Revisiting Event Study Designs: Robust and Efficient Estimation

Kirill Borusyak[1,4] Xavier Jaravel[2,4] Jann Spiess[3]

[1]UCL [2]LSE [3]Stanford [4]CEPR

DiD Reading Group, June 2021

Event study designs / diff-in-diff with staggered adoption

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$				
$t = 4$				
$t = 5$				
$t = 6$				

Outcomes Y_{it}

Treatment indicator: $D_{it} = \begin{cases} 1, & \text{treated} \\ 0, & \text{not treated} \end{cases}$

From diff-in-diff to two-way fixed effects

	$i = A$	$i = B$
$t = 1$		
$t = 2$		

Desired estimator:

$$\hat{\tau} = (Y_{A2} - Y_{A1}) - (Y_{B2} - Y_{B1})$$

is obtained from two-way FE regression:

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau + \varepsilon_{it}$$

Event study design and two-way fixed effect regression

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$				
$t = 4$				
$t = 5$				
$t = 6$				
	$E_i = 2$	$E_i = 3$	$E_i = 5$	$E_i = \infty$

Dynamic specifications:

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + \sum_{\substack{h=-a \\ h \neq -1}}^{b-1} \tau_h 1\left[t = E_i + h\right] + \tau_{b+} 1\left[t \geq E_i + b\right] + \varepsilon_{it}$$

Static specification:

$$Y_{it} = \tilde{\alpha}_i + \tilde{\beta}_t + D_{it}\tau + \varepsilon_{it}$$

Implicit assumptions

$$Y_{it} = \overbrace{\alpha_i + \beta_t}^{\text{parallel trends}} + \underbrace{D_{it}}_{\text{no anticipation}} \overbrace{\tau}^{\substack{\text{treatment effect}\\ \text{homogeneity}}} + \varepsilon_{it}$$

This paper

1. Make assumptions explicit and separate them from goals
2. Relate problems with common practice to mismatch between assumptions and convenience regressions
   (building upon original Borusyak and Jaravel (2017) version)
3. Derive robust and efficient estimator from first principles, work out "imputation" structure (in baseline case)
4. Provide large-sample theory and inference for this estimator
5. Propose approach to testing, separate from estimation
6. Compare with alternative robust estimation strategies
7. Discuss extensions beyond event studies

Related event-studies literature

- Problems with simple OLS regressions in event study designs
  Borusyak and Jaravel (2017); Goodman-Bacon (2018); Strezhnev (2018); De Chaisemartin and D'Haultfœuille (2020); Sun and Abraham (2021); Schmidheiny and Siegloch (2020); Baker et al. (2021)

- Estimators that are robust to heterogeneity
  De Chaisemartin and D'Haultfœuille (2020); Sun and Abraham (2021); Callaway and Sant'Anna (2021); Cengiz et al. (2019)

- Justification and testing of, robustness wrt parallel trends
  Roth (2018); Rambachan and Roth (2020); Roth and Sant'Anna (2020)

- Randomization-based and doubly-robust estimation
  Athey and Imbens (2018); Arkhangelsky and Imbens (2019); Callaway and Sant'Anna (2021); Roth and Sant'Anna (2021)

- Imputation-based estimation in panel data Gobillon and Magnac (2016); Xu (2017); Liu et al. (2020); Gardner (2021)

Outline

1. Setting and Framework
2. Conventional Practice and Associated Problems
3. Imputation-Based Estimation and Testing
4. Asymptotic Theory and Inference
5. Testing
6. Comparison to Other Estimators
7. Extensions

Outline

1. Setting and Framework
2. Conventional Practice and Associated Problems
3. Imputation-Based Estimation and Testing
4. Asymptotic Theory and Inference
5. Testing
6. Comparison to Other Estimators
7. Extensions

Setup: fixed sample

	$i = A$	$i = B$	$i = C$	$i = D$
$t = 1$				
$t = 2$				
$t = 3$				
$t = 4$				
$t = 5$				
$t = 6$				
	$E_i = 2$	$E_i = 3$	$E_i = 5$	$E_i = \infty$

$D_{it} = 1[t \geq E_i]$ — treatment indicator
$\Omega_1 = \{it; D_{it} = 1\}$ — set of treated observations
$\Omega_0 = \{it; D_{it} = 0\}$ — set of untreated observations
(both not-yet-treated and, if any, never-treated)

Framework (I)

Estimation Target

$$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it} \equiv {w_1}' \tau \text{ (where } \tau_{it} = Y_{it} - \underbrace{Y_{it}(0)}_{\text{outcome if never treated}})$$

Framework (I)

Estimation Target

$$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it} \equiv {w_1}' \tau \text{ (where } \tau_{it} = Y_{it} - \underbrace{Y_{it}(0)}_{\text{outcome if never treated}})$$

- Sample ATT
- Horizon-specific ATT ($E_i = t + h$)

Framework (I)

Estimation Target

$$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it} \equiv {w_1}' \tau \text{ (where } \tau_{it} = Y_{it} - \underbrace{Y_{it}(0)}_{\text{outcome if never treated}})$$

- Sample ATT
- Horizon-specific ATT ($E_i = t + h$)
- Differences btw subgroup ATTs
- Size-weighted ATT
- Balance unit composition across horizons

Framework (I)

Estimation Target

$$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it} \equiv {w_1}' \tau \text{ (where } \tau_{it} = Y_{it} - \underbrace{Y_{it}(0)}_{\text{outcome if never treated}})$$

- Sample ATT
- Horizon-specific ATT ($E_i = t + h$)
- Differences btw subgroup ATTs
- Size-weighted ATT
- Balance unit composition across horizons

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$ with $\mathbb{E}[\varepsilon_{it}] = 0$, non-stochastic α_i, β_t

Framework (I)

Estimation Target

$$\tau_w = \sum_{it \in \Omega_1} w_{it}\tau_{it} \equiv {w_1}'\tau \text{ (where } \tau_{it} = Y_{it} - \underbrace{Y_{it}(0)}_{\text{outcome if never treated}})$$

- Sample ATT
- Horizon-specific ATT ($E_i = t + h$)
- Differences btw subgroup ATTs
- Size-weighted ATT
- Balance unit composition across horizons

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$ with $\mathbb{E}[\varepsilon_{it}] = 0$, non-stochastic α_i, β_t

A1′: Generalized Fixed-Effects Model

$$Y_{it}(0) = A'_{it}\lambda_i + X'_{it}\delta + \varepsilon_{it} = Z'_{it}\pi + \varepsilon_{it}$$

Framework (II)

Estimation Target

$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

Framework (II)

Estimation Target

$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

A3: Treatment-Effect Model (optional!)

Restrictions $B\tau = 0$ hold for a known matrix B

Framework (II)

Estimation Target

$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

A3: Treatment-Effect Model (optional!)

Restrictions $B\tau = 0$ hold for a known matrix B

$\Leftrightarrow$ Model $\tau = \Gamma\theta$ holds for unknown parameters θ, known matrix Γ

Framework (II)

Estimation Target

$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

A3: Treatment-Effect Model (optional!)

Restrictions $B\tau = 0$ hold for a known matrix B
$\Leftrightarrow$ Model $\tau = \Gamma\theta$ holds for unknown parameters θ, known matrix Γ

"True" model for observed outcomes:

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau_{it} + \varepsilon_{it}, \qquad \tau = \Gamma\theta$$

Outline

1. Setting and Framework
2. Conventional Practice and Associated Problems
3. Imputation-Based Estimation and Testing
4. Asymptotic Theory and Inference
5. Testing
6. Comparison to Other Estimators
7. Extensions

Under-identification of the fully-dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{\substack{h=-\infty \\ h \neq -1}}^{\infty} \tau_h 1\left[t = E_i + h\right] + \varepsilon_{it}$$

Under-Identification

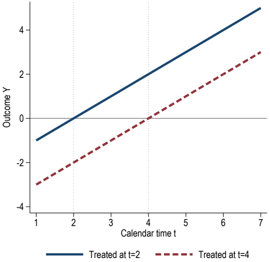
Without never-treated units, path $\{\tau_h\}_{h \neq -1}$ not point identified in fully-dynamic OLS: adding a linear trend to this path, $\{\tau_h + \kappa\,(h+1)\}$ fits the data equally well

Under-identification of the fully-dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{\substack{h=-\infty \\ h\neq -1}}^{\infty} \tau_h 1\left[t = E_i + h\right] + \varepsilon_{it}$$

Under-Identification

Without never-treated units, path $\{\tau_h\}_{h\neq -1}$ not point identified in fully-dynamic OLS: adding a linear trend to this path, $\{\tau_h + \kappa\,(h+1)\}$ fits the data equally well



Under-identification of the fully-dynamic specification

$$Y_{it} = \alpha_i + \beta_t + \sum_{\substack{h=-\infty \\ h\neq -1}}^{\infty} \tau_h 1\left[t = E_i + h\right] + \varepsilon_{it}$$

Under-Identification

Without never-treated units, path $\{\tau_h\}_{h\neq -1}$ not point identified in fully-dynamic OLS: adding a linear trend to this path, $\{\tau_h + \kappa\,(h+1)\}$ fits the data equally well

Estimation Target

$\tau_w = \sum_{it\in\Omega_1} w_{it}\tau_{it}$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

A3: Treatment-Effect Model

$\tau = \Gamma\theta$

Negative weighting in the static regression

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau + \varepsilon_{it}, \qquad D_{it} = 1[t \geq E_i]$$

OLS estimand

$\tau = \sum_{it \in \Omega_1} w_{it}^{\text{OLS}} \tau_{it}$ for some weights w_{it}^{OLS} with $\sum_{it \in \Omega_1} w_{it}^{\text{OLS}} = 1$

Negative weighting in the static regression

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau + \varepsilon_{it}, \qquad D_{it} = 1[t \geq E_i]$$

OLS estimand

$\tau = \sum_{it \in \Omega_1} w_{it}^{\text{OLS}} \tau_{it}$ for some weights w_{it}^{OLS} with $\sum_{it \in \Omega_1} w_{it}^{\text{OLS}} = 1$

Negative weights

Under A1, A2, OLS can put negative weight on treatment effects.
Here, $\tau = \tau_{A2} + \frac{1}{2}\tau_{B3} - \frac{1}{2}\tau_{A3}$

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

Why negative weights?

Negative weights

Under A1, A2, OLS can put negative weight on treatment effects

Estimation Target

$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A3: Treatment-Effect Model

$\tau = \Gamma \theta$

$$Y_{it} = \alpha_i + \beta_t + D_{it} \tau + \varepsilon_{it} \quad \longleftrightarrow \quad Y_{it} = \alpha_i + \beta_t + D_{it} \tau_{it} + \varepsilon_{it}$$

Why negative weights?

Negative weights

Under A1, A2, OLS can put negative weight on treatment effects

Estimation Target

$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A3: Treatment-Effect Model

$\tau = \Gamma \theta$

$$Y_{it} = \alpha_i + \beta_t + D_{it} \tau + \varepsilon_{it} \quad \longleftrightarrow \quad Y_{it} = \alpha_i + \beta_t + D_{it} \tau_{it} + \varepsilon_{it}$$

- Forbidden comparisons
- Efficient variance weighting

Spurious identification of long-run causal effects

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

Spurious Identification of Long-Run Effects

Without never-treated units A1, A2 do not identify long-run effects, while dynamic OLS specifications produce some estimates

Spurious identification of long-run causal effects

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

Spurious Identification of Long-Run Effects

Without never-treated units A1, A2 do not identify long-run effects, while dynamic OLS specifications produce some estimates

Estimation Target

$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

A3: Treatment-Effect Model

$\tau = \Gamma\theta$

Outline

1. Setting and Framework
2. Conventional Practice and Associated Problems
3. Imputation-Based Estimation and Testing
4. Asymptotic Theory and Inference
5. Testing
6. Comparison to Other Estimators
7. Extensions

Efficient estimation under homoskedasticity

Estimation Target

$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A3: Treatment-Effect Model

$\tau = \Gamma \theta$

A4: Homoscedasticity

$\mathbb{E}[\varepsilon_{it} \varepsilon_{js}] = \sigma^2 1[it = js]$

Analysis conditional on panel, only randomness is ε_{it}

Efficient Estimator

Among linear unbiased estimators of τ_w, the (unique) efficient estimator $\hat{\tau}_w^*$ can be written as:

1. Estimate θ by $\hat{\theta}$ from the linear regression (where we assume that θ is identified)

$$Y_{it} = \alpha_i + \beta_t + D_{it}\Gamma_{it}'\theta + \varepsilon_{it}$$

2. Estimate the vector of treatment effects τ by $\hat{\tau} = \Gamma\hat{\theta}$
3. Estimate the target τ_w by $\hat{\tau}_w^* = {w_1}'\hat{\tau}$

Imputation structure of the efficient estimator

With no restrictions on treatment effects ($\Gamma = \mathbb{I}$), this $\hat{\tau}_w^*$ can be obtained as an imputation estimator:

1. *Estimate:* Within $it \in \Omega_0$, estimate α_i, β_t from $Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$
2. *Extrapolate:* $\hat{Y}_{it}(0) = \hat{\alpha}_i + \hat{\beta}_t$ and $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$ for $it \in \Omega_1$
3. *Take averages:* $\hat{\tau}_w^* = \sum_{it \in \Omega_1} w_{it} \hat{\tau}_{it}$

Imputation structure of the efficient estimator

With no restrictions on treatment effects ($\Gamma = \mathbb{I}$), this $\hat{\tau}_w^*$ can be obtained as an imputation estimator:

1. *Estimate:* Within $it \in \Omega_0$, estimate α_i, β_t from $Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$
2. *Extrapolate:* $\hat{Y}_{it}(0) = \hat{\alpha}_i + \hat{\beta}_t$ and $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$ for $it \in \Omega_1$
3. *Take averages:* $\hat{\tau}_w^* = \sum_{it \in \Omega_1} w_{it} \hat{\tau}_{it}$

Remarks:

- Efficiency is finite-sample

Imputation structure of the efficient estimator

With no restrictions on treatment effects ($\Gamma = \mathbb{I}$), this $\hat{\tau}_w^*$ can be obtained as an imputation estimator:

1. *Estimate:* Within $it \in \Omega_0$, estimate α_i, β_t from $Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$
2. *Extrapolate:* $\hat{Y}_{it}(0) = \hat{\alpha}_i + \hat{\beta}_t$ and $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$ for $it \in \Omega_1$
3. *Take averages:* $\hat{\tau}_w^* = \sum_{it \in \Omega_1} w_{it} \hat{\tau}_{it}$

Remarks:

- Efficiency is finite-sample
- For specific w_{it} yields Liu et al. (2020); Gardner (2021)

Imputation structure of the efficient estimator

With no restrictions on treatment effects ($\Gamma = \mathbb{I}$), this $\hat{\tau}_w^*$ can be obtained as an imputation estimator:

1. *Estimate:* Within $it \in \Omega_0$, estimate α_i, β_t from $Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$
2. *Extrapolate:* $\hat{Y}_{it}(0) = \hat{\alpha}_i + \hat{\beta}_t$ and $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$ for $it \in \Omega_1$
3. *Take averages:* $\hat{\tau}_w^* = \sum_{it \in \Omega_1} w_{it} \hat{\tau}_{it}$

Remarks:

- Efficiency is finite-sample
- For specific w_{it} yields Liu et al. (2020); Gardner (2021)
- Imputation structure generalizes to non-trivial Γ (extra step: adjusted the estimand)

Imputation structure of the efficient estimator

With no restrictions on treatment effects ($\Gamma = \mathbb{I}$), this $\hat{\tau}^*_w$ can be obtained as an imputation estimator:

1. *Estimate:* Within $it \in \Omega_0$, estimate α_i, β_t from $Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$
2. *Extrapolate:* $\hat{Y}_{it}(0) = \hat{\alpha}_i + \hat{\beta}_t$ and $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$ for $it \in \Omega_1$
3. *Take averages:* $\hat{\tau}^*_w = \sum_{it \in \Omega_1} w_{it} \hat{\tau}_{it}$

Remarks:

- Efficiency is finite-sample
- For specific w_{it} yields Liu et al. (2020); Gardner (2021)
- Imputation structure generalizes to non-trivial Γ (extra step: adjusted the estimand)
- Also generalizes to *any* unbiased estimator of τ_w, with the estimation step possibly done inefficiently

Outline

1. Setting and Framework
2. Conventional Practice and Associated Problems
3. Imputation-Based Estimation and Testing
4. Asymptotic Theory and Inference
5. Testing
6. Comparison to Other Estimators
7. Extensions

Assumptions for asymptotic theory

Estimation Target

$\tau_w = \sum_{it \in \Omega_1} w_{it} \tau_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A3: Treatment-Effect Model

$\tau = \Gamma \theta$

A5: Clustered Errors

Error terms clustered by units i

Consider some $\hat{\tau}_w = \sum_{it} v_{it} Y_{it}$, with $\hat{\tau}_w^*$ as special case

Consistency and asymptotic Normality

Under a Herfindahl condition on the weights, $\hat{\tau}_w$ is consistent.
Under additional moment condition on the weights, also

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

for $\sigma_w^2 = \mathrm{Var}(\hat{\tau}_w)$

- Sufficient conditions for complete, short panels

Variance estimation with unrestricted treatment effects

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau_{it} + \varepsilon_{it}, \qquad \sigma_w^2 = \mathbb{E}\left[\sum_i \left(\sum_t v_{it}\varepsilon_{it}\right)^2\right]$$

Variance estimation with unrestricted treatment effects

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau_{it} + \varepsilon_{it}, \qquad \sigma_w^2 = \mathbb{E}\left[\sum_i \left(\sum_t v_{it}\varepsilon_{it}\right)^2\right]$$

Two challenges with plug-in estimation

- In short panels, $\hat{\alpha}_i$ and thus $\hat{\varepsilon}_{it}$ are not consistent

Variance estimation with unrestricted treatment effects

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau_{it} + \varepsilon_{it}, \qquad \sigma_w^2 = \mathbb{E}\left[\sum_i \left(\sum_t v_{it}\varepsilon_{it}\right)^2\right]$$

Two challenges with plug-in estimation

- In short panels, $\hat{\alpha}_i$ and thus $\hat{\varepsilon}_{it}$ are not consistent
- The imputation procedure fits $it \in \Omega_1$ perfectly, $\hat{\varepsilon}_{it} = 0$

Variance estimation with unrestricted treatment effects

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau_{it} + \varepsilon_{it}, \qquad \sigma_w^2 = \mathbb{E}\left[\sum_i \left(\sum_t v_{it}\varepsilon_{it}\right)^2\right]$$

Two challenges with plug-in estimation

- In short panels, $\hat{\alpha}_i$ and thus $\hat{\varepsilon}_{it}$ are not consistent
  — not a problem with clustered SE (Stock and Watson 2008)
- The imputation procedure fits $it \in \Omega_1$ perfectly, $\hat{\varepsilon}_{it} = 0$

Variance estimation with unrestricted treatment effects

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau_{it} + \varepsilon_{it}, \qquad \sigma_w^2 = \mathbb{E}\left[\sum_i \left(\sum_t v_{it}\varepsilon_{it}\right)^2\right]$$

Two challenges with plug-in estimation

- In short panels, $\hat{\alpha}_i$ and thus $\hat{\varepsilon}_{it}$ are not consistent
  — not a problem with clustered SE (Stock and Watson 2008)
- The imputation procedure fits $it \in \Omega_1$ perfectly, $\hat{\varepsilon}_{it} = 0$
  — solve by using $\tilde{\varepsilon}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_t - \hat{\bar{\tau}}_{it}$ instead,
  where $\hat{\bar{\tau}}_{it} \xrightarrow{p} \bar{\tau}_{it}$ is some average of many treatment effects

Variance estimation with unrestricted treatment effects

$$Y_{it} = \alpha_i + \beta_t + D_{it}\tau_{it} + \varepsilon_{it}, \qquad \sigma_w^2 = \mathbb{E}\left[\sum_i \left(\sum_t v_{it}\varepsilon_{it}\right)^2\right]$$

Two challenges with plug-in estimation

- In short panels, $\hat{\alpha}_i$ and thus $\hat{\varepsilon}_{it}$ are not consistent
— not a problem with clustered SE (Stock and Watson 2008)
- The imputation procedure fits $it \in \Omega_1$ perfectly, $\hat{\varepsilon}_{it} = 0$
— solve by using $\tilde{\varepsilon}_{it} = Y_{it} - \hat{\alpha}_i - \hat{\beta}_t - \hat{\bar{\tau}}_{it}$ instead,
where $\hat{\bar{\tau}}_{it} \xrightarrow{p} \bar{\tau}_{it}$ is some average of many treatment effects

Conservative Inference

Under regularity conditions on v_{it}, the variance estimator is asympotically conservatively valid (and asy. exact if $\tau_{it} \equiv \bar{\tau}_{it}$)

Outline

1. Setting and Framework
2. Conventional Practice and Associated Problems
3. Imputation-Based Estimation and Testing
4. Asymptotic Theory and Inference
5. Testing
6. Comparison to Other Estimators
7. Extensions

Testing challenges and our approach

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

Two existing approaches to testing A1 and A2:

Testing challenges and our approach

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

Two existing approaches to testing A1 and A2:

- Coefficients on leads in OLS

Testing challenges and our approach

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

Two existing approaches to testing A1 and A2:

- Coefficients on leads in OLS
- Placebo tests based on robust estimators

Testing challenges and our approach

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

Two existing approaches to testing A1 and A2:

- Coefficients on leads in OLS
  — contaminated by treatment effect heterogeneity (Sun and Abraham, 2021). We use pre-treatment observations only
- Placebo tests based on robust estimators

Testing challenges and our approach

A1: Parallel Trends

$$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$$

A2: No Anticipation

$$Y_{it} = Y_{it}(0) \text{ for all } it \in \Omega_0$$

Two existing approaches to testing A1 and A2:

- Coefficients on leads in OLS
  — contaminated by treatment effect heterogeneity (Sun and Abraham, 2021). We use pre-treatment observations only
- Placebo tests based on robust estimators
  — does not distinguish estimation and testing.

Testing challenges and our approach

A1: Parallel Trends

$Y_{it}(0) = \alpha_i + \beta_t + \varepsilon_{it}$

A2: No Anticipation

$Y_{it} = Y_{it}(0)$ for all $it \in \Omega_0$

Two existing approaches to testing A1 and A2:

- Coefficients on leads in OLS
  — contaminated by treatment effect heterogeneity (Sun and Abraham, 2021). We use pre-treatment observations only
- Placebo tests based on robust estimators
  — does not distinguish estimation and testing.

We propose to estimate by OLS:

$$y_{it} = \alpha_i + \beta_t + {W_{it}}'\gamma + \tilde{\varepsilon}_{it} \qquad \text{on } it \in \Omega_0 \text{ only}$$

where W_{it} are e.g. indicators of pre-periods $-1, \ldots, -K$

Robustness to pre-testing

OLS-based pre-trend coefficients are correlated with treatment effect estimates, biasing inference (Roth, 2018)

OLS-based pre-trend coefficients are correlated with treatment effect estimates, biasing inference (Roth, 2018)

Robustness to pre-testing

Tests based on outcome Y_{it} for $it \in \Omega_0$ alone are orthogonal to $\hat{\tau}_w^*$ under homoskedasticity. Thus, conditioning on the pre-trend passing does not affect asymptotic inference.

Outline

1. Setting and Framework
2. Conventional Practice and Associated Problems
3. Imputation-Based Estimation and Testing
4. Asymptotic Theory and Inference
5. Testing
6. Comparison to Other Estimators
7. Extensions

CATT-based approaches (1)

Several papers proposed estimators robust to heterogeneous effects (De Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021):

- Estimate cohort-average treatment effects $CATT_{et}$
- Average them in some way

CATT-based approaches (1)

Several papers proposed estimators robust to heterogeneous effects (De Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021):

- Estimate cohort-average treatment effects $CATT_{et}$
- Average them in some way
- Can be represented in an imputation way with some $\hat{Y}_{it}(0)$, but not efficient under homoskedasticity

CATT-based approaches (1)

Several papers proposed estimators robust to heterogeneous effects (De Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021):

- Estimate cohort-average treatment effects $CATT_{et}$
- Average them in some way
- Can be represented in an imputation way with some $\hat{Y}_{it}(0)$, but not efficient under homoskedasticity
- Main difference: they only use period $e-1$ for $CATT_{et}$, even if parallel trends are imposed on all not-yet-treated units

CATT-based approaches (1)

Several papers proposed estimators robust to heterogeneous effects (De Chaisemartin and D'Haultfœuille, 2020; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021):

- Estimate cohort-average treatment effects $CATT_{et}$
- Average them in some way
- Can be represented in an imputation way with some $\hat{Y}_{it}(0)$, but not efficient under homoskedasticity
- Main difference: they only use period $e-1$ for $CATT_{et}$, even if parallel trends are imposed on all not-yet-treated units

$\mathbb{E}[Y_{it}]$	$i = A$	$i = B$	$i = C$
$t = 1$	α_A	α_B	α_C
$t = 2$	$\alpha_A + \beta_2 + \tau_{A2}$	$\alpha_B + \beta_2$	$\alpha_C + \beta_2$
$t = 3$	$\alpha_A + \beta_3 + \tau_{A3}$	$\alpha_B + \beta_3 + \tau_{B3}$	$\alpha_C + \beta_3$
	$E_i = 2$	$E_i = 3$	$E_i = 4$

CATT-based approaches (2)

Additional advantages of the imputation approach:

- Transparently maps $Y_{it}(0)$ model to estimator
  $\Rightarrow$ immediately extends beyond event studies

CATT-based approaches (2)

Additional advantages of the imputation approach:

- Transparently maps $Y_{it}(0)$ model to estimator
  $\Rightarrow$ immediately extends beyond event studies
- Separates testing from estimation
  $\Rightarrow$ mitigates the pre-testing bias of Roth (2018)

CATT-based approaches (2)

Additional advantages of the imputation approach:

- Transparently maps $Y_{it}(0)$ model to estimator
  $\Rightarrow$ immediately extends beyond event studies
- Separates testing from estimation
  $\Rightarrow$ mitigates the pre-testing bias of Roth (2018)
- Does not require random sampling
  $\Rightarrow$ allows for richer class of estimands

CATT-based approaches (2)

Additional advantages of the imputation approach:

- Transparently maps $Y_{it}(0)$ model to estimator
  $\Rightarrow$ immediately extends beyond event studies
- Separates testing from estimation
  $\Rightarrow$ mitigates the pre-testing bias of Roth (2018)
- Does not require random sampling
  $\Rightarrow$ allows for richer class of estimands
- Analtyical and computationally efficient SE

Simulations

Baseline DGP

- $I = 250$ units, $T = 6$ periods
- $E_i = 2, \ldots, 7$ with equal probabilities
- Growing treatment effects: $\tau_{it} = 1 + (t - E_i)$
- Residuals $\varepsilon_{it} \sim$ iid $\mathcal{N}(0, 1)$

Simulations

Baseline DGP

- $I = 250$ units, $T = 6$ periods
- $E_i = 2, \ldots, 7$ with equal probabilities
- Growing treatment effects: $\tau_{it} = 1 + (t - E_i)$
- Residuals $\varepsilon_{it} \sim$ iid $\mathcal{N}(0, 1)$

Deviations

- Four extra pre-periods: $t = -3, \ldots, 0$
- Heteroskedasticity: $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_{it}^2)$, $\sigma_{it}^2 = t$
- Autocorrelation: AR(1) process with $\rho = 0.5$
- Anticipation effects: $1/\sqrt{I} = 0.0632$ at $t = E_i - 1$

Simulation results: 15-41% lower variance in baseline case

Horizon	Estimator	Baseline simulation *Var* (1)	Baseline simulation *Covrg* (2)	More pre-periods *Var* (3)	Heterosk. errors *Var* (4)	AR(1) errors *Var* (5)	Anticip. effects *Bias* (6)
$h = 0$	Imputation	0.0099	0.942	0.0080	0.0347	0.0072	-0.0569
	DCDH	0.0140	0.938	0.0140	0.0526	0.0070	-0.0915
	SA	0.0115	0.938	0.0115	0.0404	0.0066	-0.0753
$h = 1$	Imputation	0.0145	0.936	0.0111	0.0532	0.0143	-0.0719
	DCDH	0.0185	0.948	0.0185	0.0703	0.0151	-0.0972
	SA	0.0177	0.948	0.0177	0.0643	0.0165	-0.0812
$h = 2$	Imputation	0.0222	0.956	0.0161	0.0813	0.0240	-0.0886
	DCDH	0.0262	0.958	0.0262	0.0952	0.0257	-0.1020
	SA	0.0317	0.950	0.0317	0.1108	0.0341	-0.0850
$h = 3$	Imputation	0.0366	0.928	0.0255	0.1379	0.0394	-0.1101
	DCDH	0.0422	0.930	0.0422	0.1488	0.0446	-0.1087
	SA	0.0479	0.952	0.0479	0.1659	0.0543	-0.0932
$h = 4$	Imputation	0.0800	0.942	0.0546	0.3197	0.0773	-0.1487
	DCDH	0.0932	0.950	0.0932	0.3263	0.0903	-0.1265
	SA	0.0932	0.954	0.0932	0.3263	0.0903	-0.1265

Comparison to other imputation results

Similar imputation-based approaches:

- Gobillon and Magnac (2016); Xu (2017) for factor models
- Liu et al. (2020); Gardner (2021) applied to event studies to estimate the overall or horizon-specific ATT

Comparison to other imputation results

Similar imputation-based approaches:

- Gobillon and Magnac (2016); Xu (2017) for factor models
- Liu et al. (2020); Gardner (2021) applied to event studies to estimate the overall or horizon-specific ATT

Our contribution:

- Derive this estimator as the most efficient one
- Allow for restrictions on treatment effects
- More general class of estimands
- Extensions to other models
- Asymptotic theory with unit FEs

Outline

1. Setting and Framework
2. Conventional Practice and Associated Problems
3. Imputation-Based Estimation and Testing
4. Asymptotic Theory and Inference
5. Testing
6. Comparison to Other Estimators
7. Extensions

Extensions (1): Deviations from standard panel data

- Repeated cross-sections (i.e. different samples of i in the same groups g every year):
  $Y_{it}(0) = \alpha_{g(i)} + \beta_t + \varepsilon_{it}$

Extensions (1): Deviations from standard panel data

- Repeated cross-sections (i.e. different samples of i in the same groups g every year):
  $Y_{it}(0) = \alpha_{g(i)} + \beta_t + \varepsilon_{it}$

- Two-dimensional cross-sections: e.g. by region i and age group g in single period, different age threshoulds for eligilibity across regions ($D_{ig} = 1\,[g \geq E_i]$)
  $Y_{ig}(0) = \alpha_i + \alpha_g + \varepsilon_{ig}$

Extensions (1): Deviations from standard panel data

- Repeated cross-sections (i.e. different samples of i in the same groups g every year):
  $Y_{it}(0) = \alpha_{g(i)} + \beta_t + \varepsilon_{it}$

- Two-dimensional cross-sections: e.g. by region i and age group g in single period, different age threshoulds for eligilibity across regions ($D_{ig} = 1[g \geq E_i]$)
  $Y_{ig}(0) = \alpha_i + \alpha_g + \varepsilon_{ig}$

- Triple-differences: $Y_{igt}(0) = \alpha_{ig} + \alpha_{it} + \alpha_{gt} + \varepsilon_{it}$

Extensions (1): Deviations from standard panel data

- Repeated cross-sections (i.e. different samples of i in the same groups g every year):
  $Y_{it}(0) = \alpha_{g(i)} + \beta_t + \varepsilon_{it}$

- Two-dimensional cross-sections: e.g. by region i and age group g in single period, different age threshoulds for eligilibity across regions ($D_{ig} = 1\left[g \geq E_i\right]$)
  $Y_{ig}(0) = \alpha_i + \alpha_g + \varepsilon_{ig}$

- Triple-differences: $Y_{igt}(0) = \alpha_{ig} + \alpha_{it} + \alpha_{gt} + \varepsilon_{it}$

- Generic data: $Y_j(0) = Z_j'\pi + \varepsilon_j$

Extensions (2): Deviations from staggered timing

- Plain vanilla simultaneous treatment DiD (at $E_i = e$ or never): OLS has no negative weights but still inefficient and not robust to pre-testing — estimation and testing are conflated

Extensions (2): Deviations from staggered timing

- Plain vanilla simultaneous treatment DiD (at $E_i = e$ or never): OLS has no negative weights but still inefficient and not robust to pre-testing — estimation and testing are conflated

- Simultaneous treatment + unit-specific trends or controls: Same problems of OLS, imputation solution

Extensions (2): Deviations from staggered timing

- Plain vanilla simultaneous treatment DiD (at $E_i = e$ or never): OLS has no negative weights but still inefficient and not robust to pre-testing — estimation and testing are conflated

- Simultaneous treatment + unit-specific trends or controls: Same problems of OLS, imputation solution

- Treatment switching on and off: Imputation solution applies if no spillovers from treated to untreated periods

Extensions (2): Deviations from staggered timing

- Plain vanilla simultaneous treatment DiD (at $E_i = e$ or never): OLS has no negative weights but still inefficient and not robust to pre-testing — estimation and testing are conflated
- Simultaneous treatment + unit-specific trends or controls: Same problems of OLS, imputation solution
- Treatment switching on and off: Imputation solution applies if no spillovers from treated to untreated periods
- Multiple events per unit: Introduce appropriate A3 restrictions, e.g. that effects stabilize after P periods after the event (and before next event)

Recap: Framework to understand issues and organize solutions

1. Make assumptions explicit and separate them from goals
2. Relate problems with common practice to mismatch between assumptions and convenience regressions
   (building upon original Borusyak and Jaravel (2017) version)
3. Derive robust and efficient estimator from first principles, work out "imputation" structure (in baseline case)
4. Provide large-sample theory and inference for this estimator
5. Propose approach to testing, separate from estimation
6. Compare with alternative robust estimation strategies
7. Discuss extensions beyond event studies

Stata packages available:

```
ssc install did_imputation
ssc install event_plot
```

References

Arkhangelsky, D. and Imbens, G. W. (2019). Double-Robust Identification for Causal Panel Data Models.

Athey, S. and Imbens, G. W. (2018). Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption. *Working Paper*.

Baker, A. C., Larcker, D. F., and Wang, C. C. Y. (2021). How Much Should We Trust Staggered Difference-In-Differences Estimates ? *Working Paper*.

Borusyak, K. and Jaravel, X. (2017). Revisiting Event Study Designs. *Working Paper*.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-Differences with Multiple Time Periods and an Application on the Minimum Wage and Employment. *Journal of Econometrics*.

Cengiz, D., Dube, A., Lindner, A., and Zippere, B. (2019). The Effect of Minimum Wages on Low-Wage Jobs. *The Quarterly Journal of Economics*, pages 1405–1454.

De Chaisemartin, C. and D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–2996.

Gardner, J. (2021). Two-stage differences in differences. *Working Paper*.

Gobillon, L. and Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, 98(3):535–551.

Goodman-Bacon, A. (2018). Difference-in-Differences with Variation in Treatment Timing. *Working Paper*.

Liu, L., Wang, Y., and Xu, Y. (2020). A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data. *SSRN Electronic Journal*.

Rambachan, A. and Roth, J. (2020). An Honest Approach to Parallel Trends. *Working Paper*.

Roth, J. (2018). Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends. *Working Paper*, pages 1–55.

Roth, J. and Sant'Anna, P. H. (2020). When Is Parallel Trends Sensitive to Functional Form? *Working Paper*.

Roth, J. and Sant'Anna, P. H. C. (2021). Efficient Estimation for Staggered Rollout Designs. *Working Paper*.

Schmidheiny, K. and Siegloch, S. (2020). On Event Study Designs and Distributed-Lag Models: Equivalence , Generalization and Practical Implications. *Working Paper*.

Strezhnev, A. (2018). Semiparametric weighting estimators for multi-period difference-in-differences designs. *Working Paper*.

Sun, L. and Abraham, S. (2021). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Journal of Econometrics*.

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.