Panel Data and Program Evaluation

Zhiyuan Chen

Department of Trade Economics
Remin Business School

November 18, 2020

A Brief Review of OLS

Model, Consistency, and Asymptotics

The model we consider is:

$$Y_i = X_i'\beta + u_i$$

- The data is iid, $i = 1, 2, \cdots, N$
- $\mathbf{E}(X_i u_i) = 0$
- no perfect multi-collinearity

The OLS estimator is given by

$$\begin{aligned}\hat{\beta} &= \left(\frac{1}{N}\sum_i X_i X_i'\right)^{-1}\left(\frac{1}{N}\sum_i X_i Y_i\right)\\ &= \beta + \left(\frac{1}{N}\sum_i X_i X_i'\right)^{-1}\frac{1}{N}\sum_i X_i u_i\end{aligned}$$

- Consistency:

$$\lim_{N \to_p \infty} \hat{\beta} = \beta + \lim_{N \to_p \infty} \left(\frac{1}{N} \sum_i X_i X_i' \right)^{-1} \frac{1}{N} \sum_i X_i u_i$$
$$= \beta + \left[\mathbf{E}\left(X_i X_i' \right) \right]^{-1} \mathbf{E}\left(X_i u_i \right) = \beta$$

- Asymptotic distribution:

$$\sqrt{N} \left(\hat{\beta} - \beta \right) \sim \mathbf{N} \left(0, \left[\mathbf{E}\left(X_i X_i' \right) \right]^{-1} \mathbf{E}\left(X_i X_i' u_i^2 \right) \left[\mathbf{E}\left(X_i X_i' \right) \right]^{-1} \right)$$

- Sample approximation:

$$Var\left(\hat{\beta} \right) \approx \left(\sum_i X_i X_i' \right)^{-1} \left[\sum_i X_i X_i' \hat{u}_i^2 \right] \left(\sum_i X_i X_i' \right)^{-1}$$

Panel Data

- Now we have data for N units (people, products, firms, counties, cities, provinces, countries ...)
- For each unit we have T_i (≥ 1) observations
  - T_i typically refers to time periods
  - Could be products of a firm, students in a classroom, counties in a province
- We assume that:
  - N is large; consistency exists as N grows
  - T_i is small; T_i stays constant as N grows

Now let's consider the regression model:

$$Y_{it} = X_{it}'\beta + u_{it}$$

- *Assumption 1*: assume independence **across individuals** but **not across time**
- *Assumption 2:* $\mathbf{E}(X_{it}u_{it}) = 0$

Consistency:

$$\begin{aligned}\hat{\beta} &= \beta + \left(\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} X_{it}X_{it}'}{\sum_{i=1}^{N} T_i}\right)^{-1}\left(\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_i} X_{it}u_{it}}{\sum_{i=1}^{N} T_i}\right)\\ &\rightarrow_p \beta\end{aligned}$$

Serial Correlation

- The assumption of NO serial correlation is problematic so the asymptotic variance from before is not right
- Error terms are uncorrelated across individuals, but not within individuals
- Formally speaking, we assume that:

$$\mathbf{E}(u_{it}u_{jt}) = 0, \forall i \neq j$$
$$\mathbf{E}(u_{it}X_{it}) = 0$$

- However, the assumption of OLS is not satisfied because:

$$\mathbf{E}(u_{it}u_{is}) = 0$$

  is a CRAZY assumption!
- Ignoring the serial correlation tends to **underestimate** the size of standard errors

Clustered Standard Errors

- Recall that

$$\hat{\beta} - \beta = \left(\sum_i^N \sum_t^{T_i} X_{it} X_{it}' \right)^{-1} \sum_i^N \sum_t^{T_i} X_{it} u_{it}$$

  - Define $\eta_i \equiv \sum_{t=1}^{T_i} X_{it} u_{it}$, then η_i is iid
  - $var(\eta_i) = V_\eta$

- Then we get

$$\sqrt{N}\left(\hat{\beta} - \beta\right) = \left(\frac{1}{N} \sum_i \sum_t X_{it} X_{it}' \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \eta_i \right)$$

- Using the Central Limit Theorem we know

$$\sqrt{N}\left(\hat{\beta} - \beta\right) \sim \mathcal{N}\left(0, \left[\mathbf{E}\left(\sum_{t=1}^{T_i} X_{it} X_{it}' \right)^{-1} \right] V_\eta \left[\mathbf{E}\left(\sum_{t=1}^{T_i} X_{it} X_{it}' \right) \right]^{-1} \right)$$

We can approximate the variance of $\hat{\beta}$ as

$$Var\left(\hat{\beta}\right) \approx \left(\sum_i \sum_t X_{it} X_{it}'\right)^{-1} \left[\sum_i \left(\sum_t X_{it} \hat{u}_{it}\right) \left(\sum_t X_{it}' \hat{u}_{it}\right)\right] \left(\sum_i \sum_t X_{it} X_{it}'\right)^{-1}$$

- This is a generalization of the heteroskedastic robust standard errors
- We are allowing $[X_{i1} u_{i1}, X_{i2} u_{i2}, \cdots, X_{iT_i} u_{iT_i}]$ to have an arbitrary variance-covariance matrix

In STATA, this is done by coding `reg y x, cluster(i)`

- i is the id of group in which the errors are serially correlated

Random Effects vs. Fixed Effects

- Panel data enable us to take care of the idiosyncratic component in the error term
- Write the linear model as

$$Y_{it} = X'_{it}\beta + \theta_i + \varepsilon_{it}$$

  - In both of the models, we assume that

$$\mathbf{E}\left(\varepsilon_{it} X_{it}\right) = 0$$

  - In the Random Effects (RE) model, we assume that

$$\mathbf{E}\left(\theta_i X_{it}\right) = 0$$

  - In the Fixed Effects (FE) model, we do not assume anything about the relationship between θ_i and X_i! θ_i is an idiosyncratic constant.

Consistent Estimators for FE models

- Include individual dummies and run the regression:

$$Y_{it} = X_{it}\beta + D_{it}'\theta + \varepsilon_{it}$$

where $D_{it} = \left[D_{it}^{(j)}\right]$, $D_{it}^{(j)} = 1$ if $i = j$ and zero otherwise

  - Conceptually this is a different model, but technically it delivers consistent estimates for β

- **Standard FE (Mean-Differencing)** model:

$$\left(Y_{it} - \bar{Y}_i\right) = \left(X_{it} - \bar{X}_i\right)' \beta + \left(\varepsilon_{it} - \bar{\varepsilon}_i\right)$$

where $\bar{Z}_i \equiv \frac{1}{T_i} \sum_{t=1}^{T_i} Z_{it}$

- **First-Differencing (FD)** model:

$$Y_{it} - Y_{it-1} = \left(X_{it} - X_{it-1}\right)' \beta + \varepsilon_{it} - \varepsilon_{it-1}$$

  - With 2 periods, this is equivalent to the standard FE model (verify it by yourself)

Fixed Effects vs. First Differencing

More generally, we consider the model

$$Y_{it} = \beta \tau_{it} + \theta_i + \varepsilon_{it}$$

Assume that $T_i = T$ for everyone, for everyone the only regressor τ_{it} is given by

$$\tau_{it} = \begin{cases} 0 & t \leq t_0 \\ 1 & t > t_0 \end{cases}$$

- τ_{it} can represent some macro-level program (trade liberalization, 5-year plan, expansion of college enrollment...) begins at $t_0 + 1$
- The FE and FD estimators are:

$$FE : \hat{\beta}_{FE} = \frac{\sum_{i=1}^{N} \sum_{t=t_0+1}^{T} Y_{it}}{N(T - t_0)} - \frac{\sum_{i=1}^{N} \sum_{t=1}^{t_0} Y_{it}}{N t_0}$$

$$FD : \hat{\beta}_{FD} = \frac{\sum_{i=1}^{N} (Y_{it_0+1} - Y_{it_0})}{N}$$

Fixed Effects vs. Pooled Regression

The fixed effects estimator is NOT always better than pooled OLS regression

- The sample variance of the data is:

$$ss(X_{it}) = \sum_i \sum_t (X_{it} - \bar{X})^2$$
$$= \underbrace{\sum_i \sum_t (X_{it} - \bar{X}_i)^2}_{within\,variance} + \underbrace{\sum_i T_i (\bar{X}_i - \bar{X})^2}_{between\,variance}$$

FE estimator only uses the within variance of X_i:

1. It is inefficient. Standard errors are very large when X_{it} does not change
2. It can even a worse bias:

$$\hat{\beta}_{POLS} = \beta + \frac{cov(X_{it}, \theta_i + \varepsilon_{it})}{var(X_{it})}$$
$$\hat{\beta}_{FE} = \beta + \frac{cov(X_{it} - \bar{X}_i, \varepsilon_{it} - \bar{\varepsilon}_i)}{var(X_{it} - \bar{X}_i)}$$

Birth weight and health

Reference: Almond, Chay, and Lee (QJE, 2005)

- Their goal is to estimate the effects of birth weight on health:

$$h_{ij} = \alpha + \beta\, bw_{ij} + X_i' \gamma + a_i + \varepsilon_{ij}$$

where

  - h_{ij}-health of newborn j of mother i
  - bw_{ij} is birth weight
  - $a_i$is mother specific effect

- The pooled OLS regression of h_{ij}on bw_{ij} gives us

$$\hat{\beta}_{POLS} = \beta + \frac{\text{cov}\left(bw_{ij}, X_i'\gamma\right)}{\text{var}\left(bw_{ij}\right)} + \frac{\text{cov}\left(bw_{ij}, a_i\right)}{\text{var}\left(bw_{ij}\right)}$$

- Their clever solution is to use twins:
  - Twins share the same mother, so a_i effectively controls for race, age, education, family background...
  - Estimate model as

    $$\Delta h_{ij} = \beta \Delta bw_{ij} + \Delta \varepsilon_{ij}$$

    where they assume that $\text{cov}\left(\Delta bw_{ij}, \Delta \varepsilon_{ij}\right) = 0$

TABLE III

POOLED OLS AND TWINS FIXED EFFECTS ESTIMATES OF THE EFFECT OF BIRTH WEIGHT

Birth weight coefficient	Including congenital anomalies		Excluding congenital anomalies	
	Pooled OLS	Fixed effects	Pooled OLS	Fixed effects
Hospital costs	−29.95	−4.93	—	—
(in 2000 dollars)	(0.84)	(0.44)	—	—
	[−0.506]	[−0.083]	—	—
Adj. R^2	0.256	0.796	—	—
Sample size	44,410	44,410	—	—
Mortality, 1-year	−0.1168	−0.0222	−0.1069	−0.0082
(per 1000 births)	(0.0016)	(0.0016)	(0.0017)	(0.0012)
	[−0.412]	[−0.078]	[−0.377]	[−0.029]
Adj. R^2	0.169	0.585	0.164	0.629
Sample size	189,036	189,036	183,727	183,727
Mortality, 1-day	−0.0739	−0.0071	−0.0675	−0.0003
(per 1000 births)	(0.0015)	(0.0010)	(0.0015)	(0.0006)
	[−0.357]	[−0.034]	[−0.326]	[−0.001]
Adj. R^2	0.132	0.752	0.127	0.809
Sample size	189,036	189,036	183,727	183,727
Mortality, neonatal	−0.105	−0.0154	−0.0962	−0.0041
(per 1000 births)	(0.0016)	(0.0013)	(0.0016)	(0.0008)
	[−0.415]	[−0.061]	[−0.38]	[−0.016]
Adj. R^2	0.173	0.683	0.169	0.745
Sample size	189,036	189,036	183,727	183,727
5-min. APGAR score	0.1053	0.0117	0.1009	0.0069
(0–10 scale,	(0.0011)	(0.0012)	(0.0011)	(0.0011)
divided by 100)	[0.506]	[0.056]	[0.485]	[0.033]
Adj. R^2	0.255	0.663	0.248	0.673
Sample size	159,070	159,070	154,449	154,449
Ventilator incidence	−0.0837	−0.0039	−0.081	−0.002
(per 1000 births)	(0.0015)	(0.0017)	(0.0015)	(0.0016)
	[−0.228]	[−0.011]	[−0.221]	[−0.005]
Adj. R^2	0.052	0.706	0.05	0.716
Sample size	189,036	189,036	183,727	183,727
Ventilator ≥30 min.	−0.0724	0.0006	−0.0701	0.0016
(per 1000 births)	(0.0013)	(0.0013)	(0.0014)	(0.0012)
	[−0.252]	[0.002]	[−0.244]	[0.006]
Adj. R^2	0.063	0.724	0.062	0.739
Sample size	189,036	189,036	183,727	183,727

See notes to Tables I and II. The data come from the 1989–1991 Linked Birth-Infant Death Detail Files and the 1995–2000 HCUP Inpatient Database for New York and New Jersey. The first two columns use samples that include twin pairs in which one or both twins either had a congenital anomaly at birth or whose cause of death was a congenital anomaly. The second two columns exclude these twin pairs from the analysis. The HCUP data do not contain information on congenital anomalies. The standard errors are in parentheses and are corrected for heteroskedasticity and within-twin-pair correlation in the residuals. For APGAR score, the coefficients are scaled up by 100. Numbers in square brackets indicate effect size in terms of standard deviations of the outcome per one standard deviation in birth weight (667 grams). There are no other variables included in the regressions.

Difference in Differences

Simple policy evaluation:

- Data on a bunch of people right before and after the policy is enacted
- Two years of data 0 and 1 and that the policy is enacted in between
- We can simply identify the effect as:

$$\hat{\alpha} = \bar{Y}_1 - \bar{Y}_0$$

We could justify this using a FE model:

$$Y_{it} = \alpha_0 + \alpha \tau_{it} + \theta_i + \varepsilon_{it}$$

where

$$\tau_{it} = \begin{cases} 0 & t = 0 \\ 1 & t = 1 \end{cases}$$

We assume that $\mathbf{E}(\tau_{it}\varepsilon_{it}) = 0$ but no particularly assumption on θ_i

In the two-period case, FE model is equivalent to FD model, which is

$$Y_{i1} - Y_{i0} = \alpha + u_{i1} - u_{i0}$$

The estimator for the policy effect is

$$\hat{\alpha} = \bar{Y}_{i1} - \bar{Y}_{i0}$$

- We assume no other changes in time between and attribute whatever that is to the program
- Add time dummies into the model the treatment effect is not separated from the policy's impact

After adding time dummy variables, the model becomes

$$Y_{it} = \alpha_0 + \alpha\tau_{it} + \delta t + \theta_i + \varepsilon_{it}$$

Then the difference estimator delivers

$$\begin{aligned} \mathbf{E}(Y_{i1} - Y_{i0}) &= \mathbf{E}\left[(\alpha_0 + \alpha + \delta + \theta_i + \varepsilon_{i1}) - (\alpha_0 + \theta_i + \varepsilon_{i0})\right] \\ &= \alpha + \delta \end{aligned}$$

- The Diff-in-Diff estimator is designed to solve this problem

Instead of having one group, now we have two groups:

1. People who are affected by the policy changes (treated), with outcome Y_{it}^1
2. People who are not affected by the policy changes (controls), with outcome Y_{it}^0

Still two time periods: before ($t = 0$) and after ($t = 1$)
We can think:

1. Using the treated to pick up the time changes and policy effects:

$$\hat{\alpha} + \hat{\delta} = \bar{Y}_{i1}^1 - \bar{Y}_{i0}^1$$

2. Under the **common trend** assumption controls pick up the time changes:

$$\hat{\delta} = \bar{Y}_{i1}^0 - \bar{Y}_{i0}^0$$

We can then estimate the policy effect as a difference in difference

$$\hat{\alpha} = \left(\bar{Y}_1^1 - \bar{Y}_0^1\right) - \left(\bar{Y}_1^0 - \bar{Y}_0^0\right)$$

Formally, we can write the DGP (data generating process) for the DID estimator as

$$Y_{it} = \alpha_0 + \alpha \tau_{it}^{g_i} + \delta t + \theta_i + \varepsilon_{it}$$

where

$$\tau_{it}^{g} = \begin{cases} 1 & t = 1, g_i = 1 \\ 0 & otherwise \end{cases}$$

τ_{it}^{g} is usually written as $\tau_{it} \times t$

- Now we run a fixed effect regression and get a consistent estimate of the treatment effects α

Recall that the FE estimator for two-periods data is equivalent to the FD estimator:

$$\hat{\alpha} = \frac{\sum_{i=1}^{N}\left[\left(\tau_{i1}^{g_i} - \tau_{i0}^{g_i}\right) - \overline{\left(\tau_{i1}^{g_i} - \tau_{i0}^{g_i}\right)}\right]\left(Y_{i1} - Y_{i0}\right)}{\sum_{i=1}^{N}\left[\left(\tau_{i1}^{g_i} - \tau_{i0}^{g_i}\right) - \overline{\left(\tau_{i1}^{g_i} - \tau_{i0}^{g_i}\right)}\right]^2}$$
$$= \left(\bar{Y}_1^1 - \bar{Y}_0^1\right) - \left(\bar{Y}_1^0 - \bar{Y}_0^0\right)$$

- Notice that

$$\overline{\left(\tau_{i1}^{g_i} - \tau_{i0}^{g_i}\right)} = \frac{1}{N_1 + N_0}\sum_i \left(\tau_{i1}^{g_i} - \tau_{i0}^{g_i}\right)$$
$$= \frac{1}{N_1 + N_0}\left[\sum_{\{i:\, g_i = 1\}} \left(\tau_{i1}^{1} - \tau_{i0}^{1}\right) + \sum_{\{i:\, g_i = 0\}} \left(\tau_{i1}^{0} - \tau_{i0}^{0}\right)\right]$$
$$= \frac{N_1}{N_1 + N_0}$$

In principle, you don't need panel data to implement DID regression; repeated cross section data are fine!

- In general, we can write the regression as

$$Y_i = \alpha_0 + \alpha \tau_{t_i}^{g_i} + \delta t_i + \gamma g_i + \varepsilon_i$$

  - where g_i is the group indicator of person i and t_i is the time period in which i exists in the data
  - $\tau_{t_i}^{g_i} = g_i \times t_i$
- We have 4 categories of people: (before, treated), (before, controls), (after, treated), (after, controls)
- We end up with having 4 groups of moment conditions:

$$\begin{aligned} \mathbf{E}(\varepsilon_i) &= 0 \\ \mathbf{E}\left(\tau_{t_i}^{g_i} \varepsilon_i\right) &= 0 \\ \mathbf{E}(t_i \varepsilon_i) &= 0 \\ \mathbf{E}(g_i \varepsilon_i) &= 0 \end{aligned}$$

Using

$$Y_i = \hat{\alpha}_0 + \hat{\alpha}\tau_{t_i}^{g_i} + \hat{\delta} t_i + \hat{\gamma} g_i + \hat{\varepsilon}_i$$

We can rewrite the sample analog of the moment conditions as

$$\begin{aligned}
\bar{Y}_0^1 &= \hat{\alpha}_0 + \hat{\gamma} \\
\bar{Y}_1^1 &= \hat{\alpha}_0 + \hat{\alpha} + \hat{\delta} + \hat{\gamma} \\
\bar{Y}_0^0 &= \hat{\alpha}_0 \\
Y_1^0 &= \hat{\alpha}_0 + \hat{\delta}
\end{aligned}$$

We can solve for the parameters as

$$\begin{aligned}
\hat{\alpha}_0 &= \bar{Y}_0^0 \\
\hat{\gamma} &= \bar{Y}_0^1 - \bar{Y}_0^0 \\
\hat{\delta} &= \bar{Y}_0^1 - \bar{Y}_0^0 \\
\hat{\alpha} &= \left(Y_1^1 - \bar{Y}_0^1\right) - \left(\bar{Y}_1^0 - \bar{Y}_0^0\right)
\end{aligned}$$

More generally, we can add more control variables and specify the DID model as

$$Y_i = \alpha\tau_{t_i}^{g_i} + X_{it}'\beta + \delta_{t_i} + \theta_{g_i} + \varepsilon_i$$

This setting is simple yet powerful. There are many papers that do this basic stuff.

- Eissa and Liebman (QJE, 1996): Estimate the effect of the earned income tax credit on labor supply of women
- Dohahue and Levitt (QJE, 2001): Impact of legalized abortion on crime

Event Studies

- The treatment effects could be dynamic, with short-run effects differing from long-run effects
- We can easily extend the baseline model to allow for this:

$$Y_i = \beta_0 + \sum_{j=0}^{J} \alpha_j \tau_{t_i - j}^{g_i} + \delta_{g_i} + \gamma_{t_i} + \varepsilon_i$$

where

$$\tau_t^g = \begin{cases} 1 & g_i = 1 \text{ and policy started in year } t \\ 0 & \text{otherwise} \end{cases}$$

Common Trend Assumption

- Recall that the model we specified for DID is:

  $$Y_i = \beta_0 + \alpha \tau_{t_i}^{g_i} + \delta t_i + \gamma g_i + \varepsilon_i$$

  The DID estimator is

  $$\hat{\alpha}_{DID} = \alpha + \left(\bar{\varepsilon}_1^1 - \bar{\varepsilon}_0^1\right) - \left(\bar{\varepsilon}_1^0 - \bar{\varepsilon}_0^0\right)$$

  To obtain a consistent estimator, we require that

  $$\mathbf{E}\left[\left(\bar{\varepsilon}_1^1 - \bar{\varepsilon}_0^1\right) - \left(\bar{\varepsilon}_1^0 - \bar{\varepsilon}_0^0\right)\right] = 0$$

- The treated units can have different *levels* of the error term, but the *change* in the error term must be random
- Two approaches to validate the common trend assumption: **placebo policies** and **add group-specify time trends**

Placebo Policies

- A popular strategy for robustness check: if a policy was enacted in say 2010 you could pretend it was enacted in 2005 in the same place and then only use data through 2009
- The easiest (and most common) is in the Event framework: include leads as well as lags in the model (Bertrand, Duflo, Mullainathan, 2004, QJE)
- To implement it you can run a regression like:

$$Y_i = \beta_0 + \sum_{j=-\tilde{J}, j\neq -1}^{J} \alpha_j \tau_{t_i-j}^{g_i} + \delta_{g_i} + \rho_{t_i} + \varepsilon_i$$

  where $t_i = -\tilde{J}, \cdots, J$ and the period of enactment to be zero

  - $\tau_{t_i-j}^{1} = 1$ whenever $t_i - j = 0$
  - Note $\sum_{j=-\tilde{J}}^{J} \tau_{t_i-j}^{g_i} = \delta_{g_i}$, so we drop one period to avoid perfect co-linearity

If we normalize $\alpha_{-1} = 0$, α_{-2} is estimated as

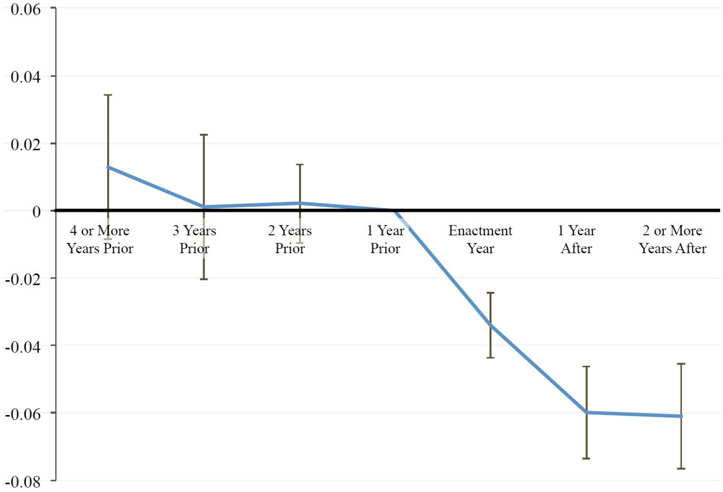
$$\hat{\alpha}_{-2} = \left(\bar{Y}^{1}_{-2} - \bar{Y}^{1}_{-1}\right) - \left(\bar{Y}^{0}_{-2} - \bar{Y}^{0}_{-1}\right)$$

where

$$\begin{aligned}
\bar{Y}^{1}_{-2} &= \hat{\beta}_0 + \hat{\alpha}_{-2} + \hat{\delta} + \hat{\rho}_{-2} \\
\bar{Y}^{1}_{-1} &= \hat{\beta}_0 + \hat{\delta} + \hat{\rho}_{-1} \\
\bar{Y}^{0}_{-2} &= \hat{\beta}_0 + \hat{\rho}_{-2} \\
\bar{Y}^{0}_{-1} &= \hat{\beta}_0 + \hat{\rho}_{-1}
\end{aligned}$$

- $\hat{\alpha}_{-2}$ should be zero under the common trend assumption
- $\hat{\alpha}_{-j} (j \geq 2)$ should also be zero under the common trend assumption

Figure: Typical event studies for the placebo policies



Group-Specific Time Trends

- One might be worried that units that are trending up or trending down are more likely to change policy
- One can include group×time dummy variables in the model to fix this problem
- Now let's consider the three-period case where the policy happens between period 1 and 2, the model becomes

$$Y_i = \beta_0 + \alpha \tau_{t_i}^{g_i} + \delta_0 t_i \times (1 - g_i) + \delta_1 t_i \times g_i + \delta_2 \mathbb{I}(t_i = 2) + \gamma g_i + \varepsilon_i$$

- We can write the estimator of α as a Triple Difference (DDD):

$$\begin{aligned}
\hat{\alpha}_{DDD} &= \left(\bar{Y}_2^1 - \bar{Y}_1^1\right) - \left(\bar{Y}_2^0 - \bar{Y}_1^0\right) - \left[\left(\bar{Y}_1^1 - \bar{Y}_0^1\right) - \left(\bar{Y}_1^0 - \bar{Y}_0^0\right)\right] \\
&\approx (\alpha + \delta_1 + \delta_2) - (\delta_0 + \delta_2) - (\delta_1 - \delta_0) \\
&= \alpha
\end{aligned}$$

If we do not add the group-specific time trends, the error term is a composite:

$$\tilde{\varepsilon}_i = \varepsilon_i + \delta_0 t_i \times (1 - g_i) + \delta_1 t_i \times g_i - \delta t_i$$

We go back to the common trend assumption and get

$$\begin{aligned}
\mathbf{E}\left(\left(\bar{\tilde{\varepsilon}}_1^1 - \bar{\tilde{\varepsilon}}_1^0\right) - \left(\bar{\tilde{\varepsilon}}_1^0 - \bar{\tilde{\varepsilon}}_0^0\right)\right) &= \mathbf{E}\left(\left(\bar{\varepsilon}_1^1 - \bar{\varepsilon}_1^0\right) - \left(\bar{\varepsilon}_1^0 - \bar{\varepsilon}_0^0\right)\right) \\
&\quad + (\delta_1 - \delta) - (\delta_0 - \delta) \\
&= \mathbf{E}\left(\left(\bar{\varepsilon}_1^1 - \bar{\varepsilon}_1^0\right) - \left(\bar{\varepsilon}_1^0 - \bar{\varepsilon}_0^0\right)\right) + \delta_1 - \delta_0 \\
&\approx \delta_1 - \delta_0
\end{aligned}$$

Inference: Get the Right Standard Errors

- In applications, we often have individual data (people, firms...) and the policy is enacted at more aggregated level (cities, provinces)
- It seems that we have more data than we need

Consider the individual-level model

$$Y_i = \alpha \tau_{t_i}^{g_i} + Z_i' \delta + X_{g_i t_i}' \beta + \theta_{g_i} + \gamma_{t_i} + u_i$$

One could aggregate the data by group and year:

$$\bar{Y}_{gt} = \alpha \bar{\tau}_{gt} + Z_{gt}' \delta + X_{gt}' \beta + \theta_g + \gamma_t + \bar{u}_{gt}$$

The second regression gives consistent estimates for the treatment effects

- HOWEVER, we obtain different standard errors, which one should we use?

- The error term for the individual level model is

$$u_i = \eta_{g_i t_i} + \varepsilon_i$$

  - The i.i.d assumption is violated is $\eta_{g_i t_i}$ exists
  - A standard solution is to cluster the standard error by group$\times$year to allow for arbitrary correlation within a group
- Bertrand, Duflo, and Mullainathan (QJE, 2004):
  - If there is serial correlation in η_{gt}, cluster by group$\times$year is problematic
- Solutions:
  - # of groups is large: block bootstrap by state works well
  - moderate # of states: asymptotic approximation of the variance-covariance matrix
  - small # of states: collapse the data into a "pre"- and "post"-period and account for the effective sample size
- A practioner's guide is offered by Cameron and Miller (2015, JHR)

Recent Developments

- Synthetic control estimator by Abadie, Diamond, and Hainmueller (2010, JASA):

$$\hat{\alpha}_t = Y_t^1 - \sum_{i=2}^{I+1} \omega_i^* Y_{it}^*$$

  - $\omega^* = \left(\omega_2^*, \cdots, \omega_{I+1}^*\right)$ is chosen to minimize $\|X_1 - X_0 \omega\|$
- Synthetic Difference in Differences estimator by Arkhangelsky et al. (2019, NBER working paper):