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

Sasha Petrov

May 3, 2022

- In the vein of Callaway & Sant'Anna, the idea is to move away from regressions, which estimate who knows what

Setting

Instead of defining the treatment alternatives set as time series, they stay binary:

$$(Y(1), Y(0), D, T, E) \sim P \tag{1}$$

Across unit heterogeneity is in P; across time heterogeneity is in conditioning with respect to T; for lag effects pick appropriate E.
Definition of a treatment effect:

$$\begin{aligned}\tau_t &= \mathbb{E}\left[Y_t - Y_t(\mathbf{0})\right] \\ &= \mathbb{E}\left[Y(1) - \mathbb{E}\left[Y(0) | E = \infty, T = t, D = 1\right] | T = t, D = 1\right]\end{aligned}$$

To add the i index, recognise non-random sampling:

$$(Y(1), Y(0), D, T, E)_i \sim P|_{A_i} \tag{2}$$

The partition (?) of the sample space into units:

$$\{A_i\}_{i \in \mathcal{S}}, \cup A_i \subset \Omega \tag{3}$$

$$\begin{aligned}
\tau_{it} &= \mathbb{E}\left[Y_{it} - Y_{it}(\mathbf{0})\right] \\
&= \mathbb{E}_{P|_{A_i}}\left[Y(1) - \mathbb{E}\left[Y(0)|E = \infty, T = t, D = 1\right] | T = t, D = 1\right]
\end{aligned}$$

Definitions and assumptions

- $\tau = \{\tau_{it}\}_{it \in \Omega_1}$
- $\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it} \equiv w_1' \tau$
  - $ATT(g,t) = \sum_{it \in \{it \in \Omega_1 : E_{it} = g\}} \frac{1}{|\{it \in \Omega_1 : E_{it} = g\}|} \tau_{it} = \sum_{it \in \Omega_1} \frac{\mathbf{1}[E_{it} = g]}{|\{it \in \Omega_1 : E_{it} = g\}|} \tau_{it}$
- Parallel trends: $\mathbb{E}[Y_{it}(0)] = \alpha_i + \beta_t, \quad \forall it \in \Omega$
  - Is this a necessary condition for parallel trends?
- No anticipation: $Y_{it} = Y_{it}(0) \quad \forall it \in \Omega_0$
- (Optional) Restricted causal effects: $B\tau = 0$
  - "Mirror image" that focuses on the number of free parameters rather than restrictions: $\tau = \Gamma\theta$
    - No hetergoeneity: $\tau = (1, \ldots, 1) \times \theta_1$, $\theta_1 \in \mathbb{R}$

Why TWFE sucks

- A somewhat weird logic to me (my issue, not theirs): we don't know what TWFE estimate $\rightarrow$ let's conjecture assumptions that would make it easy to establish what TWFE estimate $\rightarrow$ realise that these assumptions make the estimator biased
  - But there's still a possibility that TWFE estimate some other parameter without bias, no?

static OLS

Table 1: Two-Unit, Three-Period Example

$\mathbb{E}[Y_{it}]$	$i = A$	$i = B$
$t = 1$	α_A	α_B
$t = 2$	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$
$t = 3$	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$
Event date	$E_i = 2$	$E_i = 3$

$$\tau^{\text{static}} = \tau_{A2} + \frac{1}{2}\tau_{B3} - \frac{1}{2}\tau_{A3} \tag{4}$$

Imputation-based estimation

- General model of $Y(0)$: $\mathbb{E}\left[Y_{it}(0)\right] = A'_{it}\lambda_i + X'_{it}\delta$
  - A_{it} - non-time varying but allows across-unit heterogeneity of nuisance parameters; X_{it} - time-varying but homogeneous nuisance parameters
- The regression to estimate: $Y_{it} = A'_{it}\lambda_i + X'_{it}\delta + D_{it}(\Gamma'\theta)_{it} + \varepsilon_{it}$
- Need to worry about the identification of θ rather than τ (which should be easier?)

Theorem 1 (Efficient estimator). *Suppose Assumptions 1′, 2, 3′ and 4 hold. Then among the linear unbiased estimators of* τ_w*, the (unique) efficient estimator* $\hat{\tau}_w^*$ *can be obtained with the following steps:*

1. *Estimate* θ *by the OLS solution* $\hat{\theta}^*$ *from the linear regression (4) (where we assume that* θ *is identified);*
2. *Estimate the vector of treatment effects* τ *by* $\hat{\tau}^* = \Gamma\hat{\theta}^*$*;*
3. *Estimate the target* τ_w *by* $\hat{\tau}_w^* = w_1'\hat{\tau}^*$*.*

Moreover, this estimator $\hat{\tau}_w^*$ *is unbiased for* τ_w *under Assumptions 1′, 2 and 3′ alone, even when error terms are not homoskedastic.*

Theorem 2 (Imputation representation for the efficient estimator). *With a null Assumption 3′ (that is, if $\Gamma = \mathbb{I}_{N_1}$), the unique efficient linear unbiased estimator $\hat{\tau}_w^*$ of τ_w from Theorem 1 can be obtained via an imputation procedure:*

1. *Within the untreated observations only ($it \in \Omega_0$), estimate the λ_i and δ (by $\hat{\lambda}_i^*, \hat{\delta}^*$) by OLS in the regression*

$$Y_{it} = A_{it}'\lambda_i + X_{it}'\delta + \varepsilon_{it}; \tag{5}$$

2. *For each treated observation ($it \in \Omega_1$) with $w_{it} \neq 0$, set $\hat{Y}_{it}(0) = A_{it}'\hat{\lambda}_i^* + X_{it}'\hat{\delta}^*$ and $\hat{\tau}_{it}^* = Y_{it} - \hat{Y}_{it}(0)$ to obtain the estimate of τ_{it};*

3. *Estimate the target τ_w by a weighted sum $\hat{\tau}_w^* = \sum_{it \in \Omega_1} w_{it}\hat{\tau}_{it}^*$.*

Proposition 7 (Consistency of $\hat{\tau}_w$). *Under Assumptions 1′, 2, 3′, 5 and 6,* $\hat{\tau}_w - \tau_w \xrightarrow{\mathcal{L}_2} 0$ *for an unbiased estimator* $\hat{\tau}_w$ *of* τ_w*, such as* $\hat{\tau}^*_w$ *in Theorem 1.*[30]

We next consider the asymptotic distribution of the estimator around the estimand.

Proposition 8 (Asymptotic Normality). *Under the assumptions of Proposition 7, a balance assumption on higher moments of the weights (Assumption A1), and if* $\liminf n_H \sigma^2_w > 0$ *for* $\sigma^2_w = \mathrm{Var}\left[\hat{\tau}_w\right]$*, we have that*

$$\sigma_w^{-1}(\hat{\tau}_w - \tau_w) \xrightarrow{d} \mathcal{N}(0,1).$$

Table 2: Efficiency and Bias of Alternative Estimators

Horizon	Estimator	Baseline simulation		More pre-periods	Heterosk. errors	AR(1) errors	Anticipation effects
		Variance (1)	Coverage (2)	Variance (3)	Variance (4)	Variance (5)	Bias (6)
$h = 0$	Imputation	0.0099	0.942	0.0080	0.0347	0.0072	-0.0569
	dCDH	0.0140	0.944	0.0140	0.0526	0.0070	-0.0915
	SA	0.0115	0.946	0.0115	0.0404	0.0066	-0.0753
	OLS semi	0.0097	0.946	0.0078	0.0345	0.0072	-0.0550
	OLS fully	0.0115	0.940	0.0103	0.0410	0.0067	-0.0770
$h = 1$	Imputation	0.0145	0.952	0.0111	0.0532	0.0143	-0.0719
	dCDH	0.0185	0.946	0.0185	0.0703	0.0151	-0.0972
	SA	0.0177	0.956	0.0177	0.0643	0.0165	-0.0812
	OLS semi	0.0143	0.958	0.0105	0.0518	0.0144	-0.0700
	OLS fully	0.0181	0.946	0.0139	0.0607	0.0150	-0.0918
$h = 2$	Imputation	0.0222	0.942	0.0161	0.0813	0.0240	-0.0886
	dCDH	0.0262	0.954	0.0262	0.0952	0.0257	-0.1020
	SA	0.0317	0.970	0.0317	0.1108	0.0341	-0.0850
	OLS semi	0.0219	0.950	0.0151	0.0811	0.0241	-0.0866
	OLS fully	0.0293	0.946	0.0201	0.0936	0.0275	-0.1062
$h = 3$	Imputation	0.0366	0.952	0.0255	0.1379	0.0394	-0.1101
	dCDH	0.0422	0.948	0.0422	0.1488	0.0446	-0.1087
	SA	0.0479	0.964	0.0479	0.1659	0.0543	-0.0932
	OLS semi	0.0345	0.952	0.0232	0.1346	0.0388	-0.1048
	OLS fully	0.0467	0.948	0.0302	0.1508	0.0446	-0.1206
$h = 4$	Imputation	0.0800	0.950	0.0546	0.3197	0.0773	-0.1487
	dCDH	0.0932	0.952	0.0932	0.3263	0.0903	-0.1265
	SA	0.0932	0.946	0.0932	0.3263	0.0903	-0.1265
	OLS semi	0.0601	0.938	0.0418	0.2598	0.0669	-0.1263
	OLS fully	0.0777	0.936	0.0510	0.2773	0.0797	-0.1351

Notes: See Section 4.6 for a detailed description of the data-generating processes and reported statistics.

Thoughts

- Pre-testing – do they do it appropriately? What are the implications for last-stage inference?
- Imputation looks like suitable material for all of the optimal weighting of predictors stuff, right? But probably will need bootstrap to update losses?
- As we realised last time, we need to impose some sort of stationarity to make these target parameters interesting in general – then, why not be explicit about it and use it to extrapolate within sample as well to expand the support of estimated effects?