Difference-in-Differences

Causal Inference using Machine Learning
Master in Economics, UNT

Andres Mena

Spring 2024

What is Panel Data?

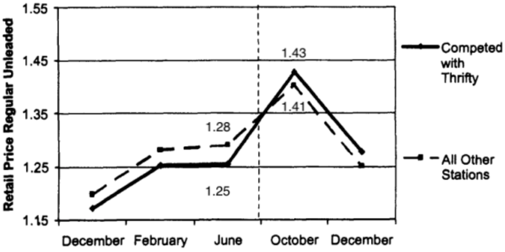
Panel Data Definition:

- Observations for each unit i (e.g., person, state) across multiple time periods t.
- Allows us to account for unobservable characteristics that are constant over time.

What is Panel Data?

Panel Data Definition:

- Observations for each unit i (e.g., person, state) across multiple time periods t.
- Allows us to account for unobservable characteristics that are constant over time.
- Example: Gas prices in neighborhoods with and without Thrifty stations, observed before and after a merger.



Defining Treatment in Panel Data

- **Treatment Definition:**
  - Let Y_{it} be the outcome for unit i at time t.
  - $D_i = 1$: Treated units.
  - $D_i = 0$: Control units.

Defining Treatment in Panel Data

- **Treatment Definition:**
  - Let Y_{it} be the outcome for unit i at time t.
  - $D_i = 1$: Treated units.
  - $D_i = 0$: Control units.
- **Key Assumptions:**
  - No anticipation: Treatment in t does not affect outcomes in earlier periods.
  - Parallel trends: Control and treatment groups would have followed the same trend in the absence of treatment.

Conditional Expectations Before and After Treatment

- **Observed Outcomes:**

$$Y_{it} = G_{it} Y_{it}(1) + (1 - G_{it}) Y_{it}(0)$$

- **Parallel Trends Assumption**

$$E[Y_{it}(0) - Y_{it-j}(0) | D_i = 1] = E[Y_{it}(0) - Y_{it-j}(0) | D_i = 0]$$

Conditional Expectations Before and After Treatment

- **Observed Outcomes:**

$$Y_{it} = G_{it} Y_{it}(1) + (1 - G_{it}) Y_{it}(0)$$

- **Parallel Trends Assumption**

$$E[Y_{it}(0) - Y_{it-j}(0) | D_i = 1] = E[Y_{it}(0) - Y_{it-j}(0) | D_i = 0]$$

- **No anticpation**

$$E(Y_{it} | D_i = 1) = E(Y_{it}(0) | D_i = 1) \quad \text{for all } t < T$$

Conditional Expectations Before and After Treatment

- **Observed Outcomes:**

$$Y_{it} = G_{it} Y_{it}(1) + (1 - G_{it}) Y_{it}(0)$$

- **Parallel Trends Assumption**

$$E[Y_{it}(0) - Y_{it-j}(0)|D_i = 1] = E[Y_{it}(0) - Y_{it-j}(0)|D_i = 0]$$

- **No anticpation**

$$E(Y_{it}|D_i = 1) = E(Y_{it}(0)|D_i = 1) \quad \text{for all } t < T$$

- **Average Treatment on The treated**

$$E[Y_{it}(1) - Y_{it}(0)|D_i = 1]$$

Identification in Differences-in-Differences (DiD) - Part 1

Step-by-Step Setup

- **Notation:**
  - Y_{it}: Observed outcome for unit i at time t.
  - D_i: Treatment indicator ($D_i = 1$ if treated, $D_i = 0$ if control).
- **Simple Difference (Within-Group Change):**

$$\Delta Y_i = Y_{i2} - Y_{i1}.$$

$$E[\Delta Y_i | D_i = 1] = E[Y_{i2} - Y_{i1} | D_i = 1], \quad E[\Delta Y_i | D_i = 0] = E[Y_{i2} - Y_{i1} | D_i$$

- **Double Difference (Between-Group Change):**

$$\text{DiD} = \big(E[Y_{i2}|D_i = 1] - E[Y_{i1}|D_i = 1]\big) - \big(E[Y_{i2}|D_i = 0] - E[Y_{i1}|D_i = 0]\big)$$

Identification in DiD - Part 1

$$\text{DiD} = \big[E(Y_2|D=1) - E(Y_1|D=1)\big] - \big[E(Y_2|D=0) - E(Y_1|D=0)\big].$$

Potential Outcomes:

$$Y_{it} = G_{it} Y_{it}(1) + (1 - G_{it}) Y_{it}(0)$$

Replacing Observed with Potential Outcomes:

- For $D_i = 1$ and $t = 2$: $E(Y_{i2}|D_i = 1) = E(Y_{i2}(1)|D_i = 1)$.
- For $D_i = 1$ and $t = 1$: $E(Y_{i1}|D_i = 1) = E(Y_{i1}(0)|D_i = 1)$. (No Anticipation)
- For $D_i = 0$: $E(Y_{it}|D_i = 0) = E(Y_{it}(0)|D_i = 0)$ for all t

Thus:

$$\text{DiD} = \big[E(Y_2(1)|D=1) - E(Y_1(0)|D=1)\big] - \big[E(Y_2(0)|D=0) - E(Y_1(0)|D=0)\big].$$

Identification in DiD - Part 2

Substrac and Add $E(Y_2(0)|D=1) - E(Y_1(0)|D=1)$**:**

$$\text{DiD} = \underbrace{\left[E(Y_2(1) - Y_2(0)|D=1)\right]}_{\text{ATT}}$$

$$+ \underbrace{\left[E(Y_1(0) - Y_1(0)|D=1)\right]}_{=0}$$

$$+ \underbrace{\left[E(Y_2(0) - Y_1(0)|D=1) - E(Y_2(0) - Y_1(0)|D=0)\right]}_{\text{Selection Bias}}.$$

Identification in DiD - Part 2

Substrac and Add $E(Y_2(0)|D=1) - E(Y_1(0)|D=1)$:

$$\text{DiD} = \underbrace{\left[E(Y_2(1) - Y_2(0)|D=1)\right]}_{\text{ATT}}$$

$$+ \underbrace{\left[E(Y_1(0) - Y_1(0)|D=1)\right]}_{=0}$$

$$+ \underbrace{\left[E(Y_2(0) - Y_1(0)|D=1) - E(Y_2(0) - Y_1(0)|D=0)\right]}_{\text{Selection Bias}}.$$

Parallel Trends Assumption:

$$E(Y_2(0) - Y_1(0)|D=1) = E(Y_2(0) - Y_1(0)|D=0) \implies \text{Selection Bias} = 0.$$

Identification in DiD - Part 2

Substrac and Add $E(Y_2(0)|D=1) - E(Y_1(0)|D=1)$:

$$\text{DiD} = \underbrace{\left[E(Y_2(1) - Y_2(0)|D=1)\right]}_{\text{ATT}}$$

$$+ \underbrace{\left[E(Y_1(0) - Y_1(0)|D=1)\right]}_{=0}$$

$$+ \underbrace{\left[E(Y_2(0) - Y_1(0)|D=1) - E(Y_2(0) - Y_1(0)|D=0)\right]}_{\text{Selection Bias}}.$$

Parallel Trends Assumption:

$$E(Y_2(0) - Y_1(0)|D=1) = E(Y_2(0) - Y_1(0)|D=0) \implies \text{Selection Bias} = 0.$$

Conclusion:

$$\text{DiD} = E(Y_2(1) - Y_2(0)|D=1) = \text{ATT}.$$

Two-Way Fixed Effects Model

Regression Specification:

$$Y_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 D_i + \beta_3 (D_i \times \text{Post}_t) + \epsilon_{it}$$

Two-Way Fixed Effects Model

Regression Specification:

$$Y_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 D_i + \beta_3 (D_i \times \text{Post}_t) + \epsilon_{it}$$

Coefficients:

- β_0: Baseline outcome for control group in pre-treatment period.
- β_1: Time trend for control group.
- β_2: Baseline difference between treated and control groups.
- β_3: Treatment effect (DiD estimate).

Two-Way Fixed Effects Model

Regression Specification:

$$Y_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 D_i + \beta_3 (D_i \times \text{Post}_t) + \epsilon_{it}$$

Coefficients:

- β_0: Baseline outcome for control group in pre-treatment period.
- β_1: Time trend for control group.
- β_2: Baseline difference between treated and control groups.
- β_3: Treatment effect (DiD estimate).

Extensions:

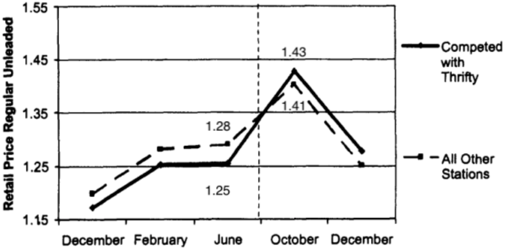
- Add unit and time fixed effects to control for unobserved heterogeneity.

Example: Hastings (2004)

- Studied gas price changes after ARCO's merger with Thrifty.
- Pre-treatment: Gas stations in areas competing with Thrifty had prices 3 cents lower.
- Post-treatment: Prices in treated areas increased by 2 cents.

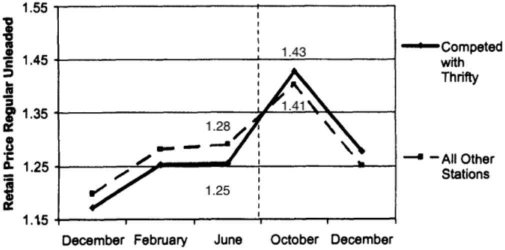
Example: Hastings (2004)

- Studied gas price changes after ARCO's merger with Thrifty.
- Pre-treatment: Gas stations in areas competing with Thrifty had prices 3 cents lower.
- Post-treatment: Prices in treated areas increased by 2 cents.



Example: Hastings (2004)

- Studied gas price changes after ARCO's merger with Thrifty.
- Pre-treatment: Gas stations in areas competing with Thrifty had prices 3 cents lower.
- Post-treatment: Prices in treated areas increased by 2 cents.



$$\text{Effect} = \text{Post-treatment difference} - \text{Pre-treatment difference}$$
$$= 2 - (-3) = 5 \text{ cents}$$

Inference in DiD: Clustered Standard Errors

$$\widehat{V}_{\text{cluster}}(\hat{\beta}) = (X^\top X)^{-1} \left(\sum_{g=1}^{G} X_g^\top \widehat{\epsilon}_g \widehat{\epsilon}_g^\top X_g \right) (X^\top X)^{-1},$$

where g indexes clusters, $\widehat{\epsilon}_g$ are residuals within cluster g, and X_g is the design matrix for cluster g.

Conclusion

- DiD is a powerful tool for causal inference in panel data.

Conclusion

- DiD is a powerful tool for causal inference in panel data.
- Assumptions: Parallel trends and no anticipation.

Conclusion

- DiD is a powerful tool for causal inference in panel data.
- Assumptions: Parallel trends and no anticipation.
- Two-way fixed effects extend DiD to handle more complex settings.

Conclusion

- DiD is a powerful tool for causal inference in panel data.
- Assumptions: Parallel trends and no anticipation.
- Two-way fixed effects extend DiD to handle more complex settings.
- Inferences should use standard errors that account for clustering.