Static Linear Panel Data Models

230347 Advanced Microeconometrics
Tilburg University

Christoph Walsh

Static Linear Panel Model

- The static linear panel data model is:

$$y_{it} = \alpha + \boldsymbol{x}'_{it}\boldsymbol{\beta} + u_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

- The one-way error component model has individual-specific effects:

$$u_{it} = \alpha_i + \varepsilon_{it}$$

- The two-way error component model adds time effects:

$$u_{it} = \alpha_i + \lambda_t + \varepsilon_{it}$$

Pooled OLS

- The pooled OLS estimator stacks the NT observations and estimates $(\alpha, \boldsymbol{\beta})$ using OLS:

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

- If $\text{Cov}\,(u_{it}, \mathbf{x}_{it}) = 0$, then pooled OLS is a consistent estimator for $(\alpha, \boldsymbol{\beta})$.
- If the $\text{Cov}\,(u_{it}, u_{is}) \neq 0$ for $t \neq s$, then OLS standard errors will be too small.
  - Correlated observations contain less information than independent observations.
- Pooled OLS is inconsistent if the true model is fixed effects and the α_i are correlated with $\mathbf{x}_{it}$ as the error term is correlated with the regressors:

$$\begin{aligned} y_{it} &= \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + u_{it} \\ &= \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + (\alpha_i - \alpha + \varepsilon_{it}) \end{aligned}$$

Between Estimator

- The between estimator estimates $(\alpha, \boldsymbol{\beta})$ with OLS using the average of all observations by individual:

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T} y_{it}}_{=\bar{y}_i} = \alpha + \underbrace{\left(\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}\right)'}_{=\bar{x}_i'} \boldsymbol{\beta} + \underbrace{\frac{1}{T}\sum_{t=1}^{T} u_{it}}_{\bar{u}_i}$$

- The between estimator only uses cross-sectional variation to estimate $\boldsymbol{\beta}$:
- When $T = 1$, this is the same as a cross-section regression.
- The between estimator is consistent only if the composite error term $\bar{u}_i$ is independent of the regressors $\bar{\boldsymbol{x}}_i$.

Within Estimator (Fixed Effects Estimator)

- The within estimator demeans by individual to remove the individual fixed effects α_i:

$$y_{it} - \bar{y}_i = \underbrace{\alpha_i - \frac{1}{T}\sum_{t=1}^{T}\alpha_i}_{=0} + (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_i$$

- The above model can then be estimated with OLS.
- The within estimator is a consistent estimator of $\boldsymbol{\beta}$ if $N \to \infty$ while T is fixed, even if the α_i are correlated with $\boldsymbol{x}_{it}$.

Limitations:

- Inability to estimate the effects of time-invariant regressors.
- If the α_i are independent of the $\boldsymbol{x}_{it}$, then random effects is more efficient.

Within Estimator for Two-Way Errors

- Suppose the composite error term is now $u_{it} = \alpha_i + \lambda_t + \varepsilon_{it}$.
- If T is fixed and $N \to \infty$, we could consistently estimate λ_t with dummies for each time period.
- However, we can also apply a double within transformation to remove both the α_i and λ_t.
  - The transformation is $y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$, where $\bar{y} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}$.
- With λ_t, we can no longer estimate effects that are individual-invariant within a time period.
  - For example, macroeconomic variables or a time trend.

First-Differences Estimator

- The first-differences estimator removes the fixed effect α_i by subtracting the lag of the model $y_{i,t-1} = \alpha_i + \mathbf{x}'_{i,t-1}\boldsymbol{\beta} + \varepsilon_{i,t-1}$:

$$y_{i,t} - y_{i,t-1} = (\mathbf{x}_{i,t} - \mathbf{x}_{i,t-1})' \boldsymbol{\beta} + \varepsilon_{i,t} - \varepsilon_{i,t-1} \quad i = 1, \ldots, N \quad t = 2, \ldots, T$$

  and estimating with OLS.
- The first-differences estimator is consistent as $N \rightarrow \infty$ with T fixed.
- When $T = 2$, both estimators produce identical estimates and standard errors.
- If the $\varepsilon_{i,t}$ are iid, the within estimator is more efficient.
  - But estimation with first differences can be efficient if using GLS.
- If $\varepsilon_{i,t} = \varepsilon_{i,t-1} + e_{i,t}$ ($\varepsilon_{i,t}$ is a random walk) and $e_{i,t} = \Delta\varepsilon_{i,t}$ is iid, then the first-differences estimator is more efficient.

Random Effects Estimator

The model is:

$$y_{it} = \mu + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}$$

- The random effects model treats $\alpha_i + \varepsilon_{it}$ as a composite error term.
  - α_i is iid and has mean and variance 0 and σ^2_α respectively.
  - ε_{it} is iid and has mean and variance 0 and σ^2_ε respectively.
- The RE model allows for the estimation of time-invariant regressors.
- If the α_i are correlated with the $\boldsymbol{x}_{it}$, the RE model is inconsistent.
- If the RE model is consistent, the FE model is also consistent but less efficient.

FGLS Estimation of RE Model (With Balanced Panels)

- *Step 1:* Estimate model with fixed effects and use the residuals to compute:

$$\widehat{\sigma}_{\varepsilon}^{2} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left[(y_{it} - \bar{y}_i) - (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)' \, \widehat{\boldsymbol{\beta}}_{FE} \right]^2}{N(T-1) - K}$$

- *Step 2:* Estimate between model and use the residuals to compute:

$$\widehat{\sigma}_{\alpha}^{2} = \frac{\sum_{i=1}^{N} (\bar{y}_i - \mu_B - \bar{\boldsymbol{x}}_i' \boldsymbol{\beta}_B)^2}{N - (K+1)} - \frac{1}{T} \widehat{\sigma}_{\varepsilon}^{2}$$

- *Step 3:* Compute: $\widehat{\lambda} = 1 - \widehat{\sigma}_{\varepsilon} / \sqrt{T \widehat{\sigma}_{\alpha}^{2} + \widehat{\sigma}_{\varepsilon}^{2}}$

- *Step 4:* Estimate the model below with OLS to obtain the RE estimates $\left(\widehat{\mu}_{RE}, \widehat{\boldsymbol{\beta}}_{RE} \right)$:

$$y_{it} - \widehat{\lambda} \bar{y}_i = \left(1 - \widehat{\lambda}\right) \mu + \left(\boldsymbol{x}_{it} - \widehat{\lambda} \bar{\boldsymbol{x}}_i \right)' \boldsymbol{\beta} + \nu_{it}$$

Hausman (1978) Test

- Since FE is consistent and RE inconsistent when the individual effects are correlated with the covariates, the difference between $\widehat{\boldsymbol{\beta}}_{FE}$ and $\widehat{\boldsymbol{\beta}}_{RE}$ can be used to test for this correlation.
- The Hausman test statistic is:

$$\left(\widehat{\boldsymbol{\beta}}_{FE}-\widehat{\boldsymbol{\beta}}_{RE}\right)'\left[\widehat{\text{Avar}}\left(\widehat{\boldsymbol{\beta}}_{FE}\right)-\widehat{\text{Avar}}\left(\widehat{\boldsymbol{\beta}}_{RE}\right)\right]^{-1}\left(\widehat{\boldsymbol{\beta}}_{FE}-\widehat{\boldsymbol{\beta}}_{RE}\right)\sim\chi^2_K$$

  where K is the number of elements in $\boldsymbol{\beta}$.
- Note that difference in variances is positive definite as RE is more efficient than FE.
- A rejection of the test suggests you should reject the RE estimates in favour of the FE estimates.

Difference-in-Differences

	Treatment $(g_i = 1)$	Control $(g_i = 0)$
Pre-treatment $(t = 1)$	$\bar{y}^{treatment}_{pre}$	$\bar{y}^{control}_{pre}$
Post-treatment $(t = 2)$	$\bar{y}^{treatment}_{post}$	$\bar{y}^{control}_{post}$

- If the treatment is randomly assigned, its effect can be estimated with:

$$\left(\bar{y}^{treatment}_{post} - \bar{y}^{treatment}_{pre}\right) - \left(\bar{y}^{control}_{post} - \bar{y}^{control}_{pre}\right)$$

- We can also estimate it in a regression, where $Post_t = \mathbb{1}\left\{t = 2\right\}$:

$$y_{it} = \beta_0 + \beta_1 g_i + \beta_2 Post_t + \beta_3 g_i Post_t + \varepsilon_{it}$$

- Or with first differences:

$$y_{i2} - y_{i1} = \beta_2 + \beta_3 g_i + (\varepsilon_{i2} - \varepsilon_{i1})$$

Generalized Difference-in-Differences

- We often observe several periods before and after treatment.
- We also may have several control groups (and treatment groups).
- This can be estimated with:

$$y_{ist} = \beta D_{st} + \boldsymbol{x}'_{ist}\boldsymbol{\gamma} + \alpha_s + \delta_t + \varepsilon_{ist}$$

  where s denotes state/region and $D_{st} \in \{0, 1\}$ denotes the having received treatment.
- This model can be estimated with repeated cross sections.
  - If panel data are available, α_s can be replaced by α_i.
- We will see, however, that estimating this generalized model has many problems.

Serial Correlation in Difference-in-Differences

Bertrand et al. (2004)

- If there are several time periods before and after the policy change, serial correlation makes standard errors too small.
- Bertrand, Duflo, and Mullainathan (2004) generate placebo laws in the CPS data and find "effects" significant at the 5% level for almost half of their placebo interventions when using conventional standard errors.
- Different methods can be used to correct the standard errors:
  - Use a state-level block bootstrap.
  - Use clustered standard errors at the state level.
  - With a small number of clusters (5-30) Cameron, Gelbach, and Miller (2008) provide a wild bootstrap method.
  - Aggregate data by state into two periods: pre- and post-intervention.
    - Also found to work well for small number (≈ 10) of states.
  - See also Cameron and Miller (2015) for an overview of cluster-robust inference.

The Bootstrap and Block Bootstrap

- NT observations: $\{x_{1c1}, \ldots, x_{1cT}, \ldots, x_{Nc'1}, \ldots, x_{Nc'T}\}$
- Each i has an associated cluster c (such as the state), with C clusters in total.

Regular Bootstrap:

1. Draw NT observations from the sample at random *with replacement*.
2. Estimate β with the bootstrap sample. Call it $\widehat{\beta}_b$. Repeat B times.
3. The standard error can then be calculated as $\sqrt{\sum_{b=1}^{B} \frac{\left[\widehat{\beta}_b - \left(\frac{1}{B}\sum_{b=1}^{B}\widehat{\beta}_b\right)\right]^2}{B-1}}$

Block Bootstrap:

- Instead of drawing observations (i, t) independently, you draw entire clusters (blocks).
- You draw C entire clusters with replacement, where for each cluster you take all observations (i, t) in that cluster.
- This maintains the structure of the correlation between observations within a cluster.

Clustered Standard Errors

- For iid errors, recall that $\widehat{\text{Var}}\left(\widehat{\boldsymbol{\beta}}\right) = \widehat{\sigma}^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$.
- For heteroskedastic but independent errors, the Huber-Ecker-White variance-covariance matrix for $\widehat{\boldsymbol{\beta}}$ is:

$$\widehat{\text{Var}}\left(\widehat{\boldsymbol{\beta}}\right) = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left[\sum_{i=1}^{N}\sum_{t=1}^{T} \widehat{\varepsilon}_{igt}^{2} \boldsymbol{x}_{ict} \boldsymbol{x}'_{ict}\right] \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

- If we want to allow correlation within a group g, i.e.:

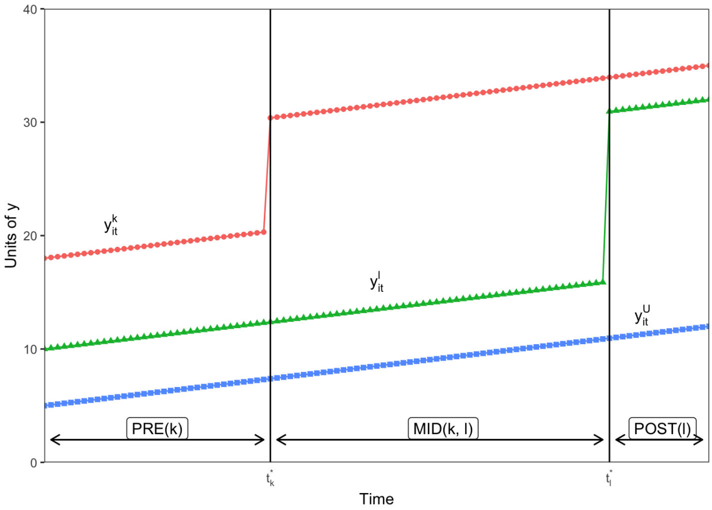
$$\mathbb{E}\left[\varepsilon_{ict}\varepsilon_{jc's}\right] = \begin{cases} \sigma_{it,js} & \text{if } c = c' \quad \text{(same group/cluster)} \\ 0 & \text{if } c \neq c' \quad \text{(different group/cluster)} \end{cases}$$

Then:

$$\widehat{\text{Var}}\left(\widehat{\boldsymbol{\beta}}\right) = \left(\frac{C}{C-1}\right)\left(\frac{NT-1}{NT-2}\right) \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \left[\sum_{c=1}^{C} \boldsymbol{x}'_c \widehat{\boldsymbol{\varepsilon}}_c \widehat{\boldsymbol{\varepsilon}}'_c \boldsymbol{x}_c\right] \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

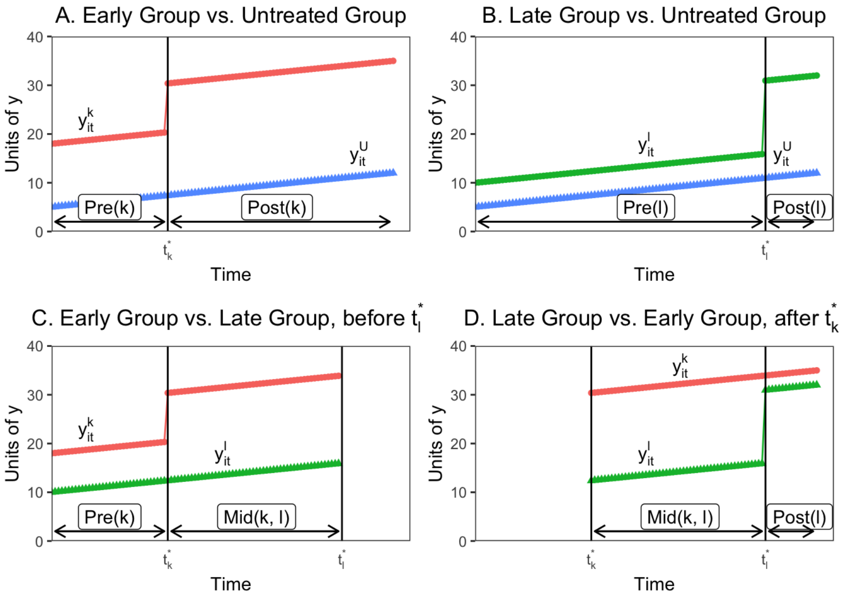
Staggered Difference-in-Differences

- We often have situations where a policy is rolled out in a staggered manner.



Source: Goodman-Bacon (2021)

Four 2 × 2 Difference-in-Differences



TWFE gives a weighted average of all 2×2 DiD estimates

- Suppose we have a model with only one observation per state and no covariates:

$$y_{st} = \beta D_{st} + \alpha_s + \delta_t + \varepsilon_{st}$$

- De Chaisemartin and d'Haultfoeuille (2020) show that the estimator for β is a weighted average of the individual 2×2 difference-in-difference estimates.
- The weight for the effect of the treatment on state s in time t is proportional to:

$$\widetilde{D}_{st} = D_{st} - \frac{1}{T}\sum_{t=1}^{T} D_{st} - \frac{1}{S}\sum_{s=1}^{S} D_{st} + \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T} D_{st}$$

- If $\frac{1}{T}\sum_{t=1}^{T} D_{st}$ and $\frac{1}{S}\sum_{s=1}^{S} D_{st}$ are large, it is possible for weights to be negative!

Negative Weighting Example

- Recall that the weights are proportional to:

$$\widetilde{D}_{st} = D_{st} - \frac{1}{T}\sum_{t=1}^{T} D_{st} - \frac{1}{S}\sum_{s=1}^{S} D_{st} + \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T} D_{st}$$

- Consider the following $S = 2$ and $T = 3$ example:

State	$t = 1$	$t = 2$	$t = 3$
$s = 1$	0	1	1
$s = 2$	0	0	1

- Then:

$$\widetilde{D}_{1,2} = 1 - \frac{2}{3} - \frac{1}{2} + \frac{1}{2} = \frac{1}{3} \qquad \widetilde{D}_{1,3} = 1 - \frac{2}{3} - 1 + \frac{1}{2} = -\frac{1}{6} \qquad \widetilde{D}_{2,3} = 1 - \frac{1}{3} - 1 + \frac{1}{2} = \frac{1}{6}$$

- The normalized weights are then $+1$, $-1/2$, $+1/2$, summing to 1.
- Suppose treatment effects are dynamic so that one year after treatment the effect on y is 4 on average, whereas the effect is only 1 in the year of treatment. Then the weighted sum is $1 \times 1 - \frac{1}{2} \times 4 + \frac{1}{2} \times 1 = -\frac{1}{2}$. The negative weight can flip the sign!

Staggered Difference-in-Differences

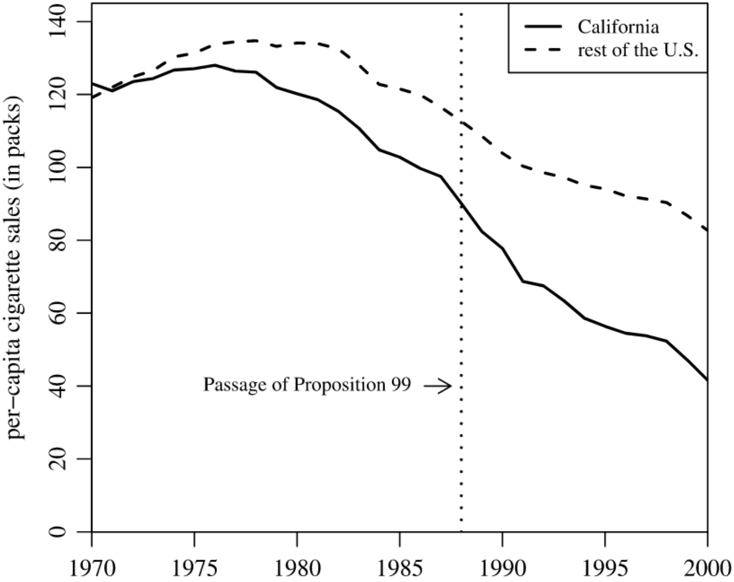
- If treatment effects are heterogeneous (across units or over time), negative weights can lead to situations where all the individual ATTs are positive but the weighted average is negative.
- Furthermore, if treatment effects are dynamic (e.g., the effect grows larger over time), already-treated units make bad comparison groups for newly treated units.
- Callaway and Sant'Anna (2021) provide a method that alleviates these concerns (`did` package in R).
  - The approach involves estimating all possible 2×2 DiD regressions with OLS.
  - It does not use earlier-treated units as a control group for later-treated units. It only uses never-treated units (if available) or not-yet-treated units as control.
  - It aggregates all these estimates with equal weights, avoiding the "negative weight problem."
  - To be estimated in the computer assignment!

Synthetic Controls

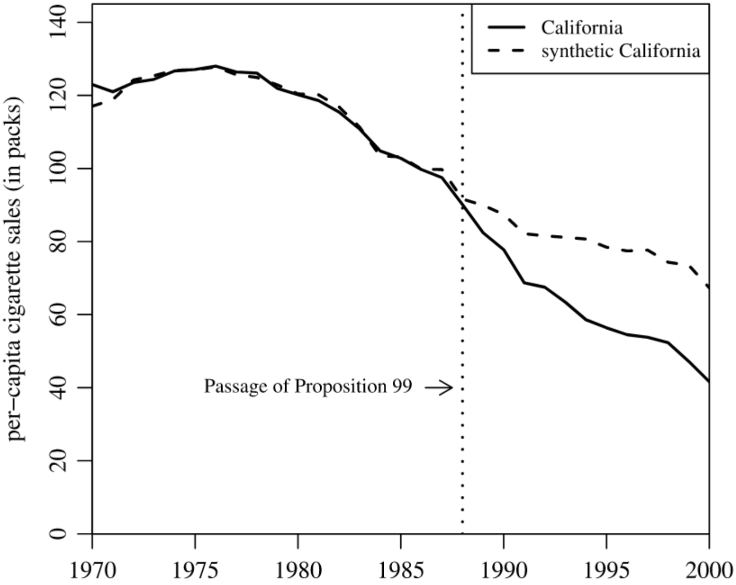
Abadie et al. (2010)

- Sometimes people use all untreated states as the control groups. Other times people use neighboring states.
- Choosing control groups can be arbitrary so it may be better to let the data choose the correct control group(s).
- The synthetic control method uses a data-drived approach to choose the optimal convex combination of potential control groups (the *donor pool*) from pre-period data.
- Abadie, Diamond, and Hainmueller (2010) study the effects of Proposition 99, a large-scale tobacco control program, on cigarette sales in California.
- They first exclude other states that had tobacco-related policy changes in the post period. They also exclude DC.

Trends in per-capita cigarette sales: Rest of US



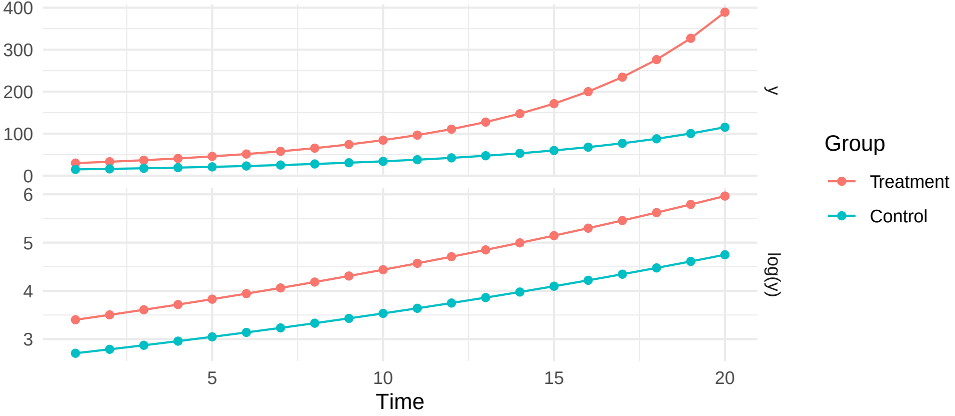
Trends in per-capita cigarette sales: Synthetic CA



Changes-in-Changes

Athey and Imbens (2006)

- DiD is functional form dependent. The common trend assumption may hold for the log but not the level:



- Athey and Imbens (2006) propose a generalization of the difference-in-differences method.

Changes-in-Changes: Setup

- The data is (Y_i, G_i, T_i) for $i = 1, \ldots, N$.
- Y_i is the outcome variable.
- $G_i \in \{0, 1\}$ denotes whether i is in the control or treatment group.
- $T_i \in \{0, 1\}$ is time (before and after treatment).
- Under Rubin's potential outcomes framework: $Y_i = G_i T_i Y_i^I + (1 - G_i T_i) Y_i^N$
  - For each individual we observe only one of Y_i^I and Y_i^N.
  - We call the unobserved outcome the counterfactual outcome.

Changes-in-Changes: Assumptions

- Write (Y, G, T, U) as random variables, where U is unobserved heterogeneity.
- They assume the following:
  - **Model:** Untreated outcomes are a function of unobserved heterogeneity and time:

$$Y^N = h(U, T)$$

  - **Strict Monotonicty:** $h(u, t)$ is strictly increasing in u for $t \in \{0, 1\}$.
  - **Time Invariance Within Groups:** $U \perp\!\!\!\perp T | G$
    - Each i's ranking in the distrubtion of Y_i^N is the same in both time periods.
  - **Support:** $\mathbb{U}_1 \subseteq \mathbb{U}_0$.

Changes-in-Changes: Counterfactual Distribution of Treated Group

- Let $F_{Y,gt}$ be the cdf of Y for group g at time t.
- Given the assumptions, they show that the distribution of Y_{11}^{N} is identified:

$$F_{Y^N,11}(y) = F_{Y,10}\left(F_{Y,00}^{-1}\left(F_{Y,01}(y)\right)\right)$$

- $F_{Y^N,11}(y)$ can be estimated by taking the empirical cdf of $\widehat{F}_{Y,01}^{-1}\left(\widehat{F}_{Y,00}(Y_{10})\right)$.
  - *Note:* The empirical cdf of n iid draws of a random variable X_i is: $\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{N} \mathbb{1}\{X_i \leq x\}$
- Since $F_{11}(y) = F_{Y^I,11}(y)$, you can use the empirical cdf of Y_{11} to estimate $\widehat{F}_{Y^I,11}(y)$.

Changes-in-Changes: Average Treatment Effects

- The average treatment effect on the treated is then:

$$\begin{aligned}\tau^{CIC} &= \mathbb{E}\left[Y_{11}^{I}\right] - \mathbb{E}\left[Y_{11}^{N}\right] \\ &= \mathbb{E}\left[Y_{11}^{I}\right] - \mathbb{E}\left[F_{Y,01}^{-1}\left(F_{Y,00}\left(Y_{10}\right)\right)\right]\end{aligned}$$

- The sample analogue is:

$$\widehat{\tau}^{CIC} = \frac{1}{N_{11}} \sum_{i=1}^{N_{11}} Y_{11,i} - \frac{1}{N_{10}} \sum_{i=1}^{N_{10}} \widehat{F}_{Y,01}^{-1}\left(\widehat{F}_{Y,00}\left(Y_{10,i}\right)\right)$$

Changes-in-Changes: Quantile Treatment Effects

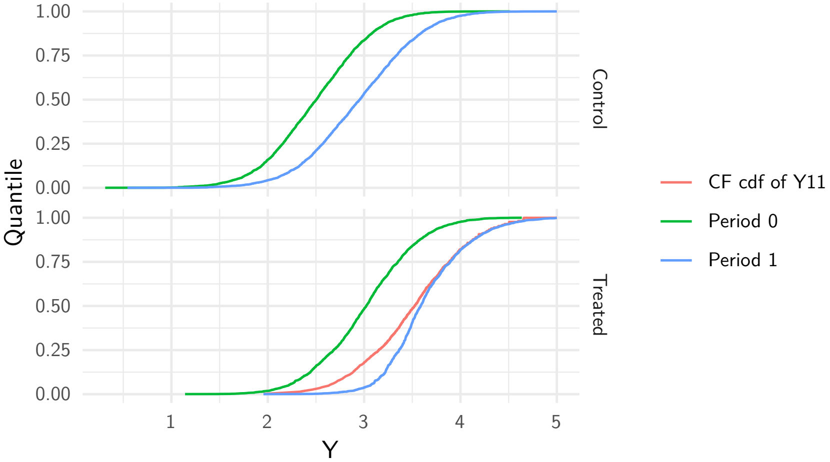
- The effect at quantile q is:

$$\tau_q^{CIC} = F^{-1}_{Y^I,11}(q) - F^{-1}_{Y^N,11}(q)$$

- Using $F_{Y^N,11}(y) = F_{Y,10}\left(F^{-1}_{Y,00}(F_{Y,01}(y))\right)$, we can calculate the sample analogue of this with:

$$\widehat{\tau}_q^{CIC} = \widehat{F}^{-1}_{Y,11}(q) - \widehat{F}^{-1}_{Y,01}\left(\widehat{F}_{Y,00}\left(\widehat{F}^{-1}_{Y,10}(q)\right)\right)$$

Changes-in-Changes: Graphical Illustration



References

- Baltagi, Ch. 2 and beginning of Ch. 3
- Cameron and Trivedi, Ch. 21

Abadie, A., A. Diamond, and J. Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American statistical Association*, 105, 493–505.

Athey, S. and G. W. Imbens (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74, 431–497.

Bertrand, M., E. Duflo, and S. Mullainathan (2004): "How Much Should we Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119, 249–275.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225, 200–230.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-Based Improvements for Inference with Clustered Errors," *The Review of Economics and Statistics*, 90, 414–427.

Cameron, A. C. and D. L. Miller (2015): "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 50, 317–372.

De Chaisemartin, C. and X. d'Haultfoeuille (2020): "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110, 2964–96.

Goodman-Bacon, A. (2021): "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225, 254–277.

Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 1251–1271.

Note: Probability Limit & Consistency

- **Probability Limit:**
  A sequence of random variables $\{x_N : N = 1, 2, \ldots\}$ *converges in probability* to the constant a if for all $\varepsilon > 0$, $\Pr\left(|x_N - a| > \varepsilon\right) \to 0$ as $N \to \infty$.
  - We can write this as: $x_N \xrightarrow{p} a$ or $\text{plim } x_N = a$.
- **Consistency:**
  Let $\{\boldsymbol{\theta}_N : N = 1, 2, \ldots\}$ be a sequence of estimators of $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where N indexes sample size. If $\boldsymbol{\theta}_N \xrightarrow{p} \boldsymbol{\theta}$ for any value of $\boldsymbol{\theta}$, then we say $\widehat{\boldsymbol{\theta}}_N$ is a *consistent estimator* of $\boldsymbol{\theta}$.
  - *Example:* Suppose x_N is the average outcome of N coin tosses. Then $x_N \xrightarrow{p} \frac{1}{2}$.