

Estimating Sibling Spillover Effects with Unobserved
Confounding Using Gain-Scores
Appendix

David C. Mallinson & Felix Elwert

Contents

1	Overview	2
2	Models with One-Sided Spillover	3
3	Models with Two-Sided Spillover	6
4	Models with Spillovers from Outcomes	8
5	Data Assembly for the Empirical Application	11
6	Appendix Tables	12
7	Simulation Code	14
	References	33

1 Overview

Consider linear data generating models, represented by **Figures 1-3** in the main text, with exposures T_{ij} , outcomes Y_{ij} , and unobserved confounders U_i for $i = 1, \dots, N$ families consisting of $j = 1, 2$ siblings. Write $D_i = Y_{i2} - Y_{i1}$ for the gain-score. Henceforth, we suppress i for readability.

To estimate spillover effects from the exposure of one sibling onto the other sibling's outcome in these models, we use gain-score regression,

$$D = b_1 T_1 + b_2 T_2 + e,$$

where b_1 and b_2 are partial regression coefficients for T_1 and T_2 , respectively, e is the residual, and $SC = b_1 + b_2$ is the spillover coefficient. The partial regression coefficients are given by,

$$b_1 = \frac{\sigma_{DT_1} - \sigma_{DT_2} * \sigma_{T_1 T_2}}{1 - \sigma_{T_1 T_2}^2},$$

$$b_2 = \frac{\sigma_{DT_2} - \sigma_{DT_1} * \sigma_{T_1 T_2}}{1 - \sigma_{T_1 T_2}^2},$$

where σ denotes covariance. We link the partial regression coefficients to the data generating models of **Figures 1-3** using Wright's path rules, where the covariance between two variables is equal to the sum of the products of path parameters on the open paths between the two variables. We assume standardized variables (mean zero and unit variance) without loss of generality. The identification results in the following sections hold for both continuous and binary exposures. See Pearl (2013) for details [1].

2 Models with One-Sided Spillover

Base model with one-sided spillover (Figure 1A)

We first consider models with one-sided exposure-to-outcome spillover ($T_1 \rightarrow Y_2$, but T_2 does not directly affect Y_1). Under the data generating process in **Figure 1A**, we point identify the spillover effect θ . We compute σ_{DT_1} , σ_{DT_2} , and $\sigma_{T_1T_2}$:

$$\begin{aligned}\sigma_{DT_1} &= -\psi\chi - \delta + \theta + \delta\chi\gamma + \psi\chi \\ &= -\delta + \theta + \delta\chi\gamma\end{aligned}$$

$$\begin{aligned}\sigma_{DT_2} &= -\psi\gamma - \delta\chi\gamma + \theta\chi\gamma + \delta + \psi\gamma \\ &= -\delta\chi\gamma + \theta\chi\gamma + \delta\end{aligned}$$

$$\sigma_{T_1T_2} = \chi\gamma$$

We plug these covariances into the formulas for our partial regression coefficients:

$$\begin{aligned}b_1 &= \frac{-\delta + \theta + \delta\chi\gamma - (-\delta\chi\gamma + \theta\chi\gamma + \delta) * (\chi\gamma)}{1 - (\chi\gamma)^2} \\ &= \frac{-\delta + \theta + \delta\chi\gamma - (-\delta\chi^2\gamma^2 + \theta\chi^2\gamma^2 + \delta\chi\gamma)}{1 - \chi^2\gamma^2} \\ &= \frac{-\delta + \theta + \delta\chi^2\gamma^2 - \theta\chi^2\gamma^2}{1 - \chi^2\gamma^2} \\ &= \frac{-\delta(1 - \chi^2\gamma^2) + \theta(1 - \chi^2\gamma^2)}{1 - \chi^2\gamma^2} \\ &= \theta - \delta\end{aligned}$$

$$\begin{aligned}b_2 &= \frac{-\delta\chi\gamma + \theta\chi\gamma + \delta - (-\delta + \theta + \delta\chi\gamma) * (\chi\gamma)}{1 - (\chi\gamma)^2} \\ &= \frac{-\delta\chi\gamma + \theta\chi\gamma + \delta - (-\delta\chi\gamma + \theta\chi\gamma + \delta\chi^2\gamma^2)}{1 - \chi^2\gamma^2} \\ &= \frac{\delta - \delta\chi^2\gamma^2}{1 - \chi^2\gamma^2} \\ &= \frac{\delta(1 - \chi^2\gamma^2)}{1 - \chi^2\gamma^2} \\ &= \delta\end{aligned}$$

We then compute SC , which point identifies θ :

$$\begin{aligned} SC &= b_1 + b_2 \\ &= \theta - \delta + \delta \\ &= \theta \end{aligned}$$

One-sided spillover and exposure-to-exposure spillover (Figures 1B-1C)

Under the data generating process in **Figure 1B**, sibling 2's exposure affects sibling 1's exposure ($T_2 \rightarrow T_1$). Regardless, we still point identify the spillover effect θ . We compute σ_{DT_1} , σ_{DT_2} , and $\sigma_{T_1T_2}$:

$$\begin{aligned} \sigma_{DT_1} &= -\psi\chi - \psi\gamma\tau - \delta + \theta + \delta\tau + \delta\chi\gamma + \psi\gamma\tau + \psi\chi \\ &= -\delta + \theta + \delta\tau + \delta\chi\gamma \end{aligned}$$

$$\begin{aligned} \sigma_{DT_2} &= -\psi\gamma - \delta\chi\gamma - \delta\tau + \theta\chi\gamma + \theta\tau + \delta + \psi\gamma \\ &= -\delta\chi\gamma - \delta\tau + \theta\chi\gamma + \theta\tau + \delta \end{aligned}$$

$$\sigma_{T_1T_2} = \chi\gamma + \tau$$

We then compute b_1 and b_2 :

$$\begin{aligned} b_1 &= \frac{-\delta + \theta + \delta\tau + \delta\chi\gamma - (-\delta\chi\gamma - \delta\tau + \theta\chi\gamma + \theta\tau + \delta) * (\chi\gamma + \tau)}{1 - (\chi\gamma + \tau)^2} \\ &= \theta - \delta \end{aligned}$$

$$\begin{aligned} b_2 &= \frac{-\delta\chi\gamma - \delta\tau + \theta\chi\gamma + \theta\tau + \delta - (-\delta + \theta + \delta\tau + \delta\chi\gamma) * (\chi\gamma + \tau)}{1 - (\chi\gamma + \tau)^2} \\ &= \delta \end{aligned}$$

Finally, we compute SC :

$$\begin{aligned} SC &= b_1 + b_2 \\ &= \theta - \delta + \delta \\ &= \theta \end{aligned}$$

The data generating process in **Figure 1C** has exposure-to-exposure spillover, where sibling 1's exposure affects sibling 2's exposure ($T_1 \rightarrow T_2$). Still, we point identify the spillover effect θ . We compute σ_{DT_1} , σ_{DT_2} ,

and $\sigma_{T_1T_2}$:

$$\begin{aligned}\sigma_{DT_1} &= -\psi\chi - \delta + \theta + \delta\phi + \delta\chi\gamma + \psi\chi \\ &= -\delta + \theta + \delta\phi + \delta\chi\gamma\end{aligned}$$

$$\begin{aligned}\sigma_{DT_2} &= -\psi\gamma - \psi\chi\phi - \delta\chi\gamma - \delta\phi + \theta\chi\gamma + \theta\phi + \delta + \psi\chi\phi + \psi\gamma \\ &= -\delta\chi\gamma - \delta\phi + \theta\chi\gamma + \theta\phi + \delta\end{aligned}$$

$$\sigma_{T_1T_2} = \chi\gamma + \phi$$

We then compute b_1 and b_2 :

$$\begin{aligned}b_1 &= \frac{-\delta + \theta + \delta\phi + \delta\chi\gamma - (-\delta\chi\gamma - \delta\phi + \theta\chi\gamma + \theta\phi + \delta) * (\chi\gamma + \phi)}{1 - (\chi\gamma + \phi)^2} \\ &= \theta - \delta\end{aligned}$$

$$\begin{aligned}b_2 &= \frac{-\delta\chi\gamma - \delta\phi + \theta\chi\gamma + \theta\phi + \delta - (-\delta + \theta + \delta\phi + \delta\chi\gamma) * (\chi\gamma + \phi)}{1 - (\chi\gamma + \phi)^2} \\ &= \delta\end{aligned}$$

Finally, we compute SC :

$$\begin{aligned}SC &= b_1 + b_2 \\ &= \theta - \delta + \delta \\ &= \theta\end{aligned}$$

3 Models with Two-Sided Spillover

Base model with two-sided spillover (Figure 2A)

We now consider models with two-sided exposure-to-outcome spillover ($T_1 \rightarrow Y_2$ and $T_2 \rightarrow Y_1$). Under the data generating process in **Figure 2A**, we identify the difference between θ and κ . We compute σ_{DT_1} , σ_{DT_2} , and $\sigma_{T_1T_2}$:

$$\begin{aligned}\sigma_{DT_1} &= -\psi\chi - \delta - \kappa\chi\gamma + \theta + \delta\chi\gamma + \psi\chi \\ &= -\delta - \kappa\chi\gamma + \theta + \delta\chi\gamma\end{aligned}$$

$$\begin{aligned}\sigma_{DT_2} &= -\psi\gamma - \delta\chi\gamma - \kappa + \theta\chi\gamma + \delta + \psi\gamma \\ &= -\delta\chi\gamma - \kappa + \theta\chi\gamma + \delta\end{aligned}$$

$$\sigma_{T_1T_2} = \chi\gamma$$

We then compute b_1 and b_2 :

$$\begin{aligned}b_1 &= \frac{-\delta - \kappa\chi\gamma + \theta + \delta\chi\gamma - (-\delta\chi\gamma - \kappa + \theta\chi\gamma + \delta) * (\chi\gamma)}{1 - (\chi\gamma)^2} \\ &= \theta - \delta\end{aligned}$$

$$\begin{aligned}b_2 &= \frac{-\delta\chi\gamma - \kappa + \theta\chi\gamma + \delta - (-\delta - \kappa\chi\gamma + \theta + \delta\chi\gamma) * (\chi\gamma)}{1 - (\chi\gamma)^2} \\ &= \delta - \kappa\end{aligned}$$

Finally, we compute SC :

$$\begin{aligned}SC &= b_1 + b_2 \\ &= \theta - \delta + \delta - \kappa \\ &= \theta - \kappa\end{aligned}$$

Two-sided spillover and exposure-to-exposure spillover (Figures 2B-2C)

Even with exposure-to-exposure spillover, the spillover coefficient continues to identify the difference between the two exposure-to-outcome spillover effects ($SC = \theta - \kappa$). This is true regardless of the direction of the exposure-to-exposure spillover. In **Figure 2B**, we have exposure-to-exposure spillover from sibling 2 to

sibling 1 ($T_2 \rightarrow T_1$). We compute σ_{DT_1} , σ_{DT_2} , and $\sigma_{T_1T_2}$:

$$\begin{aligned}\sigma_{DT_1} &= -\psi\chi - \psi\gamma\tau - \delta - \kappa\tau - \kappa\chi\gamma + \theta + \delta\tau + \delta\chi\gamma + \psi\gamma\tau + \psi\chi \\ &= -\delta - \kappa\tau - \kappa\chi\gamma + \theta + \delta\tau + \delta\chi\gamma\end{aligned}$$

$$\begin{aligned}\sigma_{DT_2} &= -\psi\gamma - \delta\chi\gamma + \delta\tau - \kappa + \theta\chi\gamma + \theta\tau + \delta + \psi\gamma \\ &= -\delta\chi\gamma + \delta\tau - \kappa + \theta\chi\gamma + \theta\tau + \delta\end{aligned}$$

$$\sigma_{T_1T_2} = \chi\gamma + \tau$$

We then compute b_1 and b_2 :

$$\begin{aligned}b_1 &= \frac{-\delta - \kappa\tau - \kappa\chi\gamma + \theta + \delta\tau + \delta\chi\gamma - (-\delta\chi\gamma + \delta\tau - \kappa + \theta\chi\gamma + \theta\tau + \delta) * (\chi\gamma + \tau)}{1 - (\chi\gamma + \tau)^2} \\ &= \theta - \delta\end{aligned}$$

$$\begin{aligned}b_2 &= \frac{-\delta\chi\gamma + \delta\tau - \kappa + \theta\chi\gamma + \theta\tau + \delta - (-\delta - \kappa\tau - \kappa\chi\gamma + \theta + \delta\tau + \delta\chi\gamma) * (\chi\gamma + \tau)}{1 - (\chi\gamma + \tau)^2} \\ &= \delta - \kappa\end{aligned}$$

Finally, we compute SC :

$$\begin{aligned}SC &= b_1 + b_2 \\ &= \theta - \delta + \delta - \kappa \\ &= \theta - \kappa\end{aligned}$$

In **Figure 2C**, we have exposure-to-exposure spillover from sibling 1 to sibling 2 ($T_1 \rightarrow T_2$). Changing the direction of the exposure-to-exposure spillover does not affect the partial regression coefficients nor the spillover coefficient.

4 Models with Spillovers from Outcomes

Past outcome affects future exposure (Figure 3A)

We consider models in which spillovers originate from outcomes. Under the data generating process in **Figure 3A**, we do not identify the spillover effect θ . We compute σ_{DT_1} , σ_{DT_2} , and $\sigma_{T_1T_2}$:

$$\begin{aligned}\sigma_{DT_1} &= -\psi\chi - \delta + \theta + \delta\chi\gamma + \delta^2\omega + \delta\omega\psi\chi + \psi\chi \\ &= -\delta + \theta + \delta\chi\gamma + \delta^2\omega + \delta\omega\psi\chi\end{aligned}$$

$$\begin{aligned}\sigma_{DT_2} &= -\psi\gamma - \psi\chi\delta\omega - \delta\chi\gamma - \omega + \theta\chi\gamma + \theta\delta\omega + \delta + \psi^2\omega + \psi\chi\delta\omega + \psi\gamma \\ &= -\delta\chi\gamma - \omega + \theta\chi\gamma + \theta\delta\omega + \delta + \psi^2\omega\end{aligned}$$

$$\sigma_{T_1T_2} = \chi\gamma + \delta\omega + \omega\psi\chi$$

We then compute b_1 and b_2 :

$$b_1 = \frac{-\delta + \theta + \delta\chi\gamma + \delta^2\omega + \delta\omega\psi\chi - (-\delta\chi\gamma - \omega + \theta\chi\gamma + \theta\delta\omega + \delta + \psi^2\omega) * (\chi\gamma + \delta\omega + \omega\psi\chi)}{1 - (\chi\gamma + \delta\omega + \omega\psi\chi)^2}$$

$$b_2 = \frac{-\delta\chi\gamma - \omega + \theta\chi\gamma + \theta\delta\omega + \delta + \psi^2\omega - (-\delta + \theta + \delta\chi\gamma + \delta^2\omega + \delta\omega\psi\chi) * (\chi\gamma + \delta\omega + \omega\psi\chi)}{1 - (\chi\gamma + \delta\omega + \omega\psi\chi)^2}$$

Finally, we compute SC :

$$\begin{aligned}SC &= b_1 + b_2 \\ &= \frac{-\delta + \theta + \delta\chi\gamma + \delta^2\omega + \delta\omega\psi\chi - (-\delta\chi\gamma - \omega + \theta\chi\gamma + \theta\delta\omega + \delta + \psi^2\omega) * (\chi\gamma + \delta\omega + \omega\psi\chi)}{1 - (\chi\gamma + \delta\omega + \omega\psi\chi)^2} + \\ &\quad \frac{-\delta\chi\gamma - \omega + \theta\chi\gamma + \theta\delta\omega + \delta + \psi^2\omega - (-\delta + \theta + \delta\chi\gamma + \delta^2\omega + \delta\omega\psi\chi) * (\chi\gamma + \delta\omega + \omega\psi\chi)}{1 - (\chi\gamma + \delta\omega + \omega\psi\chi)^2} \\ &= \frac{\theta + \delta^2\omega + \delta\omega\psi\chi - \omega + \theta\chi\gamma + \theta\delta\omega + \psi^2\omega - (-\omega + \theta\chi\gamma + \theta\delta\omega + \psi^2\omega + \theta + \delta^2\omega + \delta\omega\psi\chi) * (\chi\gamma + \delta\omega + \omega\psi\chi)}{1 - (\chi\gamma + \delta\omega + \omega\psi\chi)^2}\end{aligned}$$

In this data generating process, computing $SC = b_1 + b_2$ does not identify θ nor a simple function of θ .

Outcome-to-outcome spillover (Figures 3B-3C)

Under the data generating process in **Figure 3B**, sibling 1's outcome affects sibling 2's outcome ($Y_1 \rightarrow Y_2$), and we do not identify the spillover effect θ . We compute σ_{DT_1} , σ_{DT_2} , and $\sigma_{T_1T_2}$:

$$\begin{aligned}\sigma_{DT_1} &= -\psi\chi - \delta + \eta\psi\gamma + \eta\delta + \theta + \delta\chi\gamma + \psi\chi \\ &= -\delta + \eta\psi\gamma + \eta\delta + \theta + \delta\chi\gamma\end{aligned}$$

$$\begin{aligned}\sigma_{DT_2} &= -\psi\gamma - \delta\chi\gamma + \eta\psi\gamma + \eta\delta\chi\gamma + \theta\chi\gamma + \delta + \psi\gamma \\ &= -\delta\chi\gamma + \eta\psi\gamma + \eta\delta\chi\gamma + \theta\chi\gamma + \delta\end{aligned}$$

$$\sigma_{T_1T_2} = \chi\gamma$$

We then compute b_1 and b_2 :

$$\begin{aligned}b_1 &= \frac{-\delta + \eta\psi\gamma + \eta\delta + \theta + \delta\chi\gamma - (-\delta\chi\gamma + \eta\psi\gamma + \eta\delta\chi\gamma + \theta\chi\gamma + \delta) * (\chi\gamma)}{1 - (\chi\gamma)^2} \\ &= -\delta + \theta + \eta\delta + \frac{\eta\psi\gamma - \eta\psi\chi\gamma^2}{1 - \chi^2\gamma^2}\end{aligned}$$

$$\begin{aligned}b_2 &= \frac{-\delta\chi\gamma + \eta\psi\gamma + \eta\delta\chi\gamma + \theta\chi\gamma + \delta - (-\delta + \eta\psi\gamma + \eta\delta + \theta + \delta\chi\gamma) * (\chi\gamma)}{1 - (\chi\gamma)^2} \\ &= \delta + \frac{\eta\psi\gamma - \eta\psi\chi\gamma^2}{1 - \chi^2\gamma^2}\end{aligned}$$

Finally, we compute SC :

$$\begin{aligned}SC &= b_1 + b_2 \\ &= -\delta + \theta + \eta\delta + \frac{\eta\psi\gamma - \eta\psi\chi\gamma^2}{1 - \chi^2\gamma^2} + \delta + \frac{\eta\psi\gamma - \eta\psi\chi\gamma^2}{1 - \chi^2\gamma^2} \\ &= \theta + \eta\delta + \frac{2\eta\psi\gamma - 2\eta\psi\chi\gamma^2}{1 - \chi^2\gamma^2}\end{aligned}$$

In this data generating process, computing $SC = b_1 + b_2$ does not identify θ nor a simple function of θ .

The data generating process in **Figure 3C** has outcome-to-outcome spillover in which sibling 2's outcome affects sibling 1's outcome ($Y_2 \rightarrow Y_1$). Similarly, we do not identify the spillover effect θ . We compute σ_{DT_1} , σ_{DT_2} , and $\sigma_{T_1T_2}$:

$$\begin{aligned}\sigma_{DT_1} &= -\psi\chi - \delta - \lambda\theta - \lambda\delta\chi\gamma - \lambda\psi\chi + \theta + \delta\chi\gamma + \psi\chi \\ &= -\delta - \lambda\theta - \lambda\delta\chi\gamma - \lambda\psi\chi + \theta + \delta\chi\gamma\end{aligned}$$

$$\begin{aligned}
\sigma_{DT_2} &= -\psi\gamma - \delta\chi\gamma - \lambda\theta\chi\gamma - \lambda\delta + \theta\chi\gamma + \delta + \psi\gamma \\
&= -\delta\chi\gamma - \lambda\theta\chi\gamma - \lambda\delta + \theta\chi\gamma + \delta
\end{aligned}$$

$$\sigma_{T_1T_2} = \chi\gamma$$

We then compute b_1 and b_2 :

$$\begin{aligned}
b_1 &= \frac{-\delta - \lambda\theta - \lambda\delta\chi\gamma - \lambda\psi\chi + \theta + \delta\chi\gamma - (-\delta\chi\gamma - \lambda\theta\chi\gamma - \lambda\delta + \theta\chi\gamma + \delta) * (\chi\gamma)}{1 - (\chi\gamma)^2} \\
&= -\delta - \lambda\theta + \theta - \frac{\lambda\psi\chi}{1 - \chi^2\gamma^2}
\end{aligned}$$

$$\begin{aligned}
b_2 &= \frac{-\delta\chi\gamma - \lambda\theta\chi\gamma - \lambda\delta + \theta\chi\gamma + \delta - (-\delta - \lambda\theta - \lambda\delta\chi\gamma - \lambda\psi\chi + \theta + \delta\chi\gamma) * (\chi\gamma)}{1 - (\chi\gamma)^2} \\
&= \delta + \lambda\delta + \frac{\lambda\psi\chi^2\gamma}{1 - \chi^2\gamma^2}
\end{aligned}$$

Finally, we compute SC :

$$\begin{aligned}
SC &= b_1 + b_2 \\
&= -\delta - \lambda\theta + \theta - \frac{\lambda\psi\chi}{1 - \chi^2\gamma^2} + \delta + \lambda\delta + \frac{\lambda\psi\chi^2\gamma}{1 - \chi^2\gamma^2} \\
&= \theta - \lambda\theta + \frac{\lambda\psi\chi^2\gamma - \lambda\psi\chi}{1 - \chi^2\gamma^2}
\end{aligned}$$

In this data generating process, computing $SC = b_1 + b_2$ does not identify θ nor a simple function of θ .

5 Data Assembly for the Empirical Application

For our application, we estimated the effect of a younger sibling’s preterm birth (gestational age <37 completed weeks) on their older sibling’s Phonological Awareness Literacy Screening-Kindergarten (PALS-K) test score. PALS-K evaluates fundamental literacy skills at kindergarten entry [2]. We used data from Big Data for Little Kids (BD4LK), a cohort of birth records for live resident in-state deliveries in Wisconsin during 2007-2016 ($N > 666,000$ births). Birth records link to multiple administrative sources, including paid Medicaid claims and encounters (2007-2016) and children’s PALS-K scores (2012-2016 school years). BD4LK’s linkage process has been previously described [3, 4].

We sampled sibling pairs born sequentially to the same mother between January 1, 2007-September 1, 2010 and took the English-language PALS-K test. We restricted eligibility on birthdate because children had to be at least five years old by September 1, 2015 for PALS-K testing eligibility [5]. PALS-K is available in two languages: English and Spanish [2, 6]. However, test versions were developed separately and are not directly comparable, so we only considered children who took the English-language test.

BD4LK includes 252,883 unique deliveries during January 1, 2007-September 1, 2010. We identified 806 records (0.3%) with multiple maternal or child identifiers that imperfectly matched to Medicaid claims or PALS-K scores. Among those records, we randomly selected one match and excluded the remainder. We then identified 177,863 records (70.3%) that linked to PALS-K scores, of which 46,743 records were in-sample siblings (26.3% of test-linked records). Finally, we pulled eligible sibling pairs: sequentially-born from different deliveries, took the English-language PALS-K test, and completed information on control variables (maternal age, maternal education, and Medicaid delivery coverage, all of which were measured at the older sibling’s delivery). This generated a final sample of 40,020 siblings (85.6% of tested siblings), or 20,010 sibling pairs.

6 Appendix Tables

Table A.1. Descriptive statistics of sibling pairs for the empirical application (N = 20,010 sibling pairs^a)

	Older Siblings (N = 20,010)	Younger Siblings (N = 20,010)
PALS-K Score (Points) ^c , Mean (SD)	63.58 (24.12)	64.22 (23.83)
Preterm Birth ^b , N (%)	1,357 (6.78%)	1,331 (6.65%)
Maternal Age (Years) ^d , Mean (SD)	26.01 (5.12)	–
Maternal Education ^d , N (%)		
<i>No high school degree</i>	2,865 (14.32%)	–
<i>High school degree/equivalent only</i>	5,789 (28.93%)	–
<i>1-3 years college</i>	5,036 (25.17%)	–
<i>4+ years college</i>	6,320 (31.58%)	–
Medicaid-Paid Delivery ^d , N (%)	7,528 (37.62%)	–

^aEach sibling pair includes one older sibling and one younger sibling. The sample consists of 40,020 children in total.

^bPreterm birth is defined as gestational age <37 completed weeks.

^cPALS-K has a score range of 0-102 points.

^dMeasured at the older sibling's delivery.

Notes: "PALS-K" Phonological Awareness Literacy Screening-Kindergarten, "SD" standard deviation.

Table A.2. Cross-tabulation of preterm birth^a by sibling within two-sibling clusters (N = 20,010 pairs^b)

	Younger Sibling Not Preterm	Younger Sibling Preterm	Total
Older Sibling Not Preterm	17,652 (94.63%) (94.50%)	1,001 (5.37%) (74.21%)	18,653 (100.00%) (93.22%)
Older Sibling Preterm	1,027 (75.68%) (5.50%)	330 (24.32%) (24.79%)	1,357 (100.00%) (6.78%)
Total	18,679 (93.35%) (100.00%)	1,331 (6.65%) (100.00%)	20,010 (100.00%) (100.00%)

^aPreterm birth is defined as gestational age <37 completed weeks.

^bEach sibling pair includes one older sibling and one younger sibling. The sample consists of 40,020 children in total.

Notes: Whole numbers indicate the frequency of sibling pairs. Within a cell, the first percentage indicates the row percentage (i.e., the percent of younger siblings who are preterm or not preterm by the older sibling's preterm birth status), and the second percentage is a column percentage (i.e., the percent of older siblings who are preterm or not preterm by the younger sibling's preterm birth status).

7 Simulation Code

We conducted simulations in Stata Statistical Software: Release 16 [7].

```
1 /*
2 Estimating Sibling Spillovers with
3 Unobserved Confounding Using Gain-Scores
4
5 Simulation Code
6
7 David C. Mallinson, Felix Elwert
8 */
9
10 *****
11 *Preliminaries
12 *****
13
14 /*
15 Defintions:
16
17 t_1 = sibling 1's treatment
18 t_2 = sibling 2's treatment
19 y_1 = sibling 1's outcome
20 y_2 = sibling 2's outcome
21 u = unobserved fixed effect
22 d = gain-score (y_2-y_1)
23 */
24
25 **set working directory
26
27 /*
28 NOTE: This simulation saves datasets to compile results across models. Saved
29 files will be in the working directory. To specify the directory, run the cd
30 command with the desired file path (without the *).
31
32 At the end of this file, there is code to delete all resulting files from
33 this simulation analysis. This code is optional (for cleanup) and must be run
34 manually. See end of code for details.
35 */
36
37 *cd "[FILE PATH HERE]"
38
39
40 *****
41 *Simulate Model 1A
42 *****
43
44 **clear previous data
45 clear all
46 set more off
```

```

47
48 **monte carlo simulation settings
49 global nobs=5000 /*observations per run*/
50 global nmc=1000 /*number of runs*/
51 set seed 54385 /*seed for random number generator -- can change*/
52 set obs $nobs /*sets observations within each run*/
53
54 **set parameters (causal effects in models)
55 scalar delta=1 /*t_1 on y_1, t_2 on y_2*/
56
57 scalar psi=1 /*u on y_1, u on y_2*/
58 scalar chi=2 /*u on t_1*/
59 scalar gamma=3 /*u on t_2*/
60
61 scalar theta=0.5 /*t_1 on y_2*/
62 scalar kappa=0.3 /*t_2 on y_1*/
63 scalar tau=0.3 /*t_2 on t_1*/
64 scalar phi=0.3 /*t_1 on t_2*/
65
66 scalar omega=0.3 /*y_1 on t_2*/
67 scalar eta=0.3 /*y_1 on y_2*/
68 scalar lambda=0.3 /*y_2 on y_1*/
69
70 **generate variables
71
72 gen u=. /*unobserved fixed effect*/
73 gen t_1=. /*sibling 1's treatment*/
74 gen t_2=. /*sibling 2's treatment*/
75 gen y_1=. /*sibling 1's outcome*/
76 gen y_2=. /*sibling 2's outcome*/
77 gen d=. /*gain-score (y_2-y_1)*/
78
79 **simulation
80
81 tempname mla
82 postfile `mla' sc_mc1 sc_ub_mc1 sc_lb_mc1 using mla, replace
83 quietly {
84
85 forvalues i=1/$nmc {
86
87 /*note: for y_1 and y_2 equations, "rnormal(0,1)" is residual*/
88
89 replace u=rnormal(0,1)
90 replace t_1=(chi*u)>0.5
91 replace t_2=(gamma*u)>0.2
92 replace y_1=(delta*t_1)+(psi*u)+rnormal(0,1)
93 replace y_2=(delta*t_2)+(theta*t_1)+(psi*u)+rnormal(0,1)
94 replace d=y_2-y_1
95
96 /*gain-score regression*/

```

```

97  reg d t_1 t_2
98
99  /*lincom to compute spillover coefficient*/
100 lincom t_1+t_2
101
102 /*store lincom output*/
103 scalar sc_mc1=r(estimate) /*spillover coefficient*/
104 scalar sc_ub_mc1=(r(estimate)+(1.96*r(se))) /*95% CI upper bound*/
105 scalar sc_lb_mc1=(r(estimate)-(1.96*r(se))) /*95% CI lower bound*/
106
107 post `mla' (sc_mc1) (sc_ub_mc1) (sc_lb_mc1)
108 }
109 }
110 postclose `mla'
111
112 use mla, clear
113 summarize
114
115 **store mean of results, save to append with other simulation results
116 egen sc=mean(sc_mc1)
117 egen ub=mean(sc_ub_mc1)
118 egen lb=mean(sc_lb_mc1)
119 gen id=1 /*first simulation*/
120
121 keep sc ub lb id
122 duplicates drop
123 duplicates report
124
125 **save data
126 save "sim1.dta", replace
127
128
129 /******
130 *Simulate Model 1B
131 *****
132
133 **clear previous data
134 clear all
135 set more off
136
137 **monte carlo simulation settings
138 global nobs=5000 /*observations per run*/
139 global nmc=1000 /*number of runs*/
140 set seed 2414830 /*seed for random number generator -- can change*/
141 set obs $nobs /*sets observations within each run*/
142
143 **set parameters (causal effects in models)
144 scalar delta=1 /*t_1 on y_1, t_2 on y_2*/
145
146 scalar psi=1 /*u on y_1, u on y_2*/

```



```

147 scalar chi=2 /*u on t_1*/
148 scalar gamma=3 /*u on t_2*/
149
150 scalar theta=0.5 /*t_1 on y_2*/
151 scalar kappa=0.3 /*t_2 on y_1*/
152 scalar tau=0.3 /*t_2 on t_1*/
153 scalar phi=0.3 /*t_1 on t_2*/
154
155 scalar omega=0.3 /*y_1 on t_2*/
156 scalar eta=0.3 /*y_1 on y_2*/
157 scalar lambda=0.3 /*y_2 on y_1*/
158
159 **generate variables
160
161 gen u=. /*unobserved fixed effect*/
162 gen t_1=. /*sibling 1's treatment*/
163 gen t_2=. /*sibling 2's treatment*/
164 gen y_1=. /*sibling 1's outcome*/
165 gen y_2=. /*sibling 2's outcome*/
166 gen d=. /*gain-score (y_2-y_1)*/
167
168 **simulation
169
170 tempname m1b
171 postfile `m1b' sc_mc2 sc_ub_mc2 sc_lb_mc2 using m1b, replace
172 quietly {
173
174 forvalues i=1/$nmc {
175
176     /*note: for y_1 and y_2 equations, "rnormal(0,1)" is residual*/
177
178     replace u=rnormal(0,1)
179     replace t_2=(gamma*u)>0.2
180     replace t_1=(tau*t_2)+(chi*u)>0.5
181     replace y_1=(delta*t_1)+(psi*u)+rnormal(0,1)
182     replace y_2=(delta*t_2)+(theta*t_1)+(psi*u)+rnormal(0,1)
183     replace d=y_2-y_1
184
185     /*gain-score regression*/
186     reg d t_1 t_2
187
188     /*lincom to compute spillover coefficient*/
189     lincom t_1+t_2
190
191     /*store lincom output*/
192     scalar sc_mc2=r(estimate) /*spillover coefficient*/
193     scalar sc_ub_mc2=(r(estimate)+(1.96*r(se))) /*95% CI upper bound*/
194     scalar sc_lb_mc2=(r(estimate)-(1.96*r(se))) /*95% CI lower bound*/
195
196     post `m1b' (sc_mc2) (sc_ub_mc2) (sc_lb_mc2)

```

```

197 }
198 }
199 postclose `mlb'
200
201 use mlb, clear
202 summarize
203
204 **store mean of results, save to append with other simulation results
205 egen sc=mean(sc_mc2)
206 egen ub=mean(sc_ub_mc2)
207 egen lb=mean(sc_lb_mc2)
208 gen id=2 /*second simulation*/
209
210 keep sc ub lb id
211 duplicates drop
212 duplicates report
213
214 **save data
215 save "sim2.dta", replace
216
217
218 /******
219 *Simulate Model 1C
220 /******
221
222 **clear previous data
223 clear all
224 set more off
225
226 **monte carlo simulation settings
227 global nobs=5000 /*observations per run*/
228 global nmc=1000 /*number of runs*/
229 set seed 48910 /*seed for random number generator -- can change*/
230 set obs $nobs /*sets observations within each run*/
231
232 **set parameters (causal effects in models)
233 scalar delta=1 /*t_1 on y_1, t_2 on y_2*/
234
235 scalar psi=1 /*u on y_1, u on y_2*/
236 scalar chi=2 /*u on t_1*/
237 scalar gamma=3 /*u on t_2*/
238
239 scalar theta=0.5 /*t_1 on y_2*/
240 scalar kappa=0.3 /*t_2 on y_1*/
241 scalar tau=0.3 /*t_2 on t_1*/
242 scalar phi=0.3 /*t_1 on t_2*/
243
244 scalar omega=0.3 /*y_1 on t_2*/
245 scalar eta=0.3 /*y_1 on y_2*/
246 scalar lambda=0.3 /*y_2 on y_1*/

```

```

247
248 **generate variables
249
250 gen u=. /*unobserved fixed effect*/
251 gen t_1=. /*sibling 1's treatment*/
252 gen t_2=. /*sibling 2's treatment*/
253 gen y_1=. /*sibling 1's outcome*/
254 gen y_2=. /*sibling 2's outcome*/
255 gen d=. /*gain-score (y_2-y_1)*/
256
257 **simulation
258
259 tempname mlc
260 postfile `mlc' sc_mc3 sc_ub_mc3 sc_lb_mc3 using mlc, replace
261 quietly {
262
263 forvalues i=1/`$nmc' {
264
265 /*note: for y_1 and y_2 equations, "rnormal(0,1)" is residual*/
266
267 replace u=rnormal(0,1)
268 replace t_1=(chi*u)>0.5
269 replace t_2=(phi*t_1)+(gamma*u)>0.2
270 replace y_1=(delta*t_1)+(psi*u)+rnormal(0,1)
271 replace y_2=(delta*t_2)+(theta*t_1)+(psi*u)+rnormal(0,1)
272 replace d=y_2-y_1
273
274 /*gain-score regression*/
275 reg d t_1 t_2
276
277 /*lincom to compute spillover coefficient*/
278 lincom t_1+t_2
279
280 /*store lincom output*/
281 scalar sc_mc3=r(estimate) /*spillover coefficient*/
282 scalar sc_ub_mc3=(r(estimate)+(1.96*r(se))) /*95% CI upper bound*/
283 scalar sc_lb_mc3=(r(estimate)-(1.96*r(se))) /*95% CI lower bound*/
284
285 post `mlc' (sc_mc3) (sc_ub_mc3) (sc_lb_mc3)
286 }
287 }
288 postclose `mlc'
289
290 use mlc, clear
291 summarize
292
293 **store mean of results, save to append with other simulation results
294 egen sc=mean(sc_mc3)
295 egen ub=mean(sc_ub_mc3)
296 egen lb=mean(sc_lb_mc3)

```

```

297 gen id=3 /*third simulation*/
298
299 keep sc ub lb id
300 duplicates drop
301 duplicates report
302
303 **save data
304 save "sim3.dta", replace
305
306
307 /*****/
308 *Simulate Model 2A
309 /*****/
310
311 **clear previous data
312 clear all
313 set more off
314
315 **monte carlo simulation settings
316 global nobs=5000 /*observations per run*/
317 global nmc=1000 /*number of runs*/
318 set seed 9842 /*seed for random number generator -- can change*/
319 set obs $nobs /*sets observations within each run*/
320
321 **set parameters (causal effects in models)
322 scalar delta=1 /*t_1 on y_1, t_2 on y_2*/
323
324 scalar psi=1 /*u on y_1, u on y_2*/
325 scalar chi=2 /*u on t_1*/
326 scalar gamma=3 /*u on t_2*/
327
328 scalar theta=0.5 /*t_1 on y_2*/
329 scalar kappa=0.3 /*t_2 on y_1*/
330 scalar tau=0.3 /*t_2 on t_1*/
331 scalar phi=0.3 /*t_1 on t_2*/
332
333 scalar omega=0.3 /*y_1 on t_2*/
334 scalar eta=0.3 /*y_1 on y_2*/
335 scalar lambda=0.3 /*y_2 on y_1*/
336
337 **generate variables
338
339 gen u=. /*unobserved fixed effect*/
340 gen t_1=. /*sibling 1's treatment*/
341 gen t_2=. /*sibling 2's treatment*/
342 gen y_1=. /*sibling 1's outcome*/
343 gen y_2=. /*sibling 2's outcome*/
344 gen d=. /*gain-score (y_2-y_1)*/
345
346 **simulation

```

```

347
348 tempname m2a
349 postfile `m2a' sc_mc4 sc_ub_mc4 sc_lb_mc4 using m2a, replace
350 quietly {
351
352 forvalues i=1/$nmc {
353
354     /*note: for y_1 and y_2 equations, "rnormal(0,1)" is residual*/
355
356     replace u=rnormal(0,1)
357     replace t_1=(chi*u)>0.5
358     replace t_2=(gamma*u)>0.2
359     replace y_1=(delta*t_1)+(kappa*t_2)+(psi*u)+rnormal(0,1)
360     replace y_2=(delta*t_2)+(theta*t_1)+(psi*u)+rnormal(0,1)
361     replace d=y_2-y_1
362
363     /*gain-score regression*/
364     reg d t_1 t_2
365
366     /*lincom to compute spillover coefficient*/
367     lincom t_1+t_2
368
369     /*store lincom output*/
370     scalar sc_mc4=r(estimate) /*spillover coefficient*/
371     scalar sc_ub_mc4=(r(estimate)+(1.96*r(se))) /*95% CI upper bound*/
372     scalar sc_lb_mc4=(r(estimate)-(1.96*r(se))) /*95% CI lower bound*/
373
374     post `m2a' (sc_mc4) (sc_ub_mc4) (sc_lb_mc4)
375 }
376 }
377 postclose `m2a'
378
379 use m2a, clear
380 summarize
381
382 **store mean of results, save to append with other simulation results
383 egen sc=mean(sc_mc4)
384 egen ub=mean(sc_ub_mc4)
385 egen lb=mean(sc_lb_mc4)
386 gen id=4 /*fourth simulation*/
387
388 keep sc ub lb id
389 duplicates drop
390 duplicates report
391
392 **save data
393 save "sim4.dta", replace
394
395
396 /*******/

```

```

397 *Simulate Model 2B
398 /******
399
400 **clear previous data
401 clear all
402 set more off
403
404 **monte carlo simulation settings
405 global nobs=5000 /*observations per run*/
406 global nmc=1000 /*number of runs*/
407 set seed 33305 /*seed for random number generator -- can change*/
408 set obs $nobs /*sets observations within each run*/
409
410 **set parameters (causal effects in models)
411 scalar delta=1 /*t_1 on y_1, t_2 on y_2*/
412
413 scalar psi=1 /*u on y_1, u on y_2*/
414 scalar chi=2 /*u on t_1*/
415 scalar gamma=3 /*u on t_2*/
416
417 scalar theta=0.5 /*t_1 on y_2*/
418 scalar kappa=0.3 /*t_2 on y_1*/
419 scalar tau=0.3 /*t_2 on t_1*/
420 scalar phi=0.3 /*t_1 on t_2*/
421
422 scalar omega=0.3 /*y_1 on t_2*/
423 scalar eta=0.3 /*y_1 on y_2*/
424 scalar lambda=0.3 /*y_2 on y_1*/
425
426 **generate variables
427
428 gen u=. /*unobserved fixed effect*/
429 gen t_1=. /*sibling 1's treatment*/
430 gen t_2=. /*sibling 2's treatment*/
431 gen y_1=. /*sibling 1's outcome*/
432 gen y_2=. /*sibling 2's outcome*/
433 gen d=. /*gain-score (y_2-y_1)*/
434
435 **simulation
436
437 tempname m2b
438 postfile 'm2b' sc_mc5 sc_ub_mc5 sc_lb_mc5 using m2b, replace
439 quietly {
440
441 forvalues i=1/$nmc {
442
443 /*note: for y_1 and y_2 equations, "rnormal(0,1)" is residual*/
444
445 replace u=rnormal(0,1)
446 replace t_2=(gamma*u)>0.2

```

```

447 replace t_1=(tau*t_2)+(chi*u)>0.5
448 replace y_1=(delta*t_1)+(kappa*t_2)+(psi*u)+rnormal(0,1)
449 replace y_2=(delta*t_2)+(theta*t_1)+(psi*u)+rnormal(0,1)
450 replace d=y_2-y_1
451
452 /*gain-score regression*/
453 reg d t_1 t_2
454
455 /*lincom to compute spillover coefficient*/
456 lincom t_1+t_2
457
458 /*store lincom output*/
459 scalar sc_mc5=r(estimate) /*spillover coefficient*/
460 scalar sc_ub_mc5=(r(estimate)+(1.96*r(se))) /*95% CI upper bound*/
461 scalar sc_lb_mc5=(r(estimate)-(1.96*r(se))) /*95% CI lower bound*/
462
463 post 'm2b' (sc_mc5) (sc_ub_mc5) (sc_lb_mc5)
464 }
465 }
466 postclose 'm2b'
467
468 use m2b, clear
469 summarize
470
471 **store mean of results, save to append with other simulation results
472 egen sc=mean(sc_mc5)
473 egen ub=mean(sc_ub_mc5)
474 egen lb=mean(sc_lb_mc5)
475 gen id=5 /*fifth simulation*/
476
477 keep sc ub lb id
478 duplicates drop
479 duplicates report
480
481 **save data
482 save "sim5.dta", replace
483
484
485 /******
486 *Simulate Model 2C
487 /******
488
489 **clear previous data
490 clear all
491 set more off
492
493 **monte carlo simulation settings
494 global nobs=5000 /*observations per run*/
495 global nmc=1000 /*number of runs*/
496 set seed 101678 /*seed for random number generator -- can change*/

```

```

497 set obs $nobs /*sets observations within each run*/
498
499 **set parameters (causal effects in models)
500 scalar delta=1 /*t_1 on y_1, t_2 on y_2*/
501
502 scalar psi=1 /*u on y_1, u on y_2*/
503 scalar chi=2 /*u on t_1*/
504 scalar gamma=3 /*u on t_2*/
505
506 scalar theta=0.5 /*t_1 on y_2*/
507 scalar kappa=0.3 /*t_2 on y_1*/
508 scalar tau=0.3 /*t_2 on t_1*/
509 scalar phi=0.3 /*t_1 on t_2*/
510
511 scalar omega=0.3 /*y_1 on t_2*/
512 scalar eta=0.3 /*y_1 on y_2*/
513 scalar lambda=0.3 /*y_2 on y_1*/
514
515 **generate variables
516
517 gen u=. /*unobserved fixed effect*/
518 gen t_1=. /*sibling 1's treatment*/
519 gen t_2=. /*sibling 2's treatment*/
520 gen y_1=. /*sibling 1's outcome*/
521 gen y_2=. /*sibling 2's outcome*/
522 gen d=. /*gain-score (y_2-y_1)*/
523
524 **simulation
525
526 tempname m2c
527 postfile 'm2c' sc_mc6 sc_ub_mc6 sc_lb_mc6 using m2c, replace
528 quietly {
529
530 forvalues i=1/$nmc {
531
532     /*note: for y_1 and y_2 equations, "rnormal(0,1)" is residual*/
533
534     replace u=rnormal(0,1)
535     replace t_1=(chi*u)>0.5
536     replace t_2=(phi*t_1)+(gamma*u)>0.2
537     replace y_1=(delta*t_1)+(kappa*t_2)+(psi*u)+rnormal(0,1)
538     replace y_2=(delta*t_2)+(theta*t_1)+(psi*u)+rnormal(0,1)
539     replace d=y_2-y_1
540
541     /*gain-score regression*/
542     reg d t_1 t_2
543
544     /*lincom to compute spillover coefficient*/
545     lincom t_1+t_2
546

```



```

547  /*store lincom output*/
548  scalar sc_mc6=r(estimate) /*spillover coefficient*/
549  scalar sc_ub_mc6=(r(estimate)+(1.96*r(se))) /*95% CI upper bound*/
550  scalar sc_lb_mc6=(r(estimate)-(1.96*r(se))) /*95% CI lower bound*/
551
552  post `m2c' (sc_mc6) (sc_ub_mc6) (sc_lb_mc6)
553 }
554 }
555 postclose `m2c'
556
557 use m2c, clear
558 summarize
559
560 **store mean of results, save to append with other simulation results
561 egen sc=mean(sc_mc6)
562 egen ub=mean(sc_ub_mc6)
563 egen lb=mean(sc_lb_mc6)
564 gen id=6 /*sixth simulation*/
565
566 keep sc ub lb id
567 duplicates drop
568 duplicates report
569
570 **save data
571 save "sim6.dta", replace
572
573
574 /******
575 *Simulate Model 3A
576 *****
577
578 **clear previous data
579 clear all
580 set more off
581
582 **monte carlo simulation settings
583 global nobs=5000 /*observations per run*/
584 global nmc=1000 /*number of runs*/
585 set seed 278 /*seed for random number generator -- can change*/
586 set obs $nobs /*sets observations within each run*/
587
588 **set parameters (causal effects in models)
589 scalar delta=1 /*t_1 on y_1, t_2 on y_2*/
590
591 scalar psi=1 /*u on y_1, u on y_2*/
592 scalar chi=2 /*u on t_1*/
593 scalar gamma=3 /*u on t_2*/
594
595 scalar theta=0.5 /*t_1 on y_2*/
596 scalar kappa=0.3 /*t_2 on y_1*/

```

```

597 scalar tau=0.3 /*t_2 on t_1*/
598 scalar phi=0.3 /*t_1 on t_2*/
599
600 scalar omega=0.3 /*y_1 on t_2*/
601 scalar eta=0.3 /*y_1 on y_2*/
602 scalar lambda=0.3 /*y_2 on y_1*/
603
604 **generate variables
605
606 gen u=. /*unobserved fixed effect*/
607 gen t_1=. /*sibling 1's treatment*/
608 gen t_2=. /*sibling 2's treatment*/
609 gen y_1=. /*sibling 1's outcome*/
610 gen y_2=. /*sibling 2's outcome*/
611 gen d=. /*gain-score (y_2-y_1)*/
612
613 **simulation
614
615 tempname m3a
616 postfile `m3a' sc_mc7 sc_ub_mc7 sc_lb_mc7 using m3a, replace
617 quietly {
618
619 forvalues i=1/$nmc {
620
621     /*note: for y_1 and y_2 equations, "rnormal(0,1)" is residual*/
622
623     replace u=rnormal(0,1)
624     replace t_1=(chi*u)>0.5
625     replace y_1=(delta*t_1)+(psi*u)+rnormal(0,1)
626     replace t_2=(omega*y_1)+(gamma*u)>0.2
627     replace y_2=(delta*t_2)+(theta*t_1)+(psi*u)+rnormal(0,1)
628     replace d=y_2-y_1
629
630     /*gain-score regression*/
631     reg d t_1 t_2
632
633     /*lincom to compute spillover coefficient*/
634     lincom t_1+t_2
635
636     /*store lincom output*/
637     scalar sc_mc7=r(estimate) /*spillover coefficient*/
638     scalar sc_ub_mc7=(r(estimate)+(1.96*r(se))) /*95% CI upper bound*/
639     scalar sc_lb_mc7=(r(estimate)-(1.96*r(se))) /*95% CI lower bound*/
640
641     post `m3a' (sc_mc7) (sc_ub_mc7) (sc_lb_mc7)
642 }
643 }
644 postclose `m3a'
645
646 use m3a, clear

```

```

647 summarize
648
649 **store mean of results, save to append with other simulation results
650 egen sc=mean(sc_mc7)
651 egen ub=mean(sc_ub_mc7)
652 egen lb=mean(sc_lb_mc7)
653 gen id=7 /*seventh simulation*/
654
655 keep sc ub lb id
656 duplicates drop
657 duplicates report
658
659 **save data
660 save "sim7.dta", replace
661
662
663 /******
664 *Simulate Model 3B
665 /******
666
667 **clear previous data
668 clear all
669 set more off
670
671 **monte carlo simulation settings
672 global nobs=5000 /*observations per run*/
673 global nmc=1000 /*number of runs*/
674 set seed 53893 /*seed for random number generator -- can change*/
675 set obs $nobs /*sets observations within each run*/
676
677 **set parameters (causal effects in models)
678 scalar delta=1 /*t_1 on y_1, t_2 on y_2*/
679
680 scalar psi=1 /*u on y_1, u on y_2*/
681 scalar chi=2 /*u on t_1*/
682 scalar gamma=3 /*u on t_2*/
683
684 scalar theta=0.5 /*t_1 on y_2*/
685 scalar kappa=0.3 /*t_2 on y_1*/
686 scalar tau=0.3 /*t_2 on t_1*/
687 scalar phi=0.3 /*t_1 on t_2*/
688
689 scalar omega=0.3 /*y_1 on t_2*/
690 scalar eta=0.3 /*y_1 on y_2*/
691 scalar lambda=0.3 /*y_2 on y_1*/
692
693 **generate variables
694
695 gen u=. /*unobserved fixed effect*/
696 gen t_1=. /*sibling 1's treatment*/

```

```

697 gen t_2=. /*sibling 2's treatment*/
698 gen y_1=. /*sibling 1's outcome*/
699 gen y_2=. /*sibling 2's outcome*/
700 gen d=. /*gain-score (y_2-y_1)*/
701
702 **simulation
703
704 tempname m3b
705 postfile `m3b' sc_mc8 sc_ub_mc8 sc_lb_mc8 using m3b, replace
706 quietly {
707
708 forvalues i=1/$nmc {
709
710     /*note: for y_1 and y_2 equations, "rnormal(0,1)" is residual*/
711
712     replace u=rnormal(0,1)
713     replace t_1=(chi*u)>0.5
714     replace t_2=(gamma*u)>0.2
715     replace y_1=(delta*t_1)+(psi*u)+rnormal(0,1)
716     replace y_2=(delta*t_2)+(theta*t_1)+(eta*y_1)+(psi*u)+rnormal(0,1)
717     replace d=y_2-y_1
718
719     /*gain-score regression*/
720     reg d t_1 t_2
721
722     /*lincom to compute spillover coefficient*/
723     lincom t_1+t_2
724
725     /*store lincom output*/
726     scalar sc_mc8=r(estimate) /*spillover coefficient*/
727     scalar sc_ub_mc8=(r(estimate)+(1.96*r(se))) /*95% CI upper bound*/
728     scalar sc_lb_mc8=(r(estimate)-(1.96*r(se))) /*95% CI lower bound*/
729
730     post `m3b' (sc_mc8) (sc_ub_mc8) (sc_lb_mc8)
731 }
732 }
733 postclose `m3b'
734
735 use m3b, clear
736 summarize
737
738 **store mean of results, save to append with other simulation results
739 egen sc=mean(sc_mc8)
740 egen ub=mean(sc_ub_mc8)
741 egen lb=mean(sc_lb_mc8)
742 gen id=8 /*eighth simulation*/
743
744 keep sc ub lb id
745 duplicates drop
746 duplicates report

```

```

747
748 **save data
749 save "sim8.dta", replace
750
751
752 *****
753 *Simulate Model 3C
754 *****
755
756 **clear previous data
757 clear all
758 set more off
759
760 **monte carlo simulation settings
761 global nobs=5000 /*observations per run*/
762 global nmc=1000 /*number of runs*/
763 set seed 9921 /*seed for random number generator -- can change*/
764 set obs $nobs /*sets observations within each run*/
765
766 **set parameters (causal effects in models)
767 scalar delta=1 /*t_1 on y_1, t_2 on y_2*/
768
769 scalar psi=1 /*u on y_1, u on y_2*/
770 scalar chi=2 /*u on t_1*/
771 scalar gamma=3 /*u on t_2*/
772
773 scalar theta=0.5 /*t_1 on y_2*/
774 scalar kappa=0.3 /*t_2 on y_1*/
775 scalar tau=0.3 /*t_2 on t_1*/
776 scalar phi=0.3 /*t_1 on t_2*/
777
778 scalar omega=0.3 /*y_1 on t_2*/
779 scalar eta=0.3 /*y_1 on y_2*/
780 scalar lambda=0.3 /*y_2 on y_1*/
781
782 **generate variables
783
784 gen u=. /*unobserved fixed effect*/
785 gen t_1=. /*sibling 1's treatment*/
786 gen t_2=. /*sibling 2's treatment*/
787 gen y_1=. /*sibling 1's outcome*/
788 gen y_2=. /*sibling 2's outcome*/
789 gen d=. /*gain-score (y_2-y_1)*/
790
791 **simulation
792
793 tempname m3c
794 postfile `m3c' sc_mc9 sc_ub_mc9 sc_lb_mc9 using m3c, replace
795 quietly {
796

```

```

797 forvalues i=1/`$nmc' {
798
799     /*note: for y_1 and y_2 equations, "rnormal(0,1)" is residual*/
800
801     replace u=rnormal(0,1)
802     replace t_1=(chi*u)>0.5
803     replace t_2=(gamma*u)>0.2
804     replace y_2=(delta*t_2)+(theta*t_1)+(psi*u)+rnormal(0,1)
805     replace y_1=(delta*t_1)+(lambda*y_2)+(psi*u)+rnormal(0,1)
806     replace d=y_2-y_1
807
808     /*gain-score regression*/
809     reg d t_1 t_2
810
811     /*lincom to compute spillover coefficient*/
812     lincom t_1+t_2
813
814     /*store lincom output*/
815     scalar sc_mc9=r(estimate) /*spillover coefficient*/
816     scalar sc_ub_mc9=(r(estimate)+(1.96*r(se))) /*95% CI upper bound*/
817     scalar sc_lb_mc9=(r(estimate)-(1.96*r(se))) /*95% CI lower bound*/
818
819     post `m3c' (sc_mc9) (sc_ub_mc9) (sc_lb_mc9)
820 }
821 }
822 postclose `m3c'
823
824 use m3c, clear
825 summarize
826
827 **store mean of results, save to append with other simulation results
828 egen sc=mean(sc_mc9)
829 egen ub=mean(sc_ub_mc9)
830 egen lb=mean(sc_lb_mc9)
831 gen id=9 /*ninth simulation*/
832
833 keep sc ub lb id
834 duplicates drop
835 duplicates report
836
837 **save data
838 save "sim9.dta", replace
839
840
841 /******
842 *Graph Results
843 *****
844
845 clear
846 set more off

```

```

847
848 **open first simulation dataset
849 use "sim1.dta"
850
851 **append datasets (sim2-9)
852 append using "sim2.dta"
853 append using "sim3.dta"
854 append using "sim4.dta"
855 append using "sim5.dta"
856 append using "sim6.dta"
857 append using "sim7.dta"
858 append using "sim8.dta"
859 append using "sim9.dta"
860
861 **label id (assign to models)
862 label define model_lbl 1 "Figure 2.1A" 2 "1B" 3 "1C" ///
863     4 "Figure 2.2A" 5 "2B" 6 "2C" ///
864     7 "Figure 2.3A" 8 "3B" 9 "3C"
865
866 label values id model_lbl
867
868 **display results
869
870 /*
871 NOTE:
872 id = sibling spillover model
873 sc = spillover coefficient
874 lb = 95% CI lower bound for spillover coefficient
875 ub = 95% CI upper bound for spillover coefficient
876 */
877
878 list id sc lb ub
879
880 **graph results
881
882 /*
883 NOTE: When importing Stata code into LaTeX, the compiler will not accept the en-dash
884 in "{&theta} [MINUS] {&kappa}." Replace the hyphen with an en-dash to recreate
885 the original figure.
886 */
887
888 twoway(scatter id sc, mcolor(black) ysc(reverse) ///
889     graphregion(color(white)) ///
890     xtitle(Mean Spillover Coefficient (95% CI)) ///
891     ytitle(Sibling Spillover Model) ///
892     text(8.5 0.5 "({&theta} = 0.5)", size(small) ///
893         box margin(l+0.5 r+0.5 t+0.5 b+1)) ///
894     text(8.5 0.2 "({&theta} - {&kappa} = 0.2)", size(small) ///
895         box margin(l+0.5 r+0.5 t+0.5 b+1)) ///
896     xline(0.2, lcolor(black) lpattern(shortdash_dot)) ///

```

```

895     xline(0.5, lcolor(black) lpattern(dash)) ///
896     legend(off) xtick(#21) ///
897     (rcap ub lb id, horizontal lcolor(black)), ///
898     ylabel(1(1)9, angle(horizontal) labsize(small) ///
899     valuelabel notick) ///
900     xlabel(-0.5(0.5)1.5, labsize(small)) saving(sim_results, replace)
901
902
903 /*
904
905 NOTE: The code below erases results (datasets and the figure). Highlight the
906 code below and press the "Execute (do)" button to delete files.
907
908 **erase datasets with simulation results
909 erase m1a.dta
910 erase m1b.dta
911 erase m1c.dta
912 erase m2a.dta
913 erase m2b.dta
914 erase m2c.dta
915 erase m3a.dta
916 erase m3b.dta
917 erase m3c.dta
918
919 erase sim1.dta
920 erase sim2.dta
921 erase sim3.dta
922 erase sim4.dta
923 erase sim5.dta
924 erase sim6.dta
925 erase sim7.dta
926 erase sim8.dta
927 erase sim9.dta
928
929 **erase figure with simulation results
930 erase sim_results.gph

```


References

- [1] Pearl J. Linear models: a useful "microscope" for causal analysis. *J Causal Inference*. 2013;1:155-170.
- [2] Invernizzi M, Juel C, Swank L, Meier J. PALS-K Technical Reference. Charlottesville, VA: University of Virginia Curry School of Education, 2015.
- [3] Larson A, Berger LM, Mallinson DC, Grodsky E, Ehrental DB. Variable uptake of Medicaid-covered Prenatal Care Coordination: the relevance of treatment level and service context. *J Community Health*. 2019;44:32-43.
- [4] Mallinson DC, Grodsky E, Ehrental DB. Gestational age, kindergarten-level literacy, and effect modification by maternal socio-economic and demographic factors. *Paediatr Perinat Epidemiol*. 2019;33:467-479.
- [5] Wisconsin Department of Public Instruction. Admissions and early entrance to four- and five-year-old kindergarten. <https://dpi.wi.gov/early-childhood/kind/admission> (1 September 2020, date last accessed).
- [6] Ford K, Invernizzi M. PALS Español K Technical Reference. Charlottesville, VA: University of Virginia Curry School of Education; 2014.
- [7] StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC; 2019.