Contamination Bias in Event Studies

Vasco Villas-Boas*

October 14, 2025

1 Contamination Bias in Event Studies

Consider the event study model given by

$$y_{it} = \alpha_i + \alpha_t + \gamma^{-10} d_{it}^{\leq -10} + \left(\sum_{\{k \in \mathbb{N}: -9 \leq k \leq 9,\ k \neq -1\}} \gamma^k d_{it}^k \right) + \gamma^{10} d_{it}^{\geq 10} + \epsilon_{it},\ \mathbb{E}\left[\epsilon_{it} \mid [d_i', \alpha_i, \alpha_t]'\right] = 0 \tag{1}$$

We define the variables by

- y_{it} is the outcome variable.
- α_t is a time fixed effect.
- α_i is an individual fixed effect.
- $d_{it}^k := \mathbb{1}_{\{t \text{ is } k \text{ periods relative to } i\text{'s absorbing treatment}\}}$.
- $d_{it}^{\leq -10} := \mathbb{1}_{\{t \text{ is 10 or more periods before to } i\text{'s absorbing treatment}\}}$.
- $d_{it}^{\geq 10} := \mathbb{1}_{\{t \text{ is 10 or more periods after to } i\text{'s absorbing treatment}\}}$.

Our goal is to understand why if we estimate this model by OLS when the lag effects are *homogeneous* across the population at every lag but there may be dynamic effects across lags, we get biased parameter estimates for the event study parameters (ie., γ^ks).

1.1 One Randomized Treatment Setting

Consider a simple linear regression model of an outcome Y_i on a single treatment $D_i \in \{0, 1\}$, a single binary control $W_i \in \{0, 1\}$, and an intercept:

$$Y_i = \alpha + \beta D_i + \gamma W_i + U_i,\ \mathbb{E}[U_i | D_i, W_i] = 0$$

We wish to interpret the coefficient β in terms of the causal effect of D_i on Y_i. For this, we use potential outcome notation letting $Y_i(d)$ be the test score of student i when $D_i = d$. Individual i's treatment effect is then given by $\tau_{1i} := Y_i(1) - Y_i(0)$ and we can write the realized outcome as $Y_i = Y_i(0) + \tau_{1i} D_i$. We assume that

*This is heavily based on Goldsmith-Pinkham et al. (2024).

$$(Y_i(0), Y_i(1)) \perp\!\!\!\perp D_i \mid W_i$$

1.1.1 Using Frisch-Waugh-Lovell (FWL) to Identify β

Define the following two residuals:[1]

$$\tilde{D}_i := D_i - \mathbb{L}[D_i \mid W_i]$$
$$\tilde{Y}_i := Y_i - \mathbb{L}[Y_i \mid W_i]$$

We can apply FWL to compute β:

$$\begin{aligned}
\beta &= \frac{\mathbb{E}[\tilde{D}_i\tilde{Y}_i]}{\mathbb{E}[\tilde{D}_i^2]}, \text{ by FWL} \\
&= \frac{\mathbb{E}[\tilde{D}_iY_i] - \mathbb{E}[\tilde{D}_i\mathbb{L}[Y_i \mid W_i]]}{\mathbb{E}[\tilde{D}_i^2]} \\
&= \frac{\mathbb{E}[\tilde{D}_iY_i] - \mathbb{E}[\cancelto{0}{\mathbb{E}[\tilde{D}_i \mid W_i]}\mathbb{L}[Y_i \mid W_i]]}{\mathbb{E}[\tilde{D}_i^2]}, \text{ by tower law of expectation and } \mathbb{L}[D_i \mid W_i] = \mathbb{E}[D_i \mid W_i] \text{ for binary } W_i \\
&= \frac{\mathbb{E}[\tilde{D}_iY_i]}{\mathbb{E}[\tilde{D}_i^2]} \\
&= \frac{\mathbb{E}[\tilde{D}_iY_i(0)]}{\mathbb{E}[\tilde{D}_i^2]} + \frac{\mathbb{E}[\tilde{D}_iD_i\tau_{1i}]}{\mathbb{E}[\tilde{D}_i^2]} \\
&= \frac{\mathbb{E}[\mathbb{E}[\tilde{D}_iY_i(0) \mid W_i]]}{\mathbb{E}[\tilde{D}_i^2]} + \frac{\mathbb{E}[\mathbb{E}[\tilde{D}_iD_i\tau_{1i} \mid W_i]]}{\mathbb{E}[\tilde{D}_i^2]}, \text{ by tower law of expectation} \\
&= \frac{\mathbb{E}[\cancelto{0}{\mathbb{E}[\tilde{D}_i \mid W_i]}\mathbb{E}[Y_i(0) \mid W_i]]}{\mathbb{E}[\tilde{D}_i^2]} + \frac{\mathbb{E}[\mathbb{E}[\tilde{D}_iD_i \mid W_i]\mathbb{E}[\tau_{1i} \mid W_i]]}{\mathbb{E}[\tilde{D}_i^2]}, \text{ by } D_i \perp\!\!\!\perp (Y_i(0), Y_i(1)) \mid W_i \\
&= \frac{\mathbb{E}[\mathrm{Var}(D_i \mid W_i)\mathbb{E}[\tau_{1i} \mid W_i]]}{\mathbb{E}[\tilde{D}_i^2]} \\
&= \left(\frac{\mathrm{Var}(D_i \mid W_i = 0)\Pr(W_i = 0)}{\mathbb{E}[\mathrm{Var}(D_i \mid W_i)]}\right)\mathbb{E}[\tau_{1i} \mid W_i = 0] + \left(\frac{\mathrm{Var}(D_i \mid W_i = 1)\Pr(W_i = 1)}{\mathbb{E}[\mathrm{Var}(D_i \mid W_i)]}\right)\mathbb{E}[\tau_{1i} \mid W_i = 1]
\end{aligned}$$

where the last step uses the fact that $\mathbb{E}[\tilde{D}_i^2] = \mathbb{E}[\mathrm{Var}(D_i \mid W_i)]$. This is a convex combination of treatment effects across strata of W_i. If we a-priori know there's no heterogeneity in treatment effects across strata, we can reduce the last line to $\beta = \mathbb{E}[\tau_{1i}] =: \tau_1$.

1.2 Two Randomized Treatments Setting

We now consider an augmented linear regression model with an outcome Y_i two binary and *mutually exclusive* treatments $X_{1i}, X_{2i} \in \{0, 1\}$ on a single binary control $W_i \in \{0, 1\}$, and an intercept:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \gamma W_i + U_i,\ \mathbb{E}[U_i \mid X_{i1}, X_{i2}, W_i] = 0 \tag{2}$$

[1]We define $\mathbb{L}[X \mid Y]$ to be the best linear predictor of X given Y in terms of mean squared error. That is $L[X \mid Y] = \bar{a} + \bar{b}Y$ where $(\bar{a}, \bar{b}) := \arg\min_{a,b} \mathbb{E}[(X - a + bY)^2]$.

We wish to interpret the coefficient β_1 in terms of the causal effect of X_{i1} on Y_i. For this, we use potential outcomes notation where $Y_i(d)$ for $d \in \{0, 1, 2\}$ means that individual i has treatment assignment d. Individual i's observed outcome can then be decomposed as $Y_i = Y_i(0) + \tau_{1i} X_{1i} + \tau_{2i} X_{2i}$ where $\tau_{1i} := Y_i(1) - Y_i(0)$ and $\tau_{2i} := Y_i(2) - Y_i(0)$. We assume that

$$(Y_i(0), Y_i(1), Y_i(2)) \perp\!\!\!\perp (X_{1i}, X_{2i}) \mid W_i$$

1.2.1 Using FWL to Identify β_1

Define the following two residuals:

$$\begin{aligned}
\tilde{X}_{1i} &:= X_{1i} - \mathbb{L}[X_{1i} \mid X_{2i}, W_i] \\
\tilde{Y}_i &:= Y_i - \mathbb{L}[Y_i \mid X_{2i}, W_i]
\end{aligned}$$

We can apply FWL to compute β_1:

$$\begin{aligned}
\beta_1 &= \frac{\mathbb{E}[\tilde{X}_{1i}\tilde{Y}_i]}{\mathbb{E}[\tilde{X}_{1i}^2]}, \text{ by FWL} \\
&= \frac{\mathbb{E}[\tilde{X}_{1i} Y_i]}{\mathbb{E}[\tilde{X}_{1i}^2]}, \text{ by similar algebra to above that uses tower law of expectation} \\
&= \frac{\mathbb{E}[\tilde{X}_{1i} Y_i(0)]}{\mathbb{E}[\tilde{X}_{1i}^2]} + \frac{\mathbb{E}[\tilde{X}_{1i} X_{1i}\tau_{1i}]}{\mathbb{E}[\tilde{X}_{1i}^2]} + \frac{\mathbb{E}[\tilde{X}_{1i} X_{2i}\tau_{2i}]}{\mathbb{E}[\tilde{X}_{1i}^2]} \\
&= \frac{\mathbb{E}[\mathbb{E}[\tilde{X}_{1i} Y_i(0) \mid W_i]]}{\mathbb{E}[\tilde{X}_{1i}^2]} + \frac{\mathbb{E}[\tilde{X}_{1i} X_{1i}\tau_{1i}]}{\mathbb{E}[\tilde{X}_{1i}^2]} + \frac{\mathbb{E}[\tilde{X}_{1i} X_{2i}\tau_{2i}]}{\mathbb{E}[\tilde{X}_{1i}^2]}, \text{ by tower law of expectation} \\
&= \frac{\mathbb{E}[\mathbb{E}[\tilde{X}_{1i} \mid W_i]\mathbb{E}[Y_i(0) \mid W_i]]}{\mathbb{E}[\tilde{X}_{1i}^2]} + \frac{\mathbb{E}[\tilde{X}_{1i} X_{1i}\tau_{1i}]}{\mathbb{E}[\tilde{X}_{1i}^2]} + \frac{\mathbb{E}[\tilde{X}_{1i} X_{2i}\tau_{2i}]}{\mathbb{E}[\tilde{X}_{1i}^2]}, \text{ by } Y_i(0) \perp\!\!\!\perp X_{1i}, X_{2i} \mid W_i \\
&= \frac{\mathbb{E}[\tilde{X}_{1i} X_{1i}\tau_{1i}]}{\mathbb{E}[\tilde{X}_{1i}^2]} + \frac{\mathbb{E}[\tilde{X}_{1i} X_{2i}\tau_{2i}]}{\mathbb{E}[\tilde{X}_{1i}^2]} \\
&= \frac{\mathbb{E}[\mathrm{Var}(D_i \mid W_i)\mathbb{E}[\tau_{1i} \mid W_i]]}{\mathbb{E}[\tilde{D}_i^2]} + \frac{\mathbb{E}[\tilde{X}_{1i} X_{2i}\tau_{2i}]}{\mathbb{E}[\tilde{X}_{1i}^2]}, \text{ by similar algebra to above}
\end{aligned}$$

where in the second to last step we use the fact that $\mathbb{E}[\tilde{X}_{1i} \mid W_i] = 0$ since the best linear predictor of a random variable with a binary covariate is equal to the conditional mean predictor. We wish to cancel out the second term above – the fact is that we cannot! The reason is that

$$\begin{aligned}
\mathbb{E}[\tilde{X}_{1i} \mid X_{2i}, W_i] &= \mathbb{E}[X_{1i} \mid X_{2i}, W_i] - \mathbb{E}[\mathbb{L}[X_{1i} \mid X_{2i}, W_i] \mid X_{2i}, W_i] \\
&= \mathbb{E}[X_{1i} \mid X_{2i}, W_i] - \mathbb{L}[X_{1i} \mid X_{2i}, W_i] \\
&\neq 0
\end{aligned}$$

It is generally true that $\mathbb{E}[X_{1i} \mid X_{2i}, W_i] \neq \mathbb{L}[X_{1i} \mid X_{2i}, W_i]$. The reason is that the dependence of X_{1i} on X_{2i} and W_i is non-linear. If $X_{2i} = 1$, then we know that $X_{1i} = 0$ but if $X_{2i} = 0$ then X_{1i} depends on W_i.

Again, note that this derivation considered potentially heterogeneous treatment effects across strata of the covariate W_i. If we a-priori know there's none, we can reduce the last to

$$\begin{aligned}\beta_1 &= \frac{\mathbb{E}[\mathrm{Var}(D_i \mid W_i)\mathbb{E}[\tau_{1i} \mid W_i]]}{\mathbb{E}[\tilde{D}_i^2]} + \frac{\mathbb{E}[\tilde{X}_{1i}X_{2i}\tau_{2i}]}{\mathbb{E}[\tilde{X}_{1i}^2]} \\ &= \tau_1 + \left(\frac{\mathbb{E}[\tilde{X}_{1i}X_{2i}]}{\mathbb{E}[\tilde{X}_{1i}^2]}\right)\tau_2\end{aligned}$$

This means that even under the presence of homogeneous treatment effects, we have a biased estimate for β_1.

1.2.2 Thinking more about OLS

Suppose that we observe data $\mathcal{D} := \{(Y_i, X_{1i}, X_{2i}, W_i)\}_{i=1}^n$ from the model in Equation 2 where each of the datapoints is iid. We estimate $\theta = [\alpha, \beta_1, \beta_2, \gamma]'$ with OLS under the assumption:

$$\hat{\mathbb{E}}[[1, X_{1i}, X_{2i}, W_i]'U_i] = 0$$

We know that in the population $\mathbb{E}[[1, X_{1i}, X_{2i}, W_i]'U_i] = 0$ by our conditional mean assumption in Equation 2 so why are we not estimating what we want in β_1?

When identifying the parameters using this projection assumption (and under the assumption of homogeneity of the effects), you are literally defining

$$\beta_1 = \tau_1 + \left(\frac{\mathbb{E}[\tilde{X}_{1i}X_{2i}]}{\mathbb{E}[\tilde{X}_{1i}^2]}\right)\tau_2$$

So in identification you are implicitly defining that β_1 is this combination of the treatment effects of the two treatments, which is not desirable!

1.3 Returning to Event Study Regression

In light of the analysis here, we can see that the lag variables d_{it}^k non-linearly determine each other in the presence of the fixed effects in Equation 1. As a result, estimates for the "desired lagged effects" will be biased by the "effects of the other lags". Further, under the conditional mean assumption of Equation 1, if you identify the parameters using the simple projection, you will literally be defining your parameters to take on these undesired values that are biased for the causal effects you wish to identify.

REFERENCES

Paul Goldsmith-Pinkham, Peter Hull, and Michal Kolesár. Contamination bias in linear regressions. *American Economic Review*, 114(12):4015–51, December 2024. URL: `https://www.aeaweb.org/articles?id=10.1257/aer.20221116`, `doi:10.1257/aer.20221116`.