ECON 251 Office Hours

Review + Extensions

Elird Haxhiu

Fall 2022

Outline

1. Review instrumental variables (IV)
2. Review panel methods + binary treatment diff-in-diff (DD)
3. Continuous treatment DD (Haxhiu & Helgerman, 2022)
4. Extra theory + Stata examples to sharpen tools

Mincer (1974) model of earnings $\log Y_i = \alpha + \beta \cdot S_i + U_i$

Instrumental variable (IV) Z_i decomposes S_i into S_i^{X} and S_i^{N}

Mincer (1974) model of earnings $\log Y_i = \alpha + \beta \cdot S_i + U_i$

Instrumental variable (IV) Z_i decomposes S_i into S_i^{X} and S_i^{N}

- First stage generates predicted values for treatment $\hat{S}_i \coloneqq \hat{\pi}_0 + \hat{\pi}_1 Z_i$
- We estimate returns β from model $\log Y_i = \alpha + \beta \cdot \hat{S}_i + U_i$

Mincer (1974) model of earnings $\log Y_i = \alpha + \beta \cdot S_i + U_i$

Instrumental variable (IV) Z_i decomposes S_i into S_i^{X} and S_i^{N}

- First stage generates predicted values for treatment $\hat{S}_i \coloneqq \hat{\pi}_0 + \hat{\pi}_1 Z_i$
- We estimate returns β from model $\log Y_i = \alpha + \beta \cdot \hat{S}_i + U_i$

- A valid instrument satisfies
    1. <u>Relevance</u> $\mathrm{Cov}(Z_i, S_i) \neq 0$
    2. <u>Exogeneity</u> $\mathrm{Cov}(Z_i, U_i) = 0$
    3. <u>Exclusion</u> no direct effect of Z_i on Y_i

IV in practice!

Suppose we have access to an instrument Z that is relevant, exogenous, and excluded for treatment X in the linear model $Y = \beta_0 + \beta_1 X + U$. Using *Stata*, verify that the slope β_1 in this model can be estimated in three equivalent ways:

- Ratio of instrument covariance with outcome $\widehat{\mathrm{Cov}}(Z, Y)$ and treatment $\widehat{\mathrm{Cov}}(Z, X)$
- Ratio of reduced form ($Y = \delta_0 + \delta_1 Z + V$) and first stage ($X = \pi_0 + \pi_1 Z + W$) slope coefficients estimated via ordinary least squares $\frac{\hat{\delta}_1^{\mathrm{OLS}}}{\hat{\pi}_1^{\mathrm{OLS}}}$
- Slope of alternative outcome model ($Y = \tilde{\beta}_0 + \tilde{\beta}_1 \cdot \hat{X} + \tilde{U}$) where we use predicted treatment $\hat{X} := \hat{\pi}_0 + \hat{\pi}_1 Z$ from first stage regression

Outline

1. Review instrumental variables (IV)
2. Review panel methods + binary treatment diff-in-diff (DD)
3. Continuous treatment DD (Haxhiu & Helgerman, 2022)
4. Extra theory + Stata examples to sharpen tools

Repeated sampling of population over time (usually)

Two dimensions ($i + t$) relating outcome Y_{it} to treatment X_{it}

Repeated sampling of population over time (usually)

Two dimensions ($i + t$) relating outcome Y_{it} to treatment X_{it}

- Data = <u>pooled cross-section</u> (when new units are sampled each period)
  = <u>panel</u> (when we track the same units)

Repeated sampling of population over time (usually)

Two dimensions ($i + t$) relating outcome Y_{it} to treatment X_{it}

- Data = pooled cross-section (when new units are sampled each period)
  = panel (when we track the same units)

- Random Effects: OLS on $Y_{it} = \theta_t + \beta \cdot X_{it} + U_{it}$ with dummy variables for time periods (interacting with treatment to assess structural change)

- Fixed Effects: OLS on $Y_{it} = \alpha_i + \theta_t + \beta \cdot X_{it} + U_{it}$ with dummy variables for time periods and individual units (if we have a panel) or exogenously defined groups of units (if we have a pooled cross-section)

Panel methods in practice!

Suppose we have a panel of N units across $T > 1$ periods and estimate β in the two-way fixed effects (TWFE) model $Y_{it} = \alpha_i + \theta_t + \beta \cdot X_{it} + U_{it}$. Why do multiple observations of units over time deal with <u>all time-invariant</u> omitted variables in the error term?

Discuss the assumptions needed to consistently estimate β in the two-way fixed effects (TWFE) model $Y_{it} = \alpha_i + \theta_t + \beta \cdot X_{it} + U_{it}$ when $T = 2$ via:

- OLS on <u>first-differenced</u> variables $\quad \Delta Y_{it} = \Delta\theta_t + \beta \cdot \Delta X_{it} + \Delta U_{it}$
- OLS on <u>time-demeaned</u> variables $\quad [Y_{it} - \overline{Y}_i] = \beta \cdot [X_{it} - \overline{X}_i] + [U_{it} - \overline{U}_i]$
- OLS with unit/period <u>dummy variables</u> $\quad Y_{it} = \alpha_i + \theta_t + \beta \cdot X_{it} + U_{it}$

Difference-in-differences = compare Y change of units exposed to some policy T with Y change of unexposed

Difference-in-differences = compare Y change of units exposed to some policy T with Y change of unexposed

2 periods (before/after) and 2 groups (treated/control)

Y_{it} ≔ outcome of interest

$P_t \coloneqq 1\{t \text{ is after treatment occurs}\}$

$T_i \coloneqq 1\{i \text{ is treated/exposed}\}$

$$Y_{it} = \beta_0 + \beta_1 P_t + \beta_2 T_i + \beta_3 [P_t \cdot T_i] + U_i$$

	Before	After	After – Before
Control			
Treated			
Treat – Control			

Difference-in-differences = compare Y change of units exposed to some policy T with Y change of unexposed

2 periods (before/after) and 2 groups (treated/control)

$Y_{it} \coloneqq$ outcome of interest

$P_t \coloneqq 1\{t \text{ is after treatment occurs}\}$

$T_i \coloneqq 1\{i \text{ is treated/exposed}\}$

$$Y_{it} = \beta_0 + \beta_1 P_t + \beta_2 T_i + \beta_3 [P_t \cdot T_i] + U_i$$

	Before	After	After – Before
Control	β_0	$\beta_0 + \beta_1$	β_1
Treated	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_3$
Treat – Control	β_2	$\beta_2 + \beta_3$	β_3

Difference-in-differences = compare Y change of units exposed to some policy T with Y change of unexposed

2 periods (before/after) and 2 groups (treated/control)

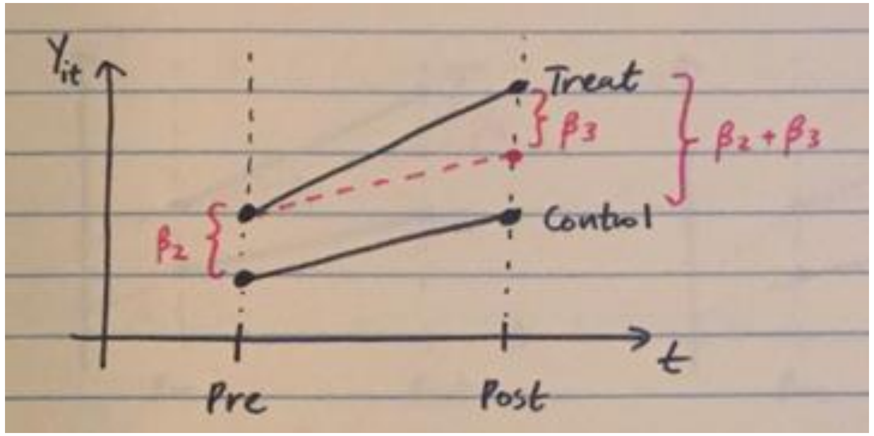
Y_{it} ≔ outcome of interest

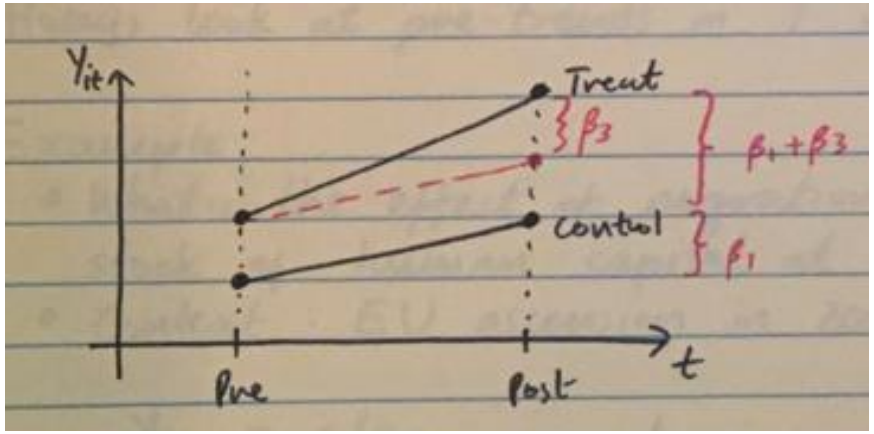
$P_t \coloneqq 1\{t \text{ is after treatment occurs}\}$

$T_i \coloneqq 1\{i \text{ is treated/exposed}\}$

$$Y_{it} = \beta_0 + \beta_1 P_t + \beta_2 T_i + \beta_3 [P_t \cdot T_i] + U_i$$

	Before	After	After – Before
Control	β_0	$\beta_0 + \beta_1$	β_1
Treated	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_1 + \beta_3$
Treat – Control	β_2	$\beta_2 + \beta_3$	β_3



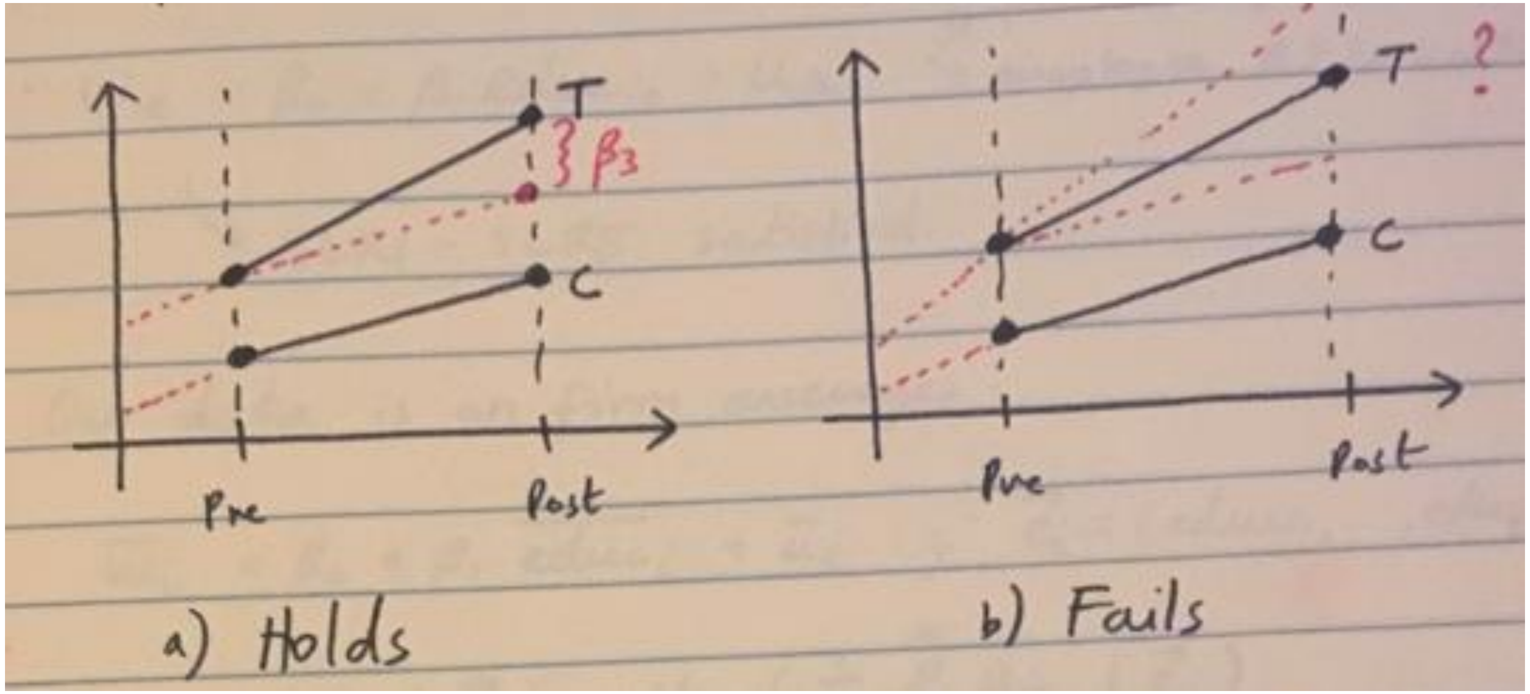


Parallel Trends Assumption = exposed units Y without policy T would have changed like unexposed units Y

- PTA is an untestable assumption, just like OLS exogeneity or IV exogeneity
- However, if we have access to more data before policy, we can assess how likely it is to hold in practice… commonly known as "checking for pre-trends"
- One reason why people seem to like DD… visual check of identifying assumption!

Parallel Trends Assumption = exposed units Y without policy T would have changed like unexposed units Y

- PTA is an untestable assumption, just like OLS exogeneity or IV exogeneity
- However, if we have access to more data before policy, we can assess how likely it is to hold in practice… commonly known as “checking for pre-trends”
- One reason why people seem to like DD… visual check of identifying assumption!



Emigration & Education: Separating Remittance from Wage Premium Effects

Elird Haxhiu*

University of Michigan

October 7, 2022

Abstract

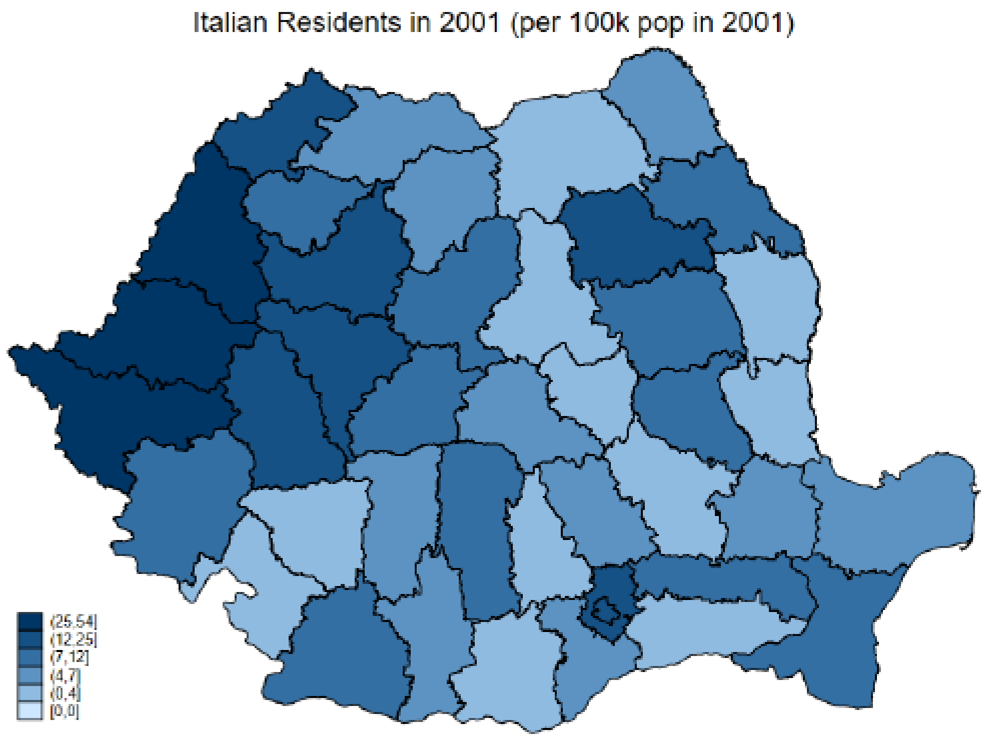
Remittances to developing countries are a large source of income but come at the cost of losing workers to destination countries. Fears of "brain drain" abound when migrants are positively selected, but may be assuaged by "brain gains" at home. These effects are usually motivated by an increased wage premium for skill but can also arise if enough constrained households use remittances to finance education investments. I argue dominance of the premium channel can make short-run gains transitory, while dominance of the remittance channel is necessary for persistent increases in origin skills. To infer their relative contributions to reduced form estimates, I show that remittances are more dominant whenever emigration increases education rates and closes skill gaps between constrained and unconstrained households. I then study Romania since the fall of Communism in 1990, where over 20% of the population (six million people) has emigrated. In 2002, Schengen visa requirements were waived for all Romanians, which generated heterogeneous opportunities for emigration that I capture with a continuous measure of foreign migrant networks. Difference-in-differences estimates show increases in enrollment and graduation rates, but no resulting increase in the stock of educated. Urban-rural skill gaps do not shrink in response to the shock, which implies changing perceptions of the skill premium generated most of the short-run education gains. These subsequently disappeared with Romania's continued European integration and higher skilled emigration rates.

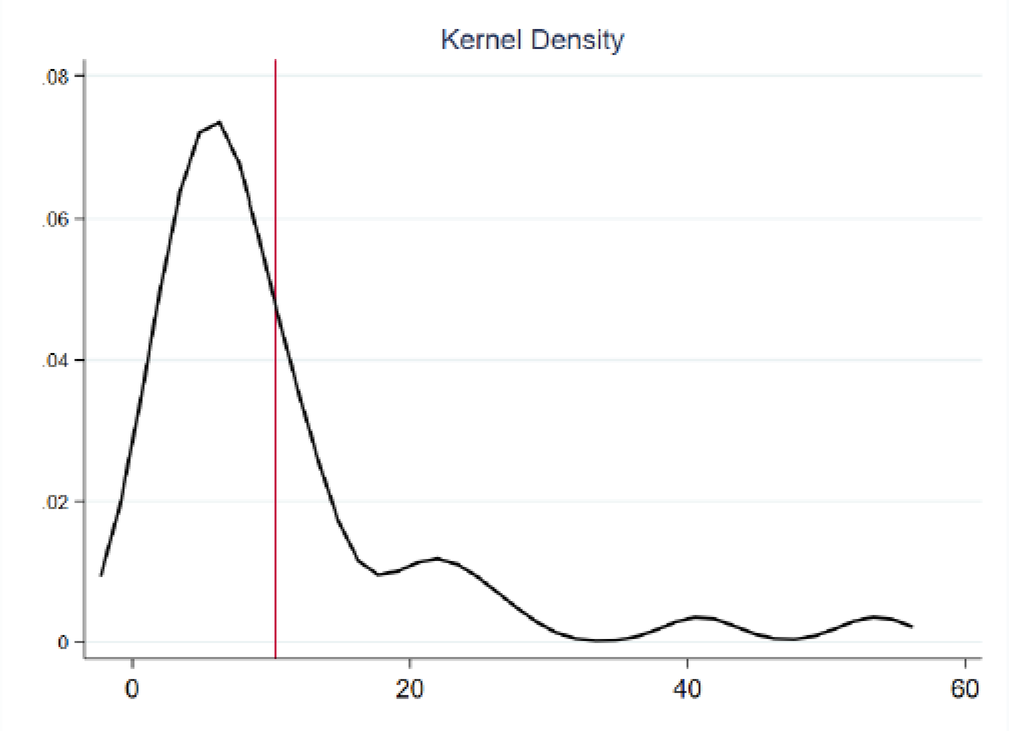
JEL codes: F22, I25, O15.

Key words: Remittances, Migration, Human Capital, Networks, Credit Constraints.

Romanian emigration following visa-free travel shock

Measuring Exposure to Emigration Opportunity in 2002





Romanian emigration following visa-free travel shock

Continuous Treatment Difference-in-Differences

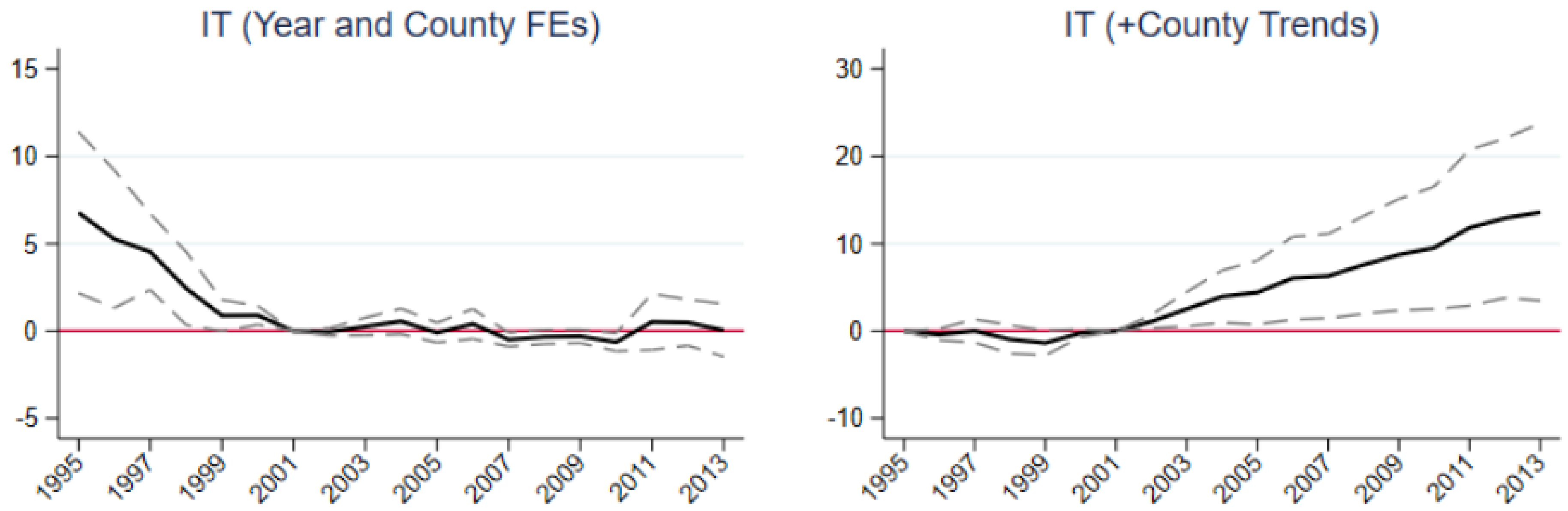
Unit of analysis	=	Romanian county (42 total)
Outcome Y_{ct}	=	EMIG (permanent) or EDUC (tertiary flows)
Exposure Z_c	=	Italians per 100k pop (>0 for all units)
Event date	=	2002

$$
\begin{aligned}
Y_{ct} &= \alpha_c + \theta_t + \sum_{\tau \in [95,17] \setminus \{01\}} \beta_\tau \cdot Z_c 1\{t = \tau\} + \varepsilon_{ct} \\
Y_{ct} &= \alpha_c + \theta_t + \delta_c t + \sum_{\tau \in [95,17] \setminus \{01,95\}} \beta_\tau \cdot Z_c 1\{t = \tau\} + \varepsilon_{ct}
\end{aligned}
$$

Romanian emigration following visa-free travel shock

Reduced form: permanent emigration rate

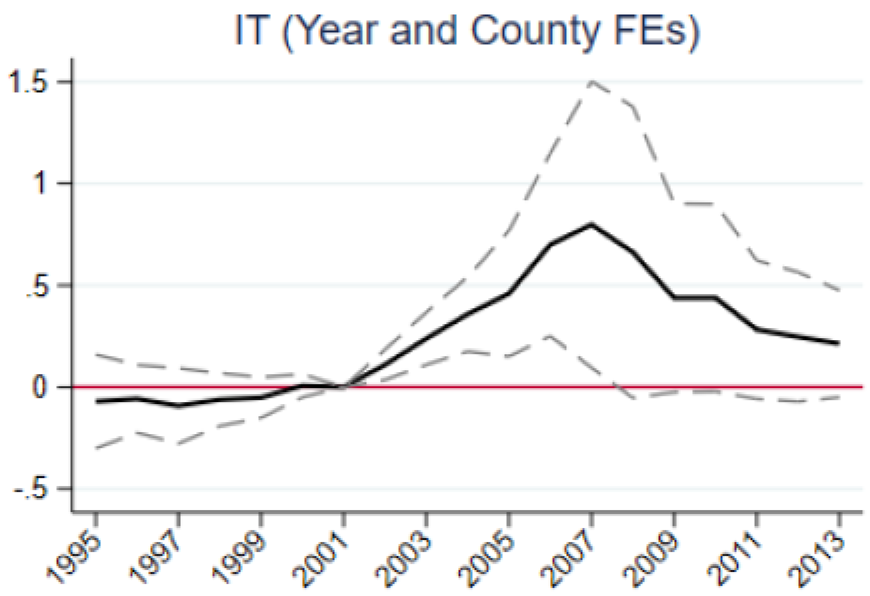
$\hat{\beta}_\tau \cdot \hat{\sigma}_Z \approx 10 \cdot 10 = 100$ more emigrants on average, given $+1$SD increase in exposure Z_c

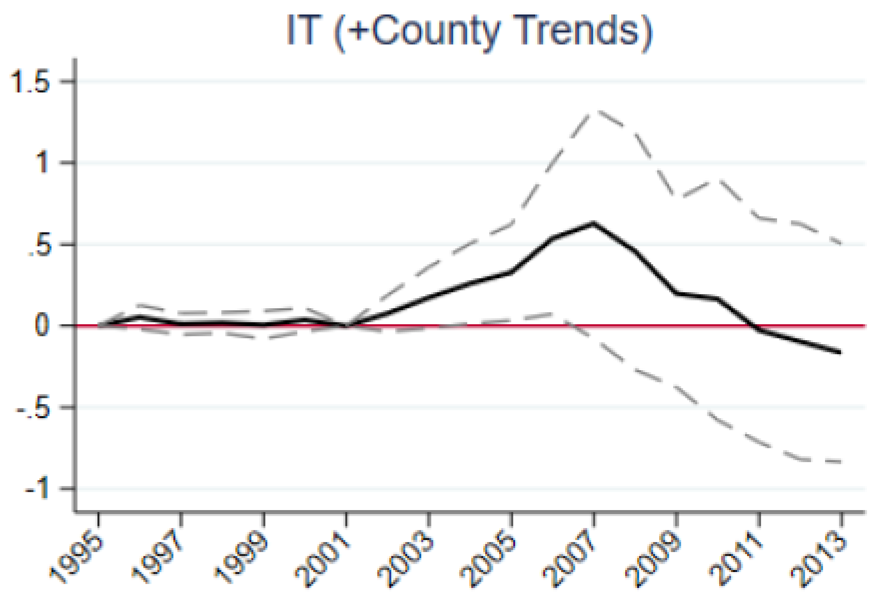


Romanian emigration following visa-free travel shock

University enrollment rate

$\hat{\beta}_{\tau} \cdot \hat{\sigma}_Z \approx 0.5 \cdot 10 = 5$pp higher enrollment

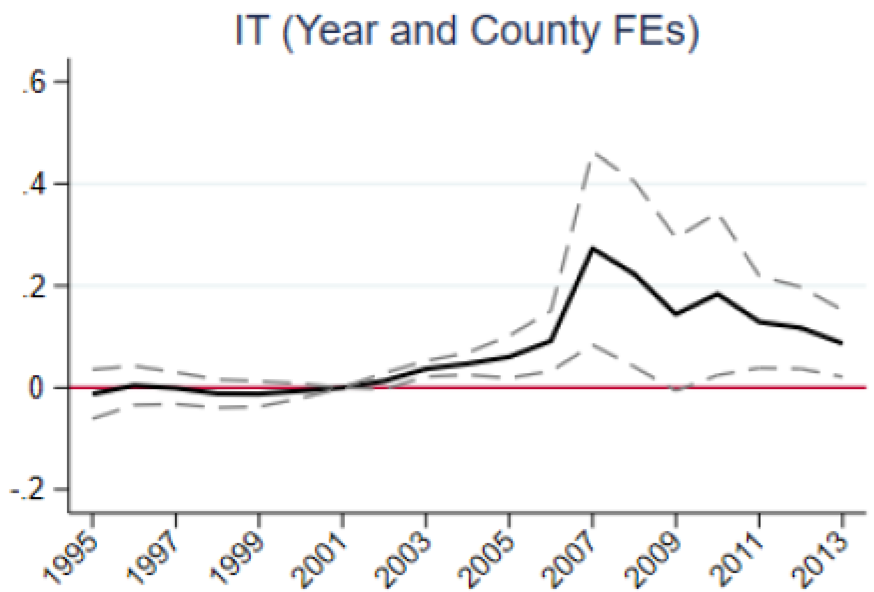


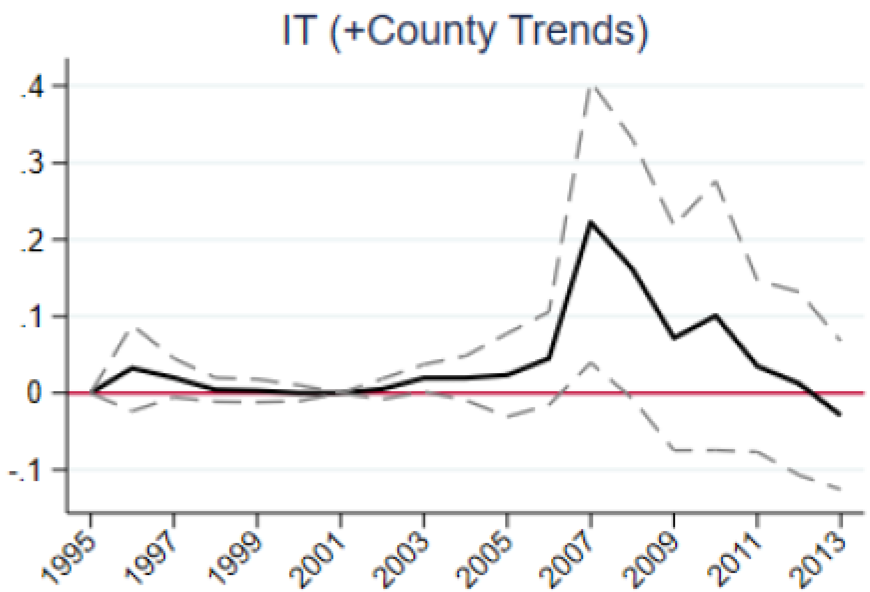


Romanian emigration following visa-free travel shock

University graduation rate

$\hat{\beta}_\tau \cdot \hat{\sigma}_Z \approx 0.2 \cdot 10 = 2$pp higher graduation





Romanian emigration following visa-free travel shock

Table 2: Romanian Census Results. Source: IPUMS Census data.

	$1\{\text{in school}\}$	$1\{\text{college grad}\}$	$1\{\text{in school}\}$	$1\{\text{college grad}\}$
$Z_c \times 1\{t = 2011\}$	0.002***	0.002**	0.003***	0.003***
	(0.0008)	(0.0008)	(0.001)	(0.009)
$Z_c \times 1\{t = 2002\}$	0.002**	0.0005***	0.004***	0.007**
	(0.0009)	(0.0002)	(0.009)	(0.0003)
$Z_c \times 1\{t = 1977\}$	-0.001	-0.001***	-0.0008	-0.002***
	(0.0006)	(0.008)	(0.007)	(0.0006)
$1\{\text{rural}\}$			-0.13***	-0.08***
			(0.01)	(0.005)
$1\{\text{rural}\} \times Z_c \times 1\{t = 2011\}$			-0.0035***	-0.0038***
			(0.001)	(0.01)
$1\{\text{rural}\} \times Z_c \times 1\{t = 2002\}$			-0.006***	-0.0007
			(0.0016)	(0.0004)
$1\{\text{rural}\} \times Z_c \times 1\{t = 1977\}$			0.006	0.003***
			(0.0007)	(0.0008)
Other Variables				
1. Region Fixed Effects	✓	✓	✓	✓
2. Year Effects	✓	✓	✓	✓
4. "Stayers" only	✓	✓	✓	✓
3. Controls	✓	✓	✓	✓
Statistics				
Age group	18-24	≥ 25	18-24	≥ 25
Observations	849,583	5,349,339	849,583	5,349,339
Adjusted-R^2	0.07	0.27	0.39	0.26

Standard errors in parentheses

Standard errors are clustered at the county level

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Outline

1. Review instrumental variables (IV)
2. Review panel methods + binary treatment diff-in-diff (DD)
3. Continuous treatment DD (Haxhiu & Helgerman, 2022)
4. Extra theory + Stata examples to sharpen tools

Low Dose or No Dose? Continuous Treatment Difference-in-Differences with Unknown Controls

Elird Haxhiu * Thomas Helgerman *

haxhiu@umich.edu tehelg@umich.edu

October 6, 2022

Abstract

This paper studies difference-in-differences research designs where all units receive a continuous treatment, or dose, so there is no group that is ex ante unexposed. We present a framework to identify and estimate average treatment effect and causal response parameters when the continuous treatment takes effect only after some cutoff value. In applied settings, this parameter is usually unknown and hence neglected from econometric analysis. Under a range of data-generating processes, we illustrate the bias from Two-Way Fixed-Effects (TWFE) estimators when treatment is defined as (i) the full dose or (ii) an indicator for units with doses above some researcher-specified value or percentile, such as the median. For large jumps or sharp discontinuities at the cutoff value, researchers should instead jointly estimate the threshold along with treatment effect parameters using existing methods. This restores identification and produces correct standard errors but fails when parametric assumptions do not hold or the dose response function is flat around the true cutoff. In these cases, we argue that researchers should instead target binned average treatment effects and document an intuitive bias-variance trade-off in recategorizing low dose units as controls in estimation. We then exploit this trade-off to derive the MSE-optimal estimator, show that it depends on the unknown cutoff, and propose a minimax constraint and partial identification procedure to make progress on inference.

JEL codes: C14, C23, C24.

Key words: Difference-in-Differences, Parallel Trends, Threshold Estimation, Dose Response curves

Low Dose or No Dose? Continuous Treatment Difference-in-Differences with Unknown Controls

Elird Haxhiu * Thomas Helgerman *

haxhiu@umich.edu tehelg@umich.edu

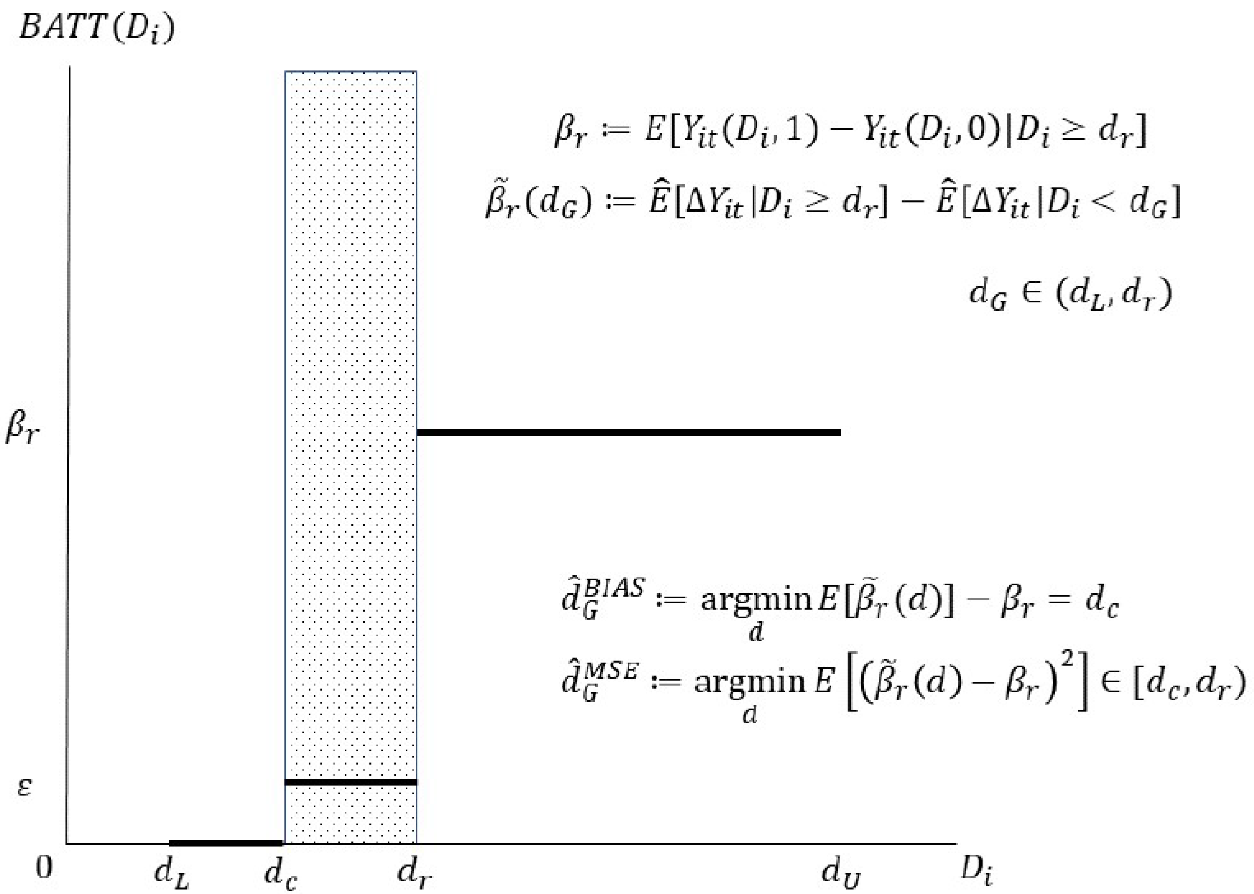
October 6, 2022

Abstract

This paper studies difference-in-differences research designs where all units receive a continuous treatment, or dose, so there is no group that is ex ante unexposed. We present a framework to identify and estimate average treatment effect and causal response parameters when the continuous treatment takes effect only after some cutoff value. In applied settings, this parameter is usually unknown and hence neglected from econometric analysis. Under a range of data-generating processes, we illustrate the bias from Two-Way Fixed-Effects (TWFE) estimators when treatment is defined as (i) the full dose or (ii) an indicator for units with doses above some researcher-specified value or percentile, such as the median. For large jumps or sharp discontinuities at the cutoff value, researchers should instead jointly estimate the threshold along with treatment effect parameters using existing methods. This restores identification and produces correct standard errors but fails when parametric assumptions do not hold or the dose response function is flat around the true cutoff. In these cases, we argue that researchers should instead target binned average treatment effects and document an intuitive bias-variance trade-off in recategorizing low dose units as controls in estimation. We then exploit this trade-off to derive the MSE-optimal estimator, show that it depends on the unknown cutoff, and propose a minimax constraint and partial identification procedure to make progress on inference.

JEL codes: C14, C23, C24.

Key words: Difference-in-Differences, Parallel Trends, Threshold Estimation, Dose Response curves



Outline

1. Review instrumental variables (IV)
2. Review panel methods + binary treatment diff-in-diff (DD)
3. Continuous treatment DD (Haxhiu & Helgerman, 2022)
4. Extra theory + Stata examples to sharpen tools

1. Prove that the sample mean is the OLS estimator when no covariates are specified.

2. Derive the OLS estimator assuming no intercept and discuss the bias-variance trade-off as an example of a very general principle in statistics and econometrics.

3. Prove that the OLS estimator is equal to the difference of means when the treatment variable is binary using its equivalence to method of moments (MM).