NUCs conditioned on X_1 :

$$\text{NUC1}' : y_2^{jk} \Pi d_1 | X_1, \quad \text{for } j = j_o, 0 \text{ and } k = k_o, 0 \quad (14)$$

$$\text{NUC2}' : y_2^{jk} \Pi d_2^j | (d_1 = j, y_1^j, x_2^j, X_1), \quad \text{for } j = j_o, 0 \text{ and } k = k_o, 0$$

Under this, the G-algorithm (1) should be modified as

$$\begin{aligned} E(y_2^{jk} | X_1) &= \int \int E(y_2 | d_1 = j, d_2 = k, y_1, x_2, X_1) f(y_1, x_2 | d_1 = j, X_1) \partial y_1 \partial x_2 \\ &= \int \int E(y_2^{jk} | d_1 = j, d_2^j = k, y_1^j, x_2^j, X_1) f(y_1^j, x_2^j | d_1 = j, X_1) \partial y_1^j \partial x_2^j \\ &= \int \int E(y_2^{jk} | d_1 = j, y_1^j, x_2^j, X_1) f(y_1^j, x_2^j | d_1 = j, X_1) \partial y_1^j \partial x_2^j \quad (\text{NUC2}' \text{ used}) \quad (15) \\ &= E(y_2^{jk} | d_1 = j, X_1) \quad (y_1^j \text{ and } x_2^j \text{ integrated out}) \\ &= E(y_2^{jk} | X_1) \quad (\text{NUC1}' \text{ used}) \end{aligned}$$

We get $E(y_2^{j_o k_o} - y_2^{00} | X_1)$, not $E(y_2^{j_o k_o} - y_2^{00} | X_2)$, from which $E(y_2^{j_o k_o} - y_2^{00})$ can be obtained by integrating out X_1 . In our data, we do not have period 2 covariates for $d_1 \rightarrow x_2 \rightarrow y_2$, and thus this generalization of the G-algorithm will not be needed.

Second, suppose that y_1 affects x_2 , which in turn affects y_2 (i.e., $y_1 \rightarrow x_2 \rightarrow y_2$). 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), and this can affect the period 2 test score. As in the preceding case, this requires the x_2^j equation, and the modified G-algorithm (15) still works. But the problem is that there might be too many such variables in our data, which makes the G-algorithm non-operational. We will assume that the effects of such variables are negligible. As the disciplinary/rewarding measure example suggests, those variables are of temporary nature rather than persistent, and if there are any effects they would be small.

To better see the problems we will encounter if we try to account for all effects in the latter case $y_1 \rightarrow x_2 \rightarrow y_2$, observe that $E(y_2^{j_o k_o} - y_2^{00})$ in (4) consists of three effects shown by three arrows pointing to y_2 in Figure 2; the part $y_1 \rightarrow d_2$ in Figure 1 is 'blocked' as $E(y_2^{j_o k_o} - y_2^{00})$

is the ‘intervention’ effect of setting $d_1 = j_o$ and $d_2 = k_o$ exogenously given the observables. In Figure 3 with $y_1 \rightarrow x_2 \rightarrow y_2$ added, there are now four arrows pointing to y_2 . The G-algorithm can accommodate this with (15), but if the dimension of x_2 is large then the G-algorithm is not operational. Our two-stage procedure can also accommodate the added arrow effect, but this will make (10) too complicated by adding too many terms.

In principle, there is no compelling reason why we should exclude the effect from $d_1 \rightarrow y_1 \rightarrow x_2 \rightarrow y_2$. But, as argued above, this effect seems negligible in our application, as one can immediately see after examining the list of home inputs at ages 6–7 and 8–9 in the Appendix which are mostly family routines represented by binary variables that these intermediate covariates are not likely to change responding to earlier test outcomes. For example, parents may reduce the number of trips to museums after seeing a bad test score, but so long as they take the child to museums at least several times a year, the value of the binary variable does not change according to the way the museum visit variables are constructed. As in any treatment effect analysis, we will never be able to account for all effects, as ‘a butterfly’s flutter in one side of the Pacific may cause a hurricane on the other side’. It is either the foreseeable negligible magnitude or the technical difficulty of gauging the effect that keeps us from pursuing all effects.

2.5. G-Algorithm with Binary Responses

We proposed a linear model approach to the G-algorithm to alleviate the implementation problem of the G-algorithm in (1). If many high-order terms should appear (particularly y_1^2, y_1^3, \dots 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, \dots any more—in which case the G-algorithm becomes a sum of two terms:

$$E(y_2^{jk} | y_0, X_2) = P(y_2 = 1 | d_1 = j, d_2 = k, y_1 = 0, X_2) \cdot P(y_1 = 0 | d_1 = j, X_2) + P(y_2 = 1 | d_1 = j, d_2 = k, y_1 = 1, X_2) \cdot P(y_1 = 1 | d_1 = j, X_2) \tag{16}$$

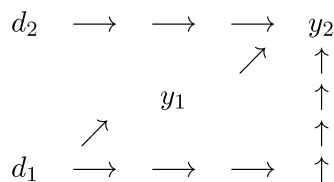


Figure 2. Dynamic Framework without Feedback

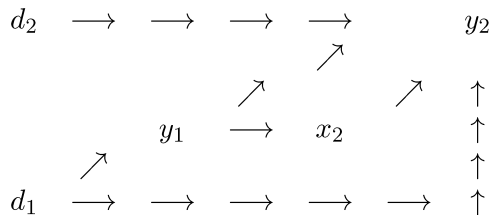


Figure 3. Dynamic Framework without Feedback from y_1 to z_2

Other than for avoiding y_1^2, y_1^3, \dots , binary response may be less susceptible to recall errors; in our data, the TV watching hours may be subject to recall errors.

For instance, apply probit (or logit) to y_2 on d_1, d_2, y_1 and X_2 to obtain the two probit probabilities for $y_2 = 1$ in (16):

$$\Phi(\psi_1 + \psi_{d1}d_1 + \psi_{d2}d_2 + \psi_{y1}y_1 + \psi'_x X_2)$$

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

$$\Phi(\eta_1 + \eta_{d1}d_1 + \eta'_x X_2)$$

where the η -parameters are to be estimated. Substituting these into (16) will do. This version will be applied to our data as well.

3. DYNAMIC PANEL DATA MODEL ESTIMATORS

Facing the task of finding the effect of (d_1, d_2) on y_2 when d_1 can influence y_2 through y_1 as well, the usual econometric panel data approach would be using a dynamic panel data model. Various dynamic panel data models are conceivable (see, e.g., Lee, 2002), but a comprehensive model is

$$y_2 = \beta_1 + x'_1\beta_{x1} + x'_2\beta_{x2} + \beta_{d1}d_1 + \beta_{d1q}d_1^2 + \beta_{d2}d_2 + \beta_{d2q}d_2^2 + \beta_{y0}y_0 + \beta_{y1}y_1 + v_2 \quad (17)$$

where v_2 is an error term. But this model cannot deliver the desired effect $E(y_2^{j_0k_0} - y_2^{00})$ in (10) because the *indirect effect* $\beta_y(\alpha_{d1}j_0 + \alpha_{d1q}j_0^2)$ of d_1 on y_2 through y_1 is missed. This, which is due to controlling y_1 , is the main disadvantage of the usual dynamic panel data approach. Our proposed two-stage method is a simple adjustment of this: estimate the y_1 equation as well, and construct (10) adding the missed indirect effect $\beta_y(\alpha_{d1}j_0 + \alpha_{d1q}j_0^2)$.

If we do not control y_1 in (17) to avoid the problem of missing the indirect effect, then the effect of d_2 on y_2 may get distorted because the omitted y_1 becomes a 'common factor' for d_2 and y_2 , causing an omitted variable bias. 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*: control y_1 to miss the indirect effect, or not control y_1 to incur an omitted variable bias.

For our empirical analysis, we will use and compare a number of estimators: first, LSE to (17), in which the test for ' H_0 : all coefficients for the treatments are zero', amounts to 'Granger noncausality test' (Granger, 1969, 1980); second, 'fixed-effect estimator' to the model (with $\Delta y_{it} \equiv y_{it} - y_{i,t-1}$ and $\Delta d_{it}^2 \equiv d_{it}^2 - d_{i,t-1}^2$):

$$\begin{aligned} y_{it} &= \alpha_t + \alpha_d d_{it} + \alpha_{dq} d_{it}^2 + x'_{it} \alpha_{xt} + \delta_i + v_{it} \\ \implies \Delta y_{i2} &= \alpha_2 - \alpha_1 + \alpha_d \Delta d_{i2} + \alpha_{dq} \Delta d_{i2}^2 + x'_{i2} \alpha_{x2} - x'_{i1} \alpha_{x1} + \Delta v_{i2} \end{aligned} \quad (18)$$

Third, in this differenced model, the lagged variables in Δy_{i2} , Δd_{i2} , and Δd_{i2}^2 appear with the restricted coefficient -1 , and if we remove these restrictions we get (17) (without y_0); the

instrumental variable estimator (IVE) to (17) is our third estimator, which is the most ‘endogeneity-conscious’ estimator. Fourth is LSE to the observed version of (9):

$$\begin{aligned} y_1 &= \alpha_1 + x'_1 \alpha_x + \alpha_{d1} d_1 + \alpha_{d1q} d_1^2 + u_1 \\ y_2 &= \beta_1 + x'_2 \beta_x + \beta_{d1} d_1 + \beta_{d1q} d_1^2 + \beta_{d2} d_2 + \beta_{d2q} d_2^2 + \beta_y y_1 + u_2 \end{aligned} \quad (19)$$

which is our main two-stage estimator. Fifth is the G-algorithm to binary response (16) after transforming y_t to the binary response taking 0 or 1, depending on whether y_t is smaller than its mean or not.

As for Granger causality, we consider this for two reasons. Firstly, although we adopt the ‘counterfactual causality’ framework as explained in the preceding section, dissenting views on causality 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. Secondly, although Granger causality does not imply nor is implied by the counterfactual causality in general, Robins *et al.* (1999) show that the two concepts do agree in some cases; see also Lechner (2006) on the comparison of counterfactual and Granger-type causalities. In our framework, the magnitude of the (d_1, d_2) effect estimated in (17) is not meaningful as (17) misses the indirect effect, as already pointed out, but *Granger noncausality of (d_1, d_2) is meaningful under the stationary environment* of the same α and β parameters, which can be seen in (4):

$$E(y_2^{j_0 k_0} - y_2^{00}) = \beta_{d \text{lag}} j_0 + \beta_d k_0 + \beta_y \beta_d j_0 = 0 \quad \text{if } \beta_{d \text{lag}} = \beta_d = 0$$

That is, the missed indirect effect is also zero when the direct effect is zero thanks to the ‘stationarity’ assumption that β_d appears in the y_1^j equation as well as in the y_2^{jk} equation in (3).

Turning to the fixed-effect estimator, it is one of the most popular estimators in panel data, which is one reason why we apply this to our data. As is well known, an important source of endogeneity in a panel data model is the ‘unobserved unit-specific effect δ_i ’ when v_{it} is redefined as $\delta_i + v_{it}$. The endogeneity is then removed by differencing the model as in (18). To further remove any remaining endogeneity, IVE can be applied to the differenced model, but we chose to apply IVE to (17) for even more generality because (17) includes the differenced model as a sub-model.

Note that removing δ_i by first-differencing requires a kind of stationarity that δ_i enters both equations with the same ‘slope’. If one believes that the TV effect is stationary, being the same at all ages, then the same slope specification for δ_i may be adopted and differencing may be the way to go. The effect found there would be applicable to all ages, leading to a ‘long-run effect’, despite that the NUCs hold only at two periods. But this scenario seems overly optimistic. Rather, the TV effects and the δ_i slopes are all likely to be non-stationary. We therefore allow the nonstationary aspect in our two-stage procedure, and this should matter greatly.

In econometrics, most ‘energy’ seems to have been spent on dealing with endogeneity, in contrast to the statistical literature for the G-algorithm, where accounting for the full effect is given more priority while the endogeneity problems are ‘swept under the rug’ with NUC conditions. It is therefore interesting to see how the dynamic panel data (IVE) econometric approaches with emphasis on endogeneity differ from G-algorithm-based approaches with emphasis on getting the full effect.

As for instruments, Angrist and Krueger (2001) show an ingenious list of instruments in various studies, but having those types 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 (17) involves already y_1 , d_2 , d_1 , x_2 , and x_1 , the only source left for instruments for endogenous regressors is x_0 , and x_0 will be the source for our instruments. As x_0 is not an ‘ideal’ instrument, this part of the empirical analysis should be taken with a grain of salt.

Related to endogeneity, we point out that the identification of the desired effect $E(y_2^{j_0k_0} - y_2^{00})$ as in (10) is a separate matter from specifying the data-generating process and getting concerned about the ensuing endogeneity there. In dynamic panel data models, often the interest falls on the initial condition y_0 and the unobserved unit-specific effect δ . This is relevant to estimating the model with observed data, but not to specifying the desired effect. To see this, observe that no matter what y_0 and δ might be, so long as they do not interact with the treatments, they get canceled out in the difference $y_2^{j_0k_0} - y_2^{00}$, and thus have no impact on the desired effect.

4. DATA DESCRIPTION

The NLSY79 child sample contains rich information on children born to women respondents of the NLSY79. Starting from 1986, a set of questionnaires was developed to collect information about the cognitive, social, and behavioral development of children. The collected 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 the Peabody Individual Achievement Test (PIAT) math and reading scores at ages 8–9 (y_2) and TV watching hours at ages 6–7 and 8–9 (d_1 , d_2). The first period is ages 6–7, and the second period is ages 8–9; the ages 4–5 are then used as the base period (period 0). When we include PIAT scores at ages 6–7 (y_1), the available sample size remains almost the same; it gets smaller when PIAT scores at ages 4–5 (y_0) are included, since only children aged 5 and above are tested with PIAT. The change of sample size is shown at the bottom of Table I. The missing data for PIAT scores at ages 4–5 are due to the simple fact that many children of ages 4–5 are below age 5 and hence not eligible for the tests.

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 10 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 I, and the complete set of variables is provided in the Appendix. Since school inputs were not available for children above 5 years old before 1992, there are many missing entries in school inputs.

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 USA in 1968. The PIAT math in our sample

Table I. TV watching and math and reading scores: summary statistics

	Daily TV watching hours		PIAT scores at ages 8–9		Group size
	At ages 8–9	At ages 6–7	Math	Reading	
Main sample	3.5 (2.0)	3.3 (2.0)	102.2 (13.9)	105 (14.6)	2180
<i>Race</i>					
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
<i>Sex</i>					
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
<i>Birth order</i>					
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
<i>How many children books a child has at home at ages 6–7</i>					
≥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
<i>Mother reads to child at ages 6–7</i>					
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
<i>Whether parents discuss TV programs with a child</i>					
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
<i>School type</i>					
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
<i>Mother's highest grade</i>					
≥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
<i>Mother's AFQT score in 1981</i>					
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
<i>Earlier PIAT scores</i>					
PIAT Math at ages 6–7			100.8 (12.1)		2180
PIAT Math at ages 4–5			99.2 (14.7)		1250
PIAT Reading at ages 6–7			103.3 (11.7)		2180
PIAT Reading at ages 4–5			106.5 (14.8)		1175

Note: The entries are group means and standard deviations (in parentheses). The main sample is composed of children watching 10 hours or less TV per day between ages 6 and 9.

has mean 102.2 and SD 13.9 around age 9, 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

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

White children on average watch about 1 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 first-borns watch about half an hour less than the others and get higher math and reading scores. A child with 10 or more children's books at home watches around 1.2 hours less TV at ages 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 1 hour more TV on a daily basis at ages 6–7 and 8–9 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, does not seem to affect much TV watching time, although it is 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, which is possibly due to multitasking, where many children watch TV while doing homework (Rideout *et al.*, 2005).

Children with mothers having 16 or more years of schooling watch about 1 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 I 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 overestimated 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 9 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 4, which causes many missing values. Whenever possible, HOME is included as a control in addition to detailed home inputs.

In discussing the dynamic treatment effects, we desire the time sequence $x_t \rightarrow d_t \rightarrow y_t$ in each period so that x_t works as the time- t baseline covariates which affect d_t and possibly y_t , and then the treatment d_t affects y_t . In our data, this temporal order is plausible for a couple of reasons.

First, y_t is measured on the interview day, which means that x_t and d_t precede y_t . Second, many family characteristics in x_t are likely to be determined independently of d_t . Third, overall, TV watching hours tend to be the ‘residual’ usage of time, and thus is unlikely to influence other time-consuming activities, although we cannot completely rule out TV watching taking precedence over the other activities.

The availability of a rich set of control variables in our data is crucial for the application of G-estimation, which relies on the credibility of the ‘no unobserved confounder’ assumption. As listed in the Appendix, our control variables contain four groups: (1) a child’s demographic information including race, sex, and birth order; (2) home inputs at different ages 4–5, 6–7, and 8–9, which cover many aspects of a child’s life at home, such as the number of children’s books at home, how often the mother reads to the child, how often the child is taken to a museum, performance, outdoor activities, meeting with relatives, having meals with parents, etc.; (3) school inputs at ages 6–7 and 8–9, including the number of hours spent on math and writing homework, whether the school is public or private, the mother’s rating of school quality such as the skills of the principal and teachers, and the safety and order of the school; (4) family backgrounds including the mother’s AFQT score, highest grade, marriage status, age at the child’s birth, whether the child was breast-fed, and family income. These variables constitute a highly comprehensive set of controls for our purpose of finding the dynamic effects of TV viewing on math and reading scores for children at ages 8–9. After controlling them, it is not easy to think of any important unobserved confounders that may affect both TV viewing time and math and reading scores. That said, a better dataset should also include more information about the child’s father, such as his AFQT score and education level, and more objective variables on school quality.

5. EMPIRICAL RESULTS

5.1. Granger Causality

Table II presents results for Granger causality for model (17). The various specifications differ mainly in the control variables used. In the first column (‘OLS’) of both math and reading score 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 background 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 substantially when the school inputs at ages 6–7 and 8–9 are controlled in the second column (‘OLS(S)’). Home inputs at ages 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 ages 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 sometimes not significant; TV hours at ages 6–7 have positive and concave effects, while those at ages 8–9 have negative and convex effects. For reading scores at ages 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 a p -value 0.009. The effects of TV watching hours at both age periods seem to have positive and concave effects.

For later comparison with the other tables, the mean differences in PIAT scores between different TV watching hours at ages 6–7 and 8–9 are listed in the lower half of Table II. More time spent

Table II. TV watching on math and reading: Granger causality

	PIAT Math at ages 8–9			PIAT Reading at ages 8–9		
	OLS	OLS(S)	OLS(SH)	OLS	OLS(S)	OLS(SH)
Daily TV hours at ages 8–9	–1.05 (0.77)	–2.15 (1.74)	–2.68 (2.52)	–0.06 (0.61)	2.21 (1.83)	2.49 (2.60)
Daily TV hours at ages 8–9 squared	0.16** (0.08)	0.36** (0.18)	0.40 (0.26)	0.03 (0.07)	–0.23 (0.18)	–0.25 (0.25)
Daily TV hours at ages 6–7	0.89 (0.68)	4.79** (1.89)	5.50** (2.31)	0.58 (0.60)	1.96 (1.79)	2.54 (2.0)
Daily TV hours at ages 6–7 squared	–0.11 (0.08)	–0.56*** (0.21)	–0.66** (0.26)	–0.07 (0.07)	–0.37* (0.21)	–0.57** (0.23)
PIAT score at ages 6–7	0.59*** (0.04)	0.58*** (0.12)	0.65*** (0.13)	0.78*** (0.03)	0.83*** (0.09)	0.88*** (0.12)
PIAT score at ages 4–5	0.10*** (0.03)	0.13 (0.09)	0.12 (0.11)	0.09*** (0.03)	0.06 (0.08)	0.09 (0.09)
Joint significance of TV at 6–9	0.07*	0.03**	0.04**	0.60	0.12	0.009***
Joint significance of TV at 8–9	0.02**	0.02**	0.15	0.50	0.46	0.61
Joint significance of TV at 6–7	0.38	0.03**	0.048**	0.60	0.04**	0.001***
Sample size	871	178	155	835	175	152
R ²	0.49	0.71	0.72	0.64	0.75	0.79

Mean difference in PIAT scores (bootstrapped SD in parentheses)

Daily TV hours at ages 6–9	PIAT Math at ages 8–9			PIAT Reading at ages 8–9		
	OLS	OLS(S)	OLS(SH)	OLS	OLS(S)	OLS(SH)
1 hour versus 0 hour	–0.12 (0.72)	2.43 (2.68)	2.56 (5.82)	0.49 (0.61)	3.57 (2.44)	4.20 (4.23)
2 hours versus 0 hour	–0.13 (1.27)	4.46 (4.73)	4.62 (9.95)	0.90 (1.07)	5.94 (4.30)	6.77 (7.34)
3 hours versus 0 hour	–0.04 (1.66)	6.08 (6.21)	6.17 (12.5)	1.25 (1.39)	7.11 (5.59)	7.69 (9.41)
4 hours versus 0 hour	0.15 (1.9)	7.28 (7.16)	7.21 (13.5)	1.52 (1.59)	7.07 (6.38)	6.97 (10.57)
5 hours versus 0 hour	0.45 (2.0)	8.08 (7.68)	7.73 (13.5)	1.72 (1.67)	5.83 (6.77)	4.61 (11.06)
6 hours versus 0 hour	0.85 (2.0)	8.47 (7.96)	7.75 (13.1)	1.86 (1.67)	3.38 (6.95)	0.60 (11.35)
4 hours versus 2 hours	0.28 (0.67)	2.82 (2.65)	2.58 (4.37)	0.62 (0.56)	1.13 (2.32)	0.20 (3.78)
6 hours versus 2 hours	0.98 (1.0)	4.01 (4.53)	3.13 (8.73)	0.95 (0.88)	–2.56 (4.04)	–6.16 (7.26)

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard deviations are in parentheses. The sample is composed of children watching 10 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 6–7 and 8–9, and family background (mother's AFQT score, her age at childbirth, whether the child was breast-fed, her marriage status, highest grade, and family income). OLS(S): school inputs at ages 6–7 and 8–9 as listed in the Appendix are included. OLS(SH). Home inputs at ages 4–5 are included as well as the above school inputs.

on watching TV (up to 6 hours daily) is associated with higher math scores, while 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 6–7 and 8–9 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 Granger causality establishing that at least there are direct effects of TV watching on the scores, the orderly thing to do is to proceed further to obtain the full dynamic effects of TV watching including the indirect effect that is missed in Granger causality. Before that, however, in the next subsection we will present some estimation results for typical econometric panel data methods.

5.2. Fixed Effect with IV

In Table III, columns 'FE(IV)' present the results based on model (17), where home inputs at ages 4–5 are used as instruments for the current home inputs at ages 8–9; as in Table II, squared treatment variables are used. As noted already, (18) may be regarded as (17) with the restriction that the lagged variables have opposite signs and the same magnitude. This proposition is supported partly by the different signs of d_1 and d_2 and their squared versions in Tables II and III.

Table III allows for endogeneity of the current home inputs, which are the G2 group of variables in the Appendix, including daily TV viewing hours, whether the child has more than 10 books, whether the child's mother reads to him/her at least three times a week, whether the child reads every day for enjoyment, whether the child's family encourages hobbies, whether the child is taken to a museum or performance at least several times a year, whether the child's family gets together with relatives or friends at least two or three times a month, whether the child does outdoor activities with the father at least once a week, whether the child eats with parents at least once a day, and whether parents discuss television programs with the child. The instruments for these variables are the home inputs at ages 4–5, which are listed in the Appendix as the G4 group of variables and are very similar to the home inputs at ages 8–9.

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

Columns 'DF' and 'DF(S)' are based on the first-difference model (18); they differ only in terms of control variables, where school inputs at ages 6–7 and 8–9 are further controlled in 'DF(S)'. For PIAT math scores at ages 8–9, TV watching time at ages 8–9 has a negative and convex effect on math scores under both specifications, which leads to an overall negative effect between TV hours and math scores, as shown in the lower half of the table. For reading scores at ages 8–9, TV watching time 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 watching is best when

Table III. TV watching on math and reading: fixed-effect model

	PIAT Math			PIAT Reading		
	At ages 8–9	Difference b/w ages 8–9 and 6–7		At ages 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		–0.18 (0.50)	–1.41 (1.26)		0.084 (0.39)	1.38 (0.90)
Difference in squared daily TV hours b/w ages 8–9 and 6–7		0.02 (0.05)	0.14 (0.13)		–0.015 (0.04)	–0.17* (0.09)
Daily TV hours at ages 8–9	–7.90 (11.3)			–1.89 (8.28)		
Daily TV hours at ages 8–9 squared	0.61 (1.68)			0.14 (1.15)		
Daily TV hours at ages 6–7	5.76 (5.41)			0.92 (3.68)		
Daily TV hours at ages 6–7 squared	–0.55 (0.46)			–.08 (0.34)		
PIAT score at ages 6–7	0.62*** (0.12)			0.72*** (0.10)		
Sample size	1050	1555	299	1043	1539	299
R ²	—	0.04	0.22	—	0.06	0.27

Mean difference in PIAT scores (bootstrapped SD in parentheses)

Daily TV hours at ages 6–9	PIAT Math at ages 8–9			PIAT Reading at ages 8–9		
	FE(IV)	DF	DF(S)	FE(IV)	DF	DF(S)
1 hour versus 0 hour	–2.08 (13.4)	–0.16 (0.47)	–1.27 (1.22)	–0.91 (24.1)	0.07 (.36)	1.21 (0.98)
2 hours versus 0 hour	–4.03 (22.8)	–0.28 (0.84)	–2.26 (2.19)	–1.70 (43.2)	0.11 (.64)	2.10 (1.77)
3 hours versus 0 hour	–5.86 (28.1)	–0.35 (1.11)	–2.96 (2.90)	–2.37 (57.3)	0.12 (.85)	2.65 (2.37)
4 hours versus 0 hour	–7.57 (29.7)	–0.38 (1.29)	–3.38 (3.36)	–2.91 (66.8)	0.10 (1.0)	2.88 (2.80)
5 hours versus 0 hour	–9.15 (28.1)	–0.37 (1.39)	–3.51 (3.61)	–3.33 (71.8)	0.05 (1.08)	2.77 (3.07)
6 hours versus 0 hour	–10.6 (25.2)	–0.31 (1.41)	–3.37 (3.68)	–3.63 (72.8)	–0.02 (1.12)	2.34 (3.22)
4 hours versus 2 hours	–3.53 (8.41)	–0.10 (0.47)	–1.12 (1.31)	–1.21 (24.3)	–0.13 (.57)	0.24 (1.65)
6 hours versus 2 hours	–6.57 (16.8)	–0.03 (0.65)	–1.11 (1.92)	–1.93 (33.5)	–0.01 (0.36)	0.78 (1.04)

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard deviations are in the parentheses. The sample is composed of children watching 10 hours or less TV per day between ages 6 and 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.

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 for math scores between the negative TV effects in the first-difference results and the positive effects under various OLS specifications in Table II, suggesting that the unobserved child-specific factor may matter more for math scores than for reading scores.

The overall results of these fixed-effect models with IVE are not significantly different from the OLS specifications for Table II, possibly because many detailed home inputs and family backgrounds are already controlled. This finding suggests that we may proceed to G-estimation and the two-stage procedure without worrying too much about the potential biases caused by the unobserved child-specific fixed effect or the endogeneity of the current inputs.

5.3. Two-Stage Procedure with Linear Models

Table IV presents results for the two-stage procedure with the linear model (19). 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 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 scores and just negative for reading scores. As shown in the lower half of Table IV, however, the overall effects of TV watching across ages 6–7 and 8–9 on math and reading scores at ages 8–9 in 'OLS(S)' are negative for math scores and concave for reading scores, where the best outcome is achieved at 2 hours in column 'TSE(S)', where the school inputs are controlled (and at 6 hours in column 'TSE'), where TSE stands for 'two-stage estimator'. Overall, however, the magnitudes of most TV effects are quite small and statistically insignificant. For example, based on estimates in columns 'TSE(S)', watching 2 hours TV per day during ages 6–7 and 8–9 reduces the math score at ages 8–9 by 4% of one standard deviation, but increases reading score at ages 8–9 by 3% of one standard deviation, compared with not watching TV at all.

To get a sense of the relative importance of the indirect effect of TV watching at ages 6–7 on the math and reading scores 2 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 1-hour TV watching at ages 6–7 on the math score at ages 6–7 is $(-0.74 + 0.01) = -0.73$, which leads to an indirect effect on the math score at ages 8–9, $-0.73 \times 0.64 = -0.47$, while its direct effect is $(-0.06 + 0.02) = -0.04$. The effect of 1-hour TV viewing at ages 8–9 is $(0.20 + 0.002) = 0.202$. Thus the indirect effect of earlier TV watching at ages 6–7 (which is -0.47) is almost three times the total direct effect of TV watching at both ages 6–7 and 8–9 (which is $0.162 = 0.202 - 0.04$), and makes the total effect of 1-hour TV watching at ages 6–7 and 8–9 negative (-0.30 , as reported at the first row in column 'TSE(S)' for math), despite the positive effect of current TV viewing at ages 8–9.

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 6–7 and 8–9 on reading scores at ages 8–9 (which is -1.78 , as reported in the last row in column 'TSE(S)' for reading) can be decomposed into three components: the indirect effect of TV viewing time at ages 6–7 is $[(-0.49 \times 6 - 0.01 \times 36) - (-0.49 \times 2 - 0.01 \times 4)] \times 0.78 = -1.78$, the direct effect of d_1 and d_1^2 is $(-0.47 \times 6 + 0.02 \times 36) - (-0.47 \times 2 + 0.02 \times 4) = -1.24$, while the direct effect of d_2 and d_2^2 is $(1.27 \times 6 - 0.12 \times 36) - (1.27 \times 2 - 0.12 \times 4) = 1.24$. Again, the indirect effect of TV viewing at ages 6–7 dominates the total direct effect of TV viewing at ages 6–7 and 8–9, and outweighs the positive effect of the current TV watching.

Table IV. TV watching on Math and Reading: two-stage with linear model

	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 ages 8–9	–0.42 (0.54)		0.20 (0.77)		0.37 (0.48)		1.27* (0.68)	
Daily TV hours at ages 8–9 squared	0.06 (0.06)		0.002 (0.08)		–0.04 (0.05)		–0.12* (0.07)	
Daily TV hours at ages 6–7	0.13 (0.50)	–0.27 (0.44)	–0.06 (0.76)	–0.74	0.18 (0.44)	–0.11 (0.43)	–0.47 (0.67)	–0.49 (0.97)
Daily TV hours at ages 6–7 squared	–0.01 (0.05)	0.02 (0.05)	0.02 (0.08)	0.01 (0.12)	–0.01 (0.05)	–0.01 (0.05)	0.02 (0.08)	–0.01 (0.10)
PIAT score at ages 6–7	0.60*** (0.03)		0.64*** (0.04)		0.75*** (0.03)		0.78*** (0.04)	
Sample size	1594	1555	663	299	1578	1539	661	299
R ²	0.46	0.25	0.50	0.42	0.54	0.22	0.57	0.35

Mean difference in PIAT scores (bootstrapped SD in parentheses)

Daily TV hours at ages 6–9	PIAT Math		PIAT Reading	
	TSE	TSE(S)	TSE	TSE(S)
1 hour versus 0 hour	–0.38 (0.57)	–0.30 (1.25)	0.42 (0.54)	0.31 (1.23)
2 hours versus 0 hour	–0.64 (1.0)	–0.54 (2.20)	0.76 (0.96)	0.40 (2.18)
3 hours versus 0 hour	–0.76 (1.31)	–0.73 (2.84)	1.02 (1.25)	0.28 (2.85)
4 hours versus 0 hour	–0.75 (1.50)	–0.86 (3.23)	1.19 (1.42)	–0.06 (3.29)
5 hours versus 0 hour	–0.62 (1.58)	–0.94 (3.41)	1.27 (1.50)	–0.62 (3.54)
6 hours versus 0 hour	–0.35 (1.61)	–0.96 (3.51)	1.28 (1.53)	–1.39 (3.68)
4 hours versus 2 hours	–0.12 (0.54)	–0.32 (1.17)	0.43 (0.51)	–0.46 (1.23)
6 hours versus 2 hours	–0.37 (0.95)	–0.56 (2.05)	0.52 (0.85)	–1.78 (2.15)

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard deviations are in parentheses. The sample is composed of children watching 10 hours or less TV per day between ages 6 and 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.

5.4. G-Algorithm with Discrete Responses

The insignificance of the total TV effects in Table IV might be due to measurement errors, since few parents can recall with much accuracy the exact hours their children watch TV every day. It is quite plausible, however, that most parents know the time range their children spend in front of the TV. This prompts us to apply 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 used 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 6–7 and 8–9, home environment indicators at ages 4–5, child demographic

information and family background variables, while detailed home inputs at ages 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 V, where the entries are the estimated marginal effects calculated at the sample means of the control variables (i.e., the derivatives of $P(y_2 = 1 | \dots)$ evaluated at the 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 ages 8–9 is significantly positive in ‘Probit (H)’. There seems to be some nonlinear effects, but most individual estimates are insignificant.

G-estimation using the above probit regressions is presented in the lower half of Table V. 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 models. Specifically, a child with High TV at ages 6–7 and 8–9 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 6–7 and 8–9 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 V, the results in the first column (‘G-est.’) show 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 ages 8–9; the corresponding numbers (unreported in Table V) are 18% and 16%, respectively, when 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 2 years earlier. More or less the same statements can be made for reading scores.

6. CONCLUSIONS

When multiple treatments are given over time with 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, developed a two-stage procedure, and applied a discretized version of the G-algorithm as well as the two-stage procedure to an important issue: the effect

Table V. TV watching on math and reading: G-estimation with discrete responses
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 ages 8–9	0.05 (0.06)		–0.006 (0.07)		0.06 (0.05)		0.10* (0.05)	
Middle TV at ages 8–9	–0.007 (0.06)		–0.07 (0.06)		–0.02 (0.05)		0.01 (0.05)	
High TV at ages 6–7	0.05 (0.06)	–0.05 (0.06)	0.06 (0.06)	–0.04 (0.06)	–0.07* (0.06)	0.01 (0.06)	–0.07 (0.06)	0.01 (0.06)
Middle TV at ages 6–7	–0.03 (0.05)	–0.04 (0.05)	–0.002 (0.06)	–0.05 (0.05)	–0.004 (0.05)	–0.001 (0.05)	–0.01 (0.05)	0.001 (0.05)
PIAT score higher than mean at ages 6–7	0.37*** (0.04)		0.40*** (0.04)		0.51*** (0.04)		0.51*** (0.05)	
PIAT score higher than mean at ages 4–5	0.03 (0.04)	0.31*** (0.04)	0.04 (0.04)	0.26*** (0.04)	0.14*** (0.05)	0.32*** (0.05)	0.13** (0.05)	0.29*** (0.05)
Sample size	720	720	651	651	687	687	622	622
Pseudo R ²	0.29	0.23	0.31	0.22	0.40	0.26	0.41	0.23
Mean difference in probability of PIAT score higher than mean at ages 8–9 (bootstrapped SD in parentheses)								
Daily TV hours at ages 6–9	PIAT Math				PIAT Reading			
	G-est.		G-est. (H)		G-est.		G-est. (H)	
High versus Low TV at ages 6–9	–0.18** (0.08)		–0.23*** (0.08)		–0.21*** (0.08)		–0.23*** (0.09)	
Middle versus Low TV at ages 6–9	–0.16** (0.07)		–0.18** (0.07)		–0.13* (0.07)		–0.14* (0.08)	
High versus Low TV at ages 6–7 with low at 8–9	–0.23* (0.13)		–0.19 (0.13)		–0.23* (0.12)		–0.26* (0.14)	
Middle versus Low TV at ages 6–7 with low at 8–9	–0.19** (0.08)		–0.20** (0.08)		–0.11 (0.08)		–0.13 (0.08)	

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. Standard deviations are in 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 children watching 10 hours or less TV per day between ages 6 and 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 ages 4–5 are included.

of watching TV on child cognitive development measured by math and reading scores. For math score, the two-stage results indicated that watching TV for any amount of time during ages 6–7 and 8–9 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

these estimators as well as typical panel data econometric approaches using continuous response variables led to coherent evidence that around 2 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.

ACKNOWLEDGEMENTS

The authors are grateful to the anonymous referees and the co-editor Manuel Arellano for their detailed comments.

REFERENCES

- Abbring JH, Heckman JJ. 2007. Econometric evaluation of social programs. Part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In *Handbook of Econometrics*, Vol. 6B, Heckman JJ, Leamer EE (eds). North-Holland: Amsterdam; 5145–5303.
- Anderson DR. 2005. Television and very young children. *American Behavioral Scientist* **48**: 505–522.
- Angrist JD, Krueger AB. 2001. Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives* **15**: 69–85.
- Chernin AR, Linebarger DL. 2005. The relationship between children's television viewing and academic performance. *Archives of Pediatrics and Adolescent Medicine* **159**: 687–689.
- Dawid AP. 2000. Causal inference without counterfactuals. *Journal of the American Statistical Association* **95**: 407–424.
- Genzckow M, Shapiro JM. 2008. Preschool television viewing and adolescent test scores: historical evidence from the Coleman study. *Quarterly Journal of Economics* **123**: 279–323.
- Gill RD, Robins JM. 2001. Causal inference for complex longitudinal data: the continuous case. *Annals of Statistics* **29**: 1785–1811.
- Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**: 424–438.
- Granger CWJ. 1980. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control* **2**: 329–352.
- Heckman JJ. 2000. Policies to foster human capital. *Research in Economics* **54**: 3–56.
- Holland PW. 1986. Statistics and causal inference. *Journal of the American Statistical Association* **81**: 945–960.
- Holtz-Eakin D, Newey W, Rosen HS. 1988. Estimating vector autoregressions with panel data. *Econometrica* **56**: 1371–1395.
- Holtz-Eakin D, Newey W, Rosen HS. 1989. The revenue–expenditure nexus: evidence from local government data. *International Economic Review* **30**: 415–429.
- Juster FT, Stafford FP. 1991. The allocation of time: empirical findings, behavioral models, and problems of measurement. *Journal of Economic Literature* **29**: 471–522.
- Keane M, Wolpin K. 1997. Career decisions of young men. *Journal of Political Economy* **105**: 473–522.
- Lechner M. 2004. *Sequential matching estimation of dynamic causal models*. Discussion paper, University of St Gallen.
- Lechner M. 2006. *The relation of different concepts of causality in econometrics*. Discussion paper, University of St Gallen.

- Lechner M. 2008. Matching estimating of dynamic treatment models: some practical issues. In *Advances in Econometrics*, Vol. 21, Millimet D, Smith J, Vytlačil E (eds). JAI Press: Greenwich, CT; 289–333.
- Lechner M, Miquel R. 2001. *A potential outcome approach to dynamic programme evaluation. Part I: Identification*. Discussion paper 2001-07, Department of Economics, University of St Gallen.
- Lee MJ. 2002. *Panel Data Econometrics: Methods-of-Moments and Limited Dependent Variables*. Academic Press: New York.
- Lee MJ, Kobayashi S. 2001. Proportional treatment effects for count response panel data: effects of binary exercise on health care demand. *Health Economics* **10**: 411–428.
- Pecora N, Murray JP, Wartella EA. 2006. *Children and Television: Fifty Years of Research*. Lawrence Erlbaum Associates: Philadelphia, PA.
- Rideout V, Roberts DF, Foehr UG. 2005. *Generation M: media in the lives of 8–18 year-olds*. A Kiser Family Foundation Study. www.kff.org [22 January 2010].
- Robins JM. 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 JM. 1998. Structural nested failure time models. In *Survival Analysis*, Vol. 6, *Encyclopedia of Biostatistics*, Armitage P, Colton T (eds). Wiley: Chichester; 4372–4389.
- Robins JM. 1999. Marginal structural models versus structural nested models as tools for causal inference. In *Statistical Models in Epidemiology: The Environment and Clinical Trials*, Halloran ME, Berry D (eds). Springer: Berlin; 95–134.
- Robins JM, Greenland S, Hu FC. 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 PR, Rubin DB. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**: 41–55.
- Van Evra JP. 2004. *Television and Child Development*. Lawrence Erlbaum Associates: Philadelphia, PA.
- Zavodny M. 2006. Does watching television rot your mind? Estimates of the effect on test scores. *Economics of Education Review* **25**: 565–573.
- Zimmerman FJ, Christakis DA. 2005. Children's television viewing and cognitive outcomes: a longitudinal analysis of national data. *Archives of Pediatrics and Adolescent Medicine* **159**: 619–625.

APPENDIX: LIST OF VARIABLES AND SUMMARY STATISTICS

Variable	Mean (SD)	Size
PIAT Math at ages 8–9	102.2 (13.9)	2180
PIAT Math at ages 6–7	100.8 (12.1)	2180
PIAT Math at ages 4–5	99.2 (14.7)	1250
PIAT Reading at ages 8–9	105.0 (14.6)	2180
PIAT Reading at ages 6–7	103.3 (11.7)	2180
PIAT Reading at ages 4–5	106.5 (14.8)	1175
<i>(G1) Child demographic information</i>		
Race of child: Black or Hispanic	0.47 (0.5)	2180
Sex of child: boy	0.50 (0.5)	2180
Birth order of child	2.0 (1.0)	2180
<i>(G2) Home inputs for 8- to 9-year-olds</i>		
Number of hours per day child sees TV	3.49 (2.03)	2180
Child has 10 or more children's books at home	0.88 (0.32)	2173
How often mother reads to child: at least 3 times per week	0.59 (0.49)	2172
How often child reads for enjoyment: every day	0.35 (0.48)	2174
Family encourages hobbies	0.93 (0.25)	2169
How often child taken to museum: at least several times per year	0.43 (0.50)	2173
How often child taken to performance: at least several times per year	0.35 (0.48)	2172
How often family gets with relatives/friends: at least 2–3 times per month	0.59 (0.49)	2177
How often child with father outdoors: at least once per week	0.49 (0.50)	2083
How often child eats with parents: at least once per day	0.57 (0.50)	2081
Parents discuss TV programs with child	0.84 (0.37)	2155
<i>(G3) Home inputs for 6- to 7-year-olds</i>		
Number of hours a day child sees TV	3.35 (2.02)	2180
Child has 10 or more children's books at home	0.86 (0.35)	2175
How often mother reads to child: at least 3 times per week	0.75 (0.43)	2176
How often child reads for enjoyment: every day	0.67 (0.47)	2170
Family encourages hobbies	0.91 (0.29)	2173
How often child taken to museum: at least several times per year	0.78 (0.41)	2177
How often child taken to performance: at least several times per year	0.61 (0.49)	2178
How often family gets with relatives/friends: at least 2–3 times per month	0.61 (0.49)	2180
How often child with father outdoors: at least once per week	0.50 (0.50)	2134
How often child eats with parents: at least once per day	0.78 (0.41)	2173
Parents discuss TV programs with child	0.83 (0.37)	2168
<i>(G4) Home inputs for 4- to 5-year-olds</i>		
Number of hours per weekday child sees TV	3.31 (3.22)	1505
Number of hours per weekend day child sees TV	3.66 (3.19)	1511
Number of hours TV is on per day	5.44 (5.64)	2002
How many books child has: 10 or more books	0.82 (0.39)	2029
How often mother reads to child: at least 3 times per week	0.58 (0.49)	2029
How often child taken to museum: at least several times per year	0.72 (0.45)	2021
How many magazines family gets: 3 or more	0.61 (0.49)	2024
How often child taken on outing: at least several times a week	0.56 (0.50)	2016
how often child eats with mom & dad: at least once a day	0.76 (0.43)	2011
Home Observation Measurement of the Environment variable (HOME)	206.1 (34.3)	2109
<i>(G5) Family background</i>		
Mother's AFQT score taken in 1981	41.3 (28.2)	2117
Mother's highest grade at 1988	12.6 (2.2)	2133
Family salary income in 1988	6895.8 (8414.8)	2112

(Continued)

Variable	Mean (SD)	Size
Mother was married in 1988	0.81 (0.39)	1819
Mother's age at child's birth	25.0 (3.0)	2179
Child is breast-fed	0.52 (0.50)	2085
<i>(G6) School inputs for 8- to 9-year-olds</i>		
Hours per week a child spends on math homework	2.24 (3.39)	970
Hours per week a child spends on writing homework	1.75 (3.06)	899
The school is public	0.93 (0.25)	951
Mother's rating of teacher caring—high	0.60 (0.49)	994
Mother's rating of principal as leader—high	0.51 (0.50)	991
Mother's rating of teacher skill—high	0.52 (0.50)	990
Mother's rating of safety of school—high	0.64 (0.48)	993
Mother's rating of school communicating with parents—high	0.57 (0.49)	994
Mother's rating of parents participating with school—high	0.32 (0.47)	991
Mother's rating of school teaching right and wrong—high	0.54 (0.50)	994
Mother's rating of school maintaining order—high	0.56 (0.50)	994
Mother's rating of teacher caring—middle	0.31 (0.46)	994
Mother's rating of principal as leader—middle	0.33 (0.47)	991
Mother's rating of teacher skill—middle	0.38 (0.49)	990
Mother's rating of safety of school—middle	0.28 (0.45)	993
Mother's rating of school communicating with parents—middle	0.31 (0.46)	994
Mother's rating of parents participating with school—middle	0.37 (0.48)	991
Mother's rating of school teaching right and wrong—middle	0.34 (0.47)	994
Mother's rating of school maintaining order—middle	0.32 (0.47)	994
<i>(G7) School inputs for 6- to 7-year-olds</i>		
Hours per week a child spends on math homework	1.15 (2.30)	414
Hours per week a child spends on writing homework	0.79 (1.62)	356
The school is public	0.95 (0.22)	450
Mother's rating of teacher caring—high	0.65 (0.48)	484
Mother's rating of principal as leader—high	0.55 (0.50)	485
Mother's rating of teacher skill—high	0.55 (0.50)	482
Mother's rating of safety of school—high	0.63 (0.48)	486
Mother's rating of school communicating with parents—high	0.58 (0.49)	484
Mother's rating of parents participating with school—high	0.31 (0.46)	485
Mother's rating of school teaching right and wrong—high	0.58 (0.49)	486
Mother's rating of school maintaining order—high	0.60 (0.49)	484
Mother's rating of teacher caring—middle	0.25 (0.43)	484
Mother's rating of principal as leader—middle	0.27 (0.45)	485
Mother's rating of teacher skill—middle	0.34 (0.48)	482
Mother's rating of safety of school—middle	0.26 (0.44)	486
Mother's rating of school communicating with parents—middle	0.29 (0.46)	484
Mother's rating of parents participating with school—middle	0.41 (0.49)	485
Mother's rating of school teaching right and wrong—middle	0.30 (0.46)	486
Mother's rating of school maintaining order—middle	0.29 (0.45)	484