Borusyak, Jaravel, Spiess (2022). Revisiting Event Study Designs: Robust and Efficient Estimation

sasha petrov

May 2023

Plan

- How does the estimator work?
- Why is/can it be better than comparable alternatives?
- When is it better that comparable alternatives?

Table of Contents

Setting: Definitions

- The set of units, indexed by i, get treated in a staggered manner;
- The average treatment effect in period t for unit i, that was treated at some $\tilde{t} \leq t$:

$$\tau_{it} = \mathbb{E}\left[Y_{it} - Y_{it}(0)\right] \tag{1}$$

- The set of ATTs:

$$\tau = \{\tau_{it}\}_{it \in \Omega_1} \tag{2}$$

- The target average of ATTs weighted by pre-chosen weights:

$$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it} \equiv w_1' \tau \tag{3}$$

▸ Notation summary

Setting: Assumptions

Assumption 1 (A1). Parallel trends:

$$\mathbb{E}\left[Y_{it}(0)\right] = \alpha_i + \beta_t, \quad \forall it \in \Omega \tag{4}$$

Assumption 2 (A2). No anticipation:

$$Y_{it} = Y_{it}(0), \quad \forall it \in \Omega_0 \tag{5}$$

Assumption 3 (A3). Model for Treatment Effects Structure

Assumption 3 (A3). An optional pre-chosen restriction on treatment effects:

$$\tau = \Gamma\theta \tag{6}$$

Equivalently:

$$B\tau = \theta \tag{7}$$

NB: Allows for no restrictions (corresponds to $\Gamma = B = \mathbf{I}$)

Illustration of causal effects restriction – 1

- Consider a simple case: $t \in \{1, 2, 3, 4, 5\}$, $i \in \{1, 2, 3\}$;
- The realized treatment allocation:

$$\Omega_1 = \{22, 23, 24, 25, 34, 35\} \tag{8}$$

- This means:
  - Unit 1 never got treated;
  - Unit 2 got treated at $t = 2$;
  - Unit 3 got treated at $t = 4$;
- The vector of ATTs:

$$\tau = \begin{bmatrix} \tau_{22} \\ \tau_{23} \\ \tau_{24} \\ \tau_{25} \\ \tau_{34} \\ \tau_{35} \end{bmatrix} \tag{9}$$

Illustration of causal effects restriction – 2

- Assumption: No heterogeneity in effects across time:

$$\tau_{22} = \tau_{23} = \tau_{24} = \tau_{25} = \theta_1, \quad \tau_{34} = \tau_{35} = \theta_2 \tag{10}$$

- Therefore, only 2 free parameters:

$$\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} \tag{11}$$

- Mapping back to τ's:

$$\tau = \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_{\Gamma} \theta \tag{12}$$

Illustration of causal effects restriction – 3

- Corresponding constraint on τ:

$$\underbrace{\begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}}_{B} \tau = \mathbf{0} \tag{13}$$

Specification of the empirical model

- General model of $Y(0)$:
$$\mathbb{E}\left[Y_{it}(0)\right] = A'_{it}\lambda_i + X'_{it}\delta \tag{14}$$
  - X_{it} – covariates that have common-across-units effects;
    - Nests time FEs (consider $X_{it} = t$);
  - A_{it} – covariates that have unit-specific effects;
    - Nests unit FEs (consider $A^{i'}_{it} = \mathbf{1}\{i = i'\}$) and unit-specific trends (consider $A_{it} = t$);
  - *NB:* Either of these can include time-varying covariates;
    - Keep in mind the ‘bad control’ risk.

Estimator Implementation

1. Estimate $\hat{\theta}$ using the following regression:

$$Y_{it} = A'_{it}\lambda_i + X'_{it}\delta + D_{it}(\Gamma'\theta)_{it} + \varepsilon_{it} \tag{15}$$

2. Apply the matrix from the treatment effects model (6):

$$\hat{\tau} = \Gamma\hat{\theta} \tag{16}$$

3. Apply the pre-chosen weights (3):

$$\hat{\tau}_w = w'_1\hat{\tau} \tag{17}$$

▸ Notation summary

Inference

The finite-sample variance of $\hat{\tau}_w$ is:

$$\sigma_w^2 = \mathbb{E}\left[\sum_i \left(\sum_{t;it\in\Omega} v_{it}\varepsilon_{it}\right)^2\right] \tag{18}$$

where v_{it} are regression weights for Y_{it}:

$$\hat{\tau}_w = \sum_{it\in\Omega} v_{it} Y_{it} \tag{19}$$

For conservative inference, impose an aggregation structure on treatment effects:

$$\Omega_1 = \cup_g G_g, \quad \tau_{it} \equiv \tau_g \quad \forall it \in G_g \tag{20}$$

▸ Notation summary

Conservative estimation of variance

1. Estimate the aggregated effect for each group:

$$\tilde{\tau}_g = \frac{\sum_i \left(\sum_{t;it\in G_g} v_{it}\right)\left(\sum_{t;it\in G_g} v_{it}\hat{\tau}_{it}\right)}{\sum_i \left(\sum_{t;it\in G_g} v_{it}\right)^2} \tag{21}$$

2. Compute residuals using $\tilde{\tau}_g$ (not $\hat{\tau}_{it}$ as it leads to a biased estimate of σ_w):

$$\tilde{\varepsilon}_{it} = Y_{it} - A'_{it}\hat{\lambda}_i - X'_{it}\hat{\delta} - D_{it}\tilde{\tau}_{g(it)} \tag{22}$$

3. Plug into the variance formula (18):

$$\hat{\sigma}_w^2 = \sum_i \left(\sum_{t;it\in\Omega} v_{it}\tilde{\varepsilon}_{it}\right)^2 \tag{23}$$

▸ Notation summary

Table of Contents

Advertised advantages of the estimator

- Necessity to explicitly pre-choose the estimand;
- Robustness to common identification issues in TWFE regressions:
  - Spurious identification of long-run effects;
  - Under-identification of ATTs if a never-treated unit is absent;
- Efficiency among linear estimators;
  - Analytically proven for the homoskedastic case;
- Tighter CIs under suitable inference compared to existing robust estimators;
  - Shown in simulations.

Summary of simulation results

- In many scenarios, the variance is smaller than that of available alternatives;
- However, the coverage rate of CIs from the alternatives is mostly as good;
- The variance advantage becomes less visible as the number of included lags goes up.

NB: The latest version of the draft (April 2023) does not contain the simulations part.

Notation

Variable	Meaning
τ_{it}	Average Treatment Effect for unit i at period t
Ω	The set of observed *it* pairs
Ω_1	The set of treated *it* pairs
Ω_0	The set of non-treated *it* pairs
w_1	Pre-chosen weights for ATTs
τ_w	The average ATT weighted with w_1
θ	Free parameters
Γ	Matrix that maps free parameters to ATTs
B	Matrix that maps ATTs to free parameters
δ	Common-across-units effects
λ_i	Unit-specific effects

Table of contents