Fixed effects identification of causal effects

Fixed effects is an ignorable treatment, or perhaps more aptly, selection on observables identification strategy for causal effects. The idea is that individual and/or time mean effects help support conditional mean independence of treatment on potential outcomes with (y_1) and without (y_0) treatment. These potential outcomes are only partially observed, thus conditional mean independence is a thought experiment. While fixed effects may make conditional mean independence more plausible, increasing the number of covariates likely makes common support more challenging. Balancing common support and a strategy for mapping observables into the quantities of interest is typical of counterfactual-based causality.

Next, we present two simple examples. One in which OLS (without fixed effects) as well as fixed effects identifies the average treatment effect and another in which individual fixed effects identifies the average treatment effect but ignoring the fixed effects (OLS) produces bias.

Example 1 (identified via OLS) *Suppose the data generating process (DGP) involves individual effects but no time effects as follows.*

y_1	y_0	TE	y	D	x	v_1	v_2	ϵ	u
9	2	7	9	1	1	3	0	1	4
3	0	3	3	1	-1	1	2	-1	0
9	2	7	2	0	1	3	0	1	1
3	0	3	0	0	-1	1	2	-1	1
7	-1	8	-1	0	1	1	-3	1	-2
0	-3	3	-3	0	-1	-2	-1	-1	-2
7	-1	8	7	1	1	1	-3	1	2
0	-3	3	0	1	-1	-2	-1	-1	-3

The DGP is homogeneous with

$$\begin{array}{rr} ATE = E\left[y_1 - y_0\right] = E_X\left[y_1 - y_0 \mid x\right] & 5.25 \\ ATT = E\left[y_1 - y_0 \mid D = 1\right] = & 5.25 \\ ATUT = E\left[y_1 - y_0 \mid D = 0\right] = & 5.25 \\ ATE\left(x = 1\right) = E\left[y_1 - y_0 \mid x = 1\right] = & 7.5 \\ ATE\left(x = -1\right) = E\left[y_1 - y_0 \mid x = -1\right] = & 3 \end{array}$$

The process is linear. OLS identifies

$$\begin{aligned} E\left[y \mid X, D\right] &= \gamma_0 + \alpha D + \beta x + \delta x * D \\ &= -0.5 + 5.25D + 1x + 2.25x * D \end{aligned}$$

and fixed effects identifies

$$\begin{aligned} E\left[y \mid X, I_1, I_2, D\right] &= \alpha D + \beta x + \delta x * D + \gamma_1 I_1 + \gamma_2 I_2 \\ &= 5.25D + 1x + 2.25x * D + 0.875I_1 - 1.875I_2 \end{aligned}$$

where $u = Dv_1 + (1-D)\,v_2 + \epsilon$, v_j is the error component for individual j, ϵ is an error, D is an indicator variable for treatment, x is a covariate, I_j is an indicator variable for individual j, and $y = Dy_1 + (1-D)\,y_0$ is observed outcome. Since the mean of X is zero, α identifies ATE. OLS (without fixed effects) supports identification as conditional mean independence as well as common support is satisfied.

$$\begin{aligned} E[y_1 \mid x=1, D=1] &= E[y_1 \mid x=1, D=0] = 8 \\ E[y_1 \mid x=-1, D=1] &= E[y_1 \mid x=-1, D=0] = 1.5 \end{aligned}$$

and

$$\begin{aligned} E[y_0 \mid x=1, D=1] &= E[y_0 \mid x=1, D=0] = 0.5 \\ E[y_0 \mid x=-1, D=1] &= E[y_0 \mid x=-1, D=0] = -1.5 \end{aligned}$$

Fixed effects supports identification as conditional mean independence as well as common support is satisfied.

$$\begin{aligned} E[y_1 \mid x=1, I_1=1, I_2=0, D=1] &= E[y_1 \mid x=1, I_1=1, I_2=0, D=0] = 9 \\ E[y_1 \mid x=-1, I_1=1, I_2=0, D=1] &= E[y_1 \mid x=-1, I_1=1, I_2=0, D=0] = 3 \\ E[y_1 \mid x=1, I_1=0, I_2=1, D=1] &= E[y_1 \mid x=1, I_1=0, I_2=1, D=0] = 7 \\ E[y_1 \mid x=-1, I_1=0, I_2=1, D=1] &= E[y_1 \mid x=-1, I_1=0, I_2=1, D=0] = 0 \end{aligned}$$

and

$$\begin{aligned} E[y_0 \mid x=1, I_1=1, I_2=0, D=1] &= E[y_0 \mid x=1, I_1=1, I_2=0, D=0] = 2 \\ E[y_0 \mid x=-1, I_1=1, I_2=0, D=1] &= E[y_0 \mid x=-1, I_1=1, I_2=0, D=0] = 0 \\ E[y_0 \mid x=1, I_1=0, I_2=1, D=1] &= E[y_0 \mid x=1, I_1=0, I_2=1, D=0] = -1 \\ E[y_0 \mid x=-1, I_1=0, I_2=1, D=1] &= E[y_0 \mid x=-1, I_1=0, I_2=1, D=0] = -3 \end{aligned}$$

Example 2 (identified via fixed effects) *Suppose the data generating process (DGP) involves individual effects but no time effects as follows.*

y_1	y_0	TE	y	D	x	v_1	v_2	ϵ	u
9	2	7	9	1	1	3	0	1	4
3	0	3	3	1	−1	1	2	−1	0
9	2	7	2	0	1	3	0	1	1
3	0	3	0	0	−1	1	2	−1	1
9	2	7	9	1	1	3	0	1	4
3	0	3	3	1	−1	1	2	−1	0
7	−1	8	−1	0	1	1	−3	1	−2
0	−3	3	−3	0	−1	−2	−1	−1	−2
7	−1	8	−1	0	1	1	−3	1	−2
0	−3	3	−3	0	−1	−2	−1	−1	−2
7	−1	8	7	1	1	1	−3	1	2
0	−3	3	0	1	−1	−2	−1	−1	−3

The DGP is heterogeneous with

$$\begin{array}{rc} ATE = E[y_1 - y_0] = E_X[y_1 - y_0 \mid x] & 5\frac{1}{4} \\ ATT = E[y_1 - y_0 \mid D = 1] = & 5\frac{1}{6} \\ ATUT = E[y_1 - y_0 \mid D = 0] = & 5\frac{1}{3} \\ ATE(x = 1) = E[y_1 - y_0 \mid x = 1] = & 7\frac{1}{2} \\ ATE(x = -1) = E[y_1 - y_0 \mid x = -1] = & 3 \end{array}$$

The process is linear. OLS identifies

$$\begin{aligned} E[y \mid X, D] &= \gamma_0 + \alpha D + \beta x + \delta x * D \\ &= -1 + 6\frac{1}{6}D + 1x + 2\frac{1}{6}x * D \end{aligned}$$

and fixed effects identifies

$$\begin{aligned} E[y \mid X, I_1, I_2, D] &= \alpha D + \beta x + \delta x * D + \gamma_1 I_1 + \gamma_2 I_2 \\ &= 5\frac{1}{4}D + 1x + 2\frac{1}{6}x * D + \frac{5}{6}I_1 - 1\frac{11}{12}I_2 \end{aligned}$$

where $u = Dv_1 + (1 - D)v_2 + \epsilon$, v_j is the error component for individual j, ϵ is an error, D is an indicator variable for treatment, x is a covariate, I_j is an indicator variable for individual j, and $y = Dy_1 + (1 - D)y_0$ is observed outcome. Since the mean of X is zero, α identifies ATE. Even though common support is satisfied, OLS (without fixed effects) fails to support identification as conditional mean independence (on X) is not satisfied.

$$\begin{aligned} E[y_1 \mid x = 1, D = 1] &= 8\frac{1}{3} \neq E[y_1 \mid x = 1, D = 0] = 7\frac{2}{3} \\ E[y_1 \mid x = -1, D = 1] &= 2 \neq E[y_1 \mid x = -1, D = 0] = 1 \end{aligned}$$

and

$$\begin{aligned} E[y_0 \mid x = 1, D = 1] &= 1 \neq E[y_0 \mid x = 1, D = 0] = 0 \\ E[y_0 \mid x = -1, D = 1] &= -1 \neq E[y_0 \mid x = -1, D = 0] = -2 \end{aligned}$$

However, fixed effects supports identification as conditional mean independence (on X, I_1, and I_2) as well as common support is satisfied.

$$\begin{aligned} E[y_1 \mid x = 1, I_1 = 1, I_2 = 0, D = 1] &= E[y_1 \mid x = 1, I_1 = 1, I_2 = 0, D = 0] = 9 \\ E[y_1 \mid x = -1, I_1 = 1, I_2 = 0, D = 1] &= E[y_1 \mid x = -1, I_1 = 1, I_2 = 0, D = 0] = 3 \\ E[y_1 \mid x = 1, I_1 = 0, I_2 = 1, D = 1] &= E[y_1 \mid x = 1, I_1 = 0, I_2 = 1, D = 0] = 7 \\ E[y_1 \mid x = -1, I_1 = 0, I_2 = 1, D = 1] &= E[y_1 \mid x = -1, I_1 = 0, I_2 = 1, D = 0] = 0 \end{aligned}$$

and

$$\begin{aligned} E[y_0 \mid x = 1, I_1 = 1, I_2 = 0, D = 1] &= E[y_0 \mid x = 1, I_1 = 1, I_2 = 0, D = 0] = 2 \\ E[y_0 \mid x = -1, I_1 = 1, I_2 = 0, D = 1] &= E[y_0 \mid x = -1, I_1 = 1, I_2 = 0, D = 0] = 0 \\ E[y_0 \mid x = 1, I_1 = 0, I_2 = 1, D = 1] &= E[y_0 \mid x = 1, I_1 = 0, I_2 = 1, D = 0] = -1 \\ E[y_0 \mid x = -1, I_1 = 0, I_2 = 1, D = 1] &= E[y_0 \mid x = -1, I_1 = 0, I_2 = 1, D = 0] = -3 \end{aligned}$$

Next, we consider a modest variation on the above DGP that results in identification failure based on fixed effects (or OLS). Modifications from the above DGP are marked in bold and placed inside braces. The variation is hidden from view as it resides in the counterfactual component of the outcome distribution. Consequently, fixed effects (as well as OLS) produce the same estimates as in the previous example.

Example 3 (fixed effects fails) *Suppose the data generating process (DGP) involves individual effects but no time effects as follows.*

y_1	y_0	TE	y	D	x	v_1	v_2	ϵ	u
9	2	7	9	1	1	3	0	1	4
3	0	3	3	1	-1	1	2	-1	0
9	2	7	2	0	1	3	0	1	1
3	0	3	0	0	-1	1	2	-1	1
9	2	7	9	1	1	3	0	1	4
3	$\{\mathbf{-2}\}$	3	3	1	-1	1	$\{\mathbf{0}\}$	-1	0
7	-1	8	-1	0	1	1	-3	1	-2
0	-3	3	-3	0	-1	-2	-1	-1	-2
7	-1	8	-1	0	1	1	-3	1	-2
0	-3	3	-3	0	-1	-2	-1	-1	-2
7	-1	8	7	1	1	1	-3	1	2
0	-3	3	0	1	-1	-2	-1	-1	-3

The DGP is heterogeneous with

$$\begin{array}{rc} ATE = E\left[y_1 - y_0\right] = E_X\left[y_1 - y_0 \mid x\right] & \left\{\mathbf{5\tfrac{5}{12}}\right\} \\ ATT = E\left[y_1 - y_0 \mid D = 1\right] = & \left\{\mathbf{5\tfrac{1}{2}}\right\} \\ ATUT = E\left[y_1 - y_0 \mid D = 0\right] = & 5\tfrac{1}{3} \\ ATE\left(x = 1\right) = E\left[y_1 - y_0 \mid x = 1\right] = & 7\tfrac{1}{2} \\ ATE\left(x = -1\right) = E\left[y_1 - y_0 \mid x = -1\right] = & \left\{\mathbf{3\tfrac{1}{3}}\right\} \end{array}$$

The process is linear. OLS identifies

$$\begin{aligned} E\left[y \mid X, D\right] &= \gamma_0 + \alpha D + \beta x + \delta x * D \\ &= -1 + 6\frac{1}{6}D + 1x + 2\frac{1}{6}x * D \end{aligned}$$

and fixed effects identifies

$$\begin{aligned} E\left[y \mid X, I_1, I_2, D\right] &= \alpha D + \beta x + \delta x * D + \gamma_1 I_1 + \gamma_2 I_2 \\ &= 5\frac{1}{4}D + 1x + 2\frac{1}{6}x * D + \frac{5}{6}I_1 - 1\frac{11}{12}I_2 \end{aligned}$$

where $u = Dv_1 + (1 - D)\,v_2 + \epsilon$*,* v_j *is the error component for individual* j*,* ϵ *is an error,* D *is an indicator variable for treatment,* x *is a covariate,* I_j *is an indicator variable for individual* j*, and* $y = Dy_1 + (1 - D)\,y_0$ *is observed* *outcome. Since the mean of X is zero, α identifies ATE. Even though common support is satisfied, OLS (without fixed effects) fails to support identification as conditional mean independence (on X) is not satisfied.*

$$\begin{aligned} E[y_1 \mid x=1, D=1] &= 8\frac{1}{3} \neq E[y_1 \mid x=1, D=0] = 7\frac{2}{3} \\ E[y_1 \mid x=-1, D=1] &= 2 \neq E[y_1 \mid x=-1, D=0] = 1 \end{aligned}$$

and

$$\begin{aligned} E[y_0 \mid x=1, D=1] &= 1 \neq E[y_0 \mid x=1, D=0] = 0 \\ E[y_0 \mid x=-1, D=1] &= \left\{-\mathbf{1}\frac{2}{3}\right\} \neq E[y_0 \mid x=-1, D=0] = -2 \end{aligned}$$

While common support is satisfied, fixed effects fails to identify ATE as conditional mean independence (on X, I_1, and I_2) is not satisfied.

$$\begin{aligned} E[y_1 \mid x=1, I_1=1, I_2=0, D=1] &= E[y_1 \mid x=1, I_1=1, I_2=0, D=0] = 9 \\ E[y_1 \mid x=-1, I_1=1, I_2=0, D=1] &= E[y_1 \mid x=-1, I_1=1, I_2=0, D=0] = 3 \\ E[y_1 \mid x=1, I_1=0, I_2=1, D=1] &= E[y_1 \mid x=1, I_1=0, I_2=1, D=0] = 7 \\ E[y_1 \mid x=-1, I_1=0, I_2=1, D=1] &= E[y_1 \mid x=-1, I_1=0, I_2=1, D=0] = 0 \end{aligned}$$

and

$$\begin{aligned} E[y_0 \mid x=1, I_1=1, I_2=0, D=1] &= E[y_0 \mid x=1, I_1=1, I_2=0, D=0] = 2 \\ E[y_0 \mid x=-1, I_1=1, I_2=0, D=1] &= \{-\mathbf{1}\} \neq E[y_0 \mid x=-1, I_1=1, I_2=0, D=0] = 0 \\ E[y_0 \mid x=1, I_1=0, I_2=1, D=1] &= E[y_0 \mid x=1, I_1=0, I_2=1, D=0] = -1 \\ E[y_0 \mid x=-1, I_1=0, I_2=1, D=1] &= E[y_0 \mid x=-1, I_1=0, I_2=1, D=0] = -3 \end{aligned}$$