Williams College ECON 523:

Program Evaluation for International Development

Lecture 5: Two-Way Fixed Effects

Professor: Pamela Jakiela

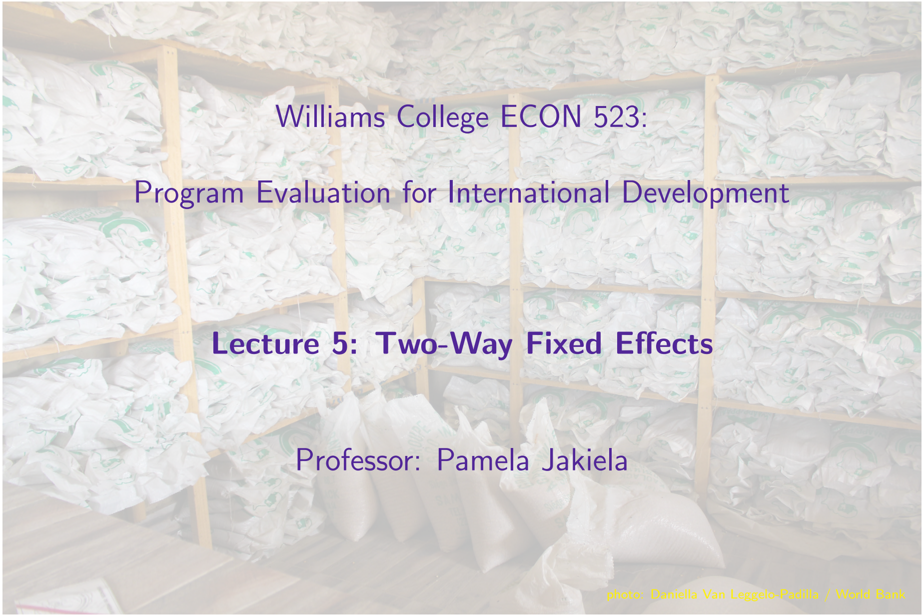


photo: Daniella Van Leggelo-Padilla / World Bank

Variation in Treatment Timing

Example: States Adopted Medicaid at Different Times

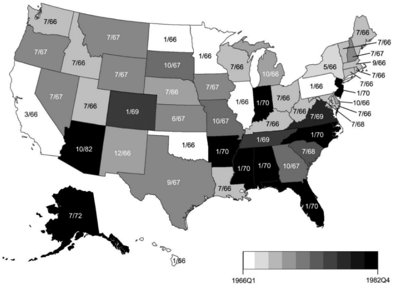


Figure 2.
Medicaid Adoption by Quarter
Notes: Adoption dates come from the Department of Health Education and Welfare (1970) & Social Security Administration (2013). The map is shaded relative to the quarter of adoption and states are labeled with the month and year of adoption.

source: Boudreaux, Golberstein, and McAlpine (Journal of Health Economics, 2016)

Example: Counties Opening Community Health Centers

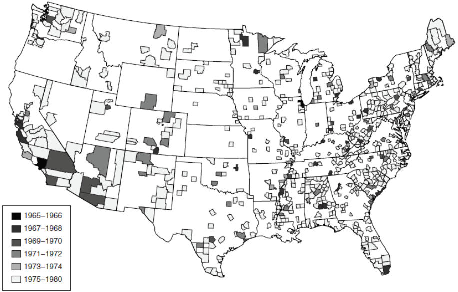


Figure 3. Establishment of Community Health Centers by County of Service Delivery, 1965–1980

Note: Dates are the first year that a CHC was established in the county.
Source: Information on CHCs drawn from NACAP and PHS reports.

source: Bailey and Goodman-Bacon (AER, 2015)

Example: African Countries Introducing Multi-Party Elections

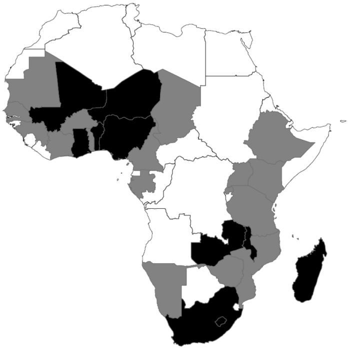


FIGURE A.2. Geographical distribution of democratized countries since 1990. Black-colored countries are democratized since 1990; grey-colored countries the other countries in the sample for infant mortality analysis. Democratized countries include the Comoros, tiny islands to the northwest of Madagascar, which may not be visible as black-colored.

source: Kudamatsu (JEEA, 2012)

Two-Way Fixed Effects Estimates of β^{DD}

What exactly is β^{DD}?

$$Y_{it} = \alpha_i + \gamma_t + \beta^{DD} D_{it} + \varepsilon_{it}$$

unit fixed effects

time fixed effects

treatment dummy

that turns "on" at different times

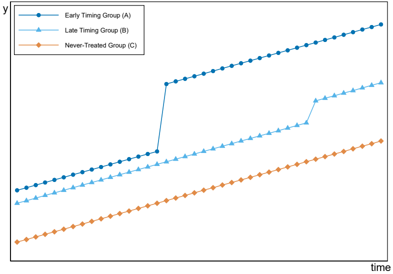
What exactly is β^{DD} in TWFE?

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	
ID 1	0	0	1	1	1	treatment
ID 2	0	0	1	1	1	
ID 3	0	0	0	0	0	comparison
ID 4	0	0	0	0	0	

Multiple Treatment and Comparison Groups

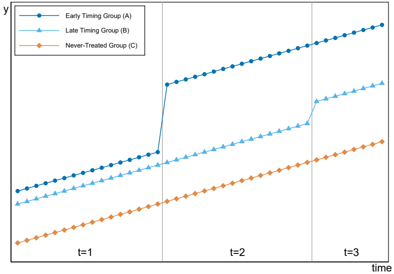
	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$	
ID 1	0	0	1	1	1	1	1	1	timing group 1 (early)
ID 2	0	0	1	1	1	1	1	1	
ID 3	0	0	0	0	0	1	1	1	timing group 2 (late)
ID 4	0	0	0	0	0	1	1	1	
ID 5	0	0	0	0	0	0	0	0	never treated
ID 6	0	0	0	0	0	0	0	0	

Decomposition into Timing Groups



Panel with variation in treatment timing can be decomposed into distinct **timing groups** reflecting observed start of treatment

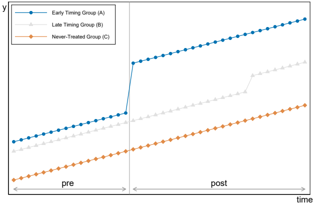
Decomposition into Timing Groups



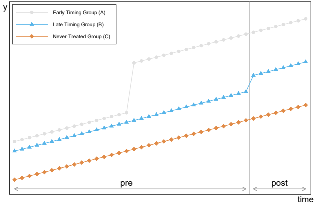
Example: with three timing groups (one of which is never treated), can construct three timing windows (pre, middle, post or $t = 1, 2, 3$)

Decomposition into Standard 2 × 2 DDs

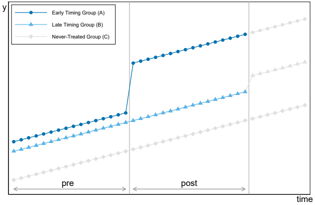
Group A vs. Group C



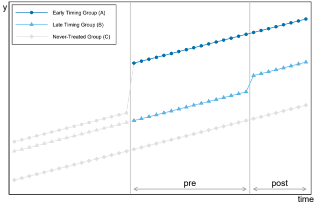
Group B vs. Group C



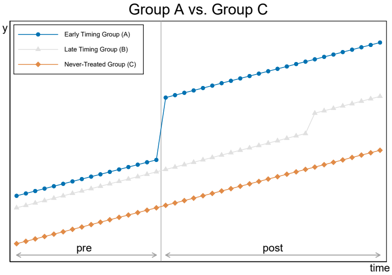
Group A vs. Group B



Group B vs. Group A



Decomposition into Standard 2 × 2 DDs



We know DD estimate of treatment effect for each timing group:

$$\hat{\beta}^{DD}_{AC} = \left(\bar{Y}^{POST}_{A} - \bar{Y}^{POST}_{C}\right) - \left(\bar{Y}^{PRE}_{A} - \bar{Y}^{PRE}_{C}\right)$$
$$= \left(\bar{Y}^{t=2,3}_{A} - \bar{Y}^{t=2,3}_{C}\right) - \left(\bar{Y}^{t=1}_{A} - \bar{Y}y^{t=1}_{C}\right)$$

Bacon Decomposition

Theorem

Consider a data set comprising K timing groups ordered by the time at which they first receive treatment and a maximum of one never-treated group, U. The OLS estimate from a two-way fixed effects regression is:

$$\hat{\beta}^{DD} = \sum_{k \neq U} s_{kU} \hat{\beta}^{DD}_{kU} + \sum_{k \neq U} \sum_{j > k} \left[s_{kj} \hat{\beta}^{DD}_{kj} + s_{jk} \hat{\beta}^{DD}_{jk} \right]$$

The two-way fixed effects estimator β^{DD} is a weighted average of 2×2 diff-in-diff estimators across all possible pairwise combinations of timing groups (Goodman-Bacon 2021)

Bacon Decomposition: Calculating the Weights

Weights depend on sample size, variance of treatment w/in each DD:

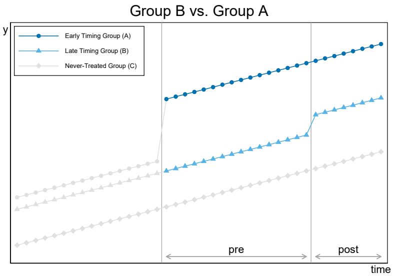
$$s_{kU} = \left[\frac{(n_k + n_U)^2}{\hat{V}^{\tilde{D}}}\right] \underbrace{n_{kU}\,(1 - n_{kU})\,\bar{D}_k(1 - \bar{D}_k)}_{\hat{Var}^{\tilde{D}}_{kU}}$$

$$s_{kj} = \left[\frac{\left((n_k + n_j)\,(1 - \bar{D}_j)\right)^2}{\hat{V}^{\tilde{D}}}\right] \underbrace{n_{kj}(1 - n_{kj}) \left(\frac{\bar{D}_k - \bar{D}_j}{1 - \bar{D}_j}\right) \left(\frac{1 - \bar{D}_k}{1 - \bar{D}_j}\right)}_{\hat{Var}^{\tilde{D}}_{kj}}$$

$$s_{jk} = \left[\frac{\left((n_k + n_j)\,\bar{D}_k\right)^2}{\hat{V}^{\tilde{D}}}\right] \underbrace{n_{kj}(1 - n_{kj})\frac{\bar{D}_j}{\bar{D}_k} \left(\frac{\bar{D}_k - \bar{D}_j}{\bar{D}_k}\right)}_{\hat{Var}^{\tilde{D}}_{jk}}$$

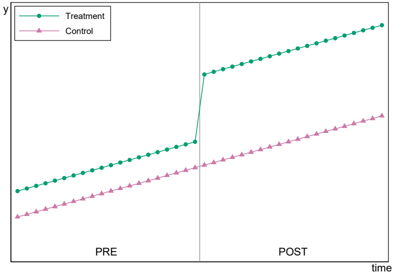
where n_k is the proportion of the sample in group timing group k for all k timing groups, $n_{kj} = n_k/(n_k + n_j)$, and $\bar{D}_k$ is the fraction of sample periods in which k is treated

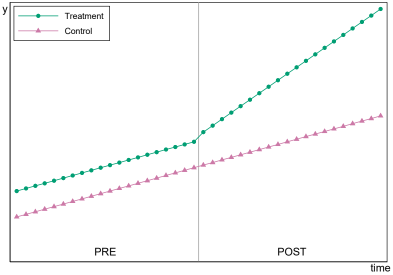
Forbidden Comparisons



Some 2×2 diff-in-diffs (implicitly) use already-treated groups as the comparison

Why Are Forbidden Comparisons Bad?





If treatment changes the level of Y **and the rate of change in** Y, already-treated units cannot be used as a comparison group (common trends does not hold)

$\rightarrow$ This problem does not arise (in the same way) in 2×2 diff-in-diff

Two-Way Fixed Effects β^{DD} as a Weighted Sum

The two-way fixed effects estimator β^{DD} is a weighted sum of 2×2 diff-in-diff estimators across all possible pairwise combinations of timing groups (Goodman-Bacon 2021)

- Some use an already-treated group as comparison
  - Creates problems if treatment effect grows/changes over time
  - TWFE imposes a model of homogeneous treatment effects
  - When treatment effects evolve over time, model is mis-specified

We can use Frisch-Waugh-Lovell to construct the TWFE/OLS weights used to generate β^{DD}

- Weights on treated units are not always positive (they are also used as comparison)

Two-Way Fixed Effects as Univariate Regression

Two-way fixed effects is equivalent to univariate regression:

$$\tilde{Y}_{it} = \alpha + \tilde{D}_{it} + \epsilon_{it}$$

where $\tilde{Y}_{it} = Y_{it} - \bar{Y}_t - (\bar{Y}_i - \bar{\bar{Y}})$ and $\tilde{D}_{it}$ defined analogously

"grand mean"

(just the mean across i and t)

Two-Way Fixed Effects as Univariate Regression

Two-way fixed effects is equivalent to univariate regression:

$$\tilde{Y}_{it} = \alpha + \tilde{D}_{it} + \epsilon_{it}$$

where $\tilde{Y}_{it} = Y_{it} - \bar{Y}_t - (\bar{Y}_i - \bar{\bar{Y}})$ and $\tilde{D}_{it}$ defined analogously

$\Rightarrow$ Treatment dummy now effectively continuous measure $\tilde{D}_{it}$

$$\hat{\beta}^{OLS} = \sum_{it} \tilde{Y}_{it} \underbrace{\left(\frac{\tilde{D}_{it} - \bar{\tilde{D}}_{it}}{\sum_i \left(\tilde{D}_{it} - \bar{\tilde{D}}_{it} \right)^2} \right)}_{\text{OLS weight}}$$

Two-Way Fixed Effects as Univariate Regression

Two-way fixed effects is equivalent to univariate regression:

$$\tilde{Y}_{it} = \alpha + \tilde{D}_{it} + \epsilon_{it}$$

where $\tilde{Y}_{it} = Y_{it} - \bar{Y}_t - \left(\bar{Y}_i - \bar{\bar{Y}} \right)$ and $\tilde{D}_{it}$ defined analogously

$\Rightarrow$ Treatment dummy now effectively continuous measure $\tilde{D}_{it}$

$$\hat{\beta}^{OLS} = \sum_{it} \tilde{Y}_{it} \underbrace{\left(\frac{\color{red}{\tilde{D}_{it}} - \bar{\tilde{D}}_{it}}{\sum_i \left(\tilde{D}_{it} - \bar{\tilde{D}}_{it} \right)^2} \right)}_{\text{OLS weight}} \quad \text{where } \bar{\tilde{D}}_{it} = 0$$

Diff-in-Diff without Staggered Treatment Timing: Review

	$t = 1$	$t = 2$	$t = 3$	$t = 4$
ID 1	0	1	1	1
ID 2	0	0	0	0

Diff-in-Diff without Staggered Treatment Timing: Review

$$D_{it} - \bar{D}_t - \left(\bar{D}_i - \bar{\bar{D}}\right)$$

	$t = 1$	$t = 2$	$t = 3$	$t = 4$
ID 1	-0.375	0.125	0.125	0.125
ID 2	0.375	-0.125	-0.125	-0.125

← Equal weight

Diff-in-Diff without Staggered Treatment Timing: Review

$$w_{it} = \tilde{D}_{it} / \left(\sum_{it} \tilde{D}_{it}^2 \right)$$

	$t = 1$	$t = 2$	$t = 3$	$t = 4$
ID 1	-1	$0.\bar{3}$	$0.\bar{3}$	$0.\bar{3}$
ID 2	1	$-0.\bar{3}$	$-0.\bar{3}$	$-0.\bar{3}$

← Equal weight

$$\hat{\beta}_{ols} = \sum_i Y_i w_i = \sum_i Y_i \frac{\tilde{D}_{it}}{\sum_{it} \tilde{D}_{it}^2}$$

Diff-in-Diff without Staggered Treatment Timing: Review

$$w_{it} = \tilde{D}_{it} / \left(\sum_{it} \tilde{D}_{it}^2 \right)$$

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	
ID 1	-1	$0.\bar{3}$	$0.\bar{3}$	$0.\bar{3}$	⟵ Equal weight
ID 2	1	$-0.\bar{3}$	$-0.\bar{3}$	$-0.\bar{3}$	⟵ Equal weight

$$\hat{\beta}_{ols} = \sum_i Y_i w_i$$

$$= \sum_{ET,pre} Y_i w_i + \sum_{ET,post} Y_i w_i + \sum_{NT,pre} Y_i w_i + \sum_{NT,post} Y_i w_i$$

Diff-in-Diff without Staggered Treatment Timing: Review

$$w_{it} = \tilde{D}_{it} / \left(\sum_{it} \tilde{D}_{it}^2 \right)$$

	$t=1$	$t=2$	$t=3$	$t=4$	
ID 1	-1	$0.\bar{3}$	$0.\bar{3}$	$0.\bar{3}$	$\longleftarrow$ Equal weight
ID 2	1	$-0.\bar{3}$	$-0.\bar{3}$	$-0.\bar{3}$	$\longleftarrow$ Equal weight

$$\begin{aligned} \hat{\beta}_{ols} &= \sum_i Y_i w_i \\ &= \sum_{ET,pre} Y_i w_i + \sum_{ET,post} Y_i w_i + \sum_{NT,pre} Y_i w_i + \sum_{NT,post} Y_i w_i \\ &= \ldots \\ &= \bar{Y}_{ET,pre}(-1) + 3\bar{Y}_{ET,post}(0.\bar{3}) + \bar{Y}_{NT,pre}(1) + 3\bar{Y}_{NT,post}(-0.\bar{3}) \end{aligned}$$

Diff-in-Diff with Staggered Treatment Timing: Example

	$t = 1$	$t = 2$	$t = 3$	$t = 4$
ID 1	0	1	1	1
ID 2	0	0	0	1
$\bar{D}_t$	0	0.5	0.5	1

mean treatment in period t

Diff-in-Diff with Staggered Treatment Timing: Example

	Y_{it}			
	$t = 1$	$t = 2$	$t = 3$	$t = 4$
ID 1	0	10	10	10
ID 2	0	0	0	10

Let $Y_{it} = \gamma_i + \lambda_t + \delta_{it}$

Treated cells

Positive weights (in treatment group)

Diff-in-Diff with Staggered Treatment Timing: Example

Y_{it} $\longrightarrow$ $\hat{\beta}_{OLS} = 10$

	$t=1$	$t=2$	$t=3$	$t=4$
ID 1	0	10	10	10
ID 2	0	0	0	10

homogeneous impacts:
$E[\hat{\beta}_{OLS}] = \text{ATE}$

Treated cells

Positive weights (in treatment group)

Diff-in-Diff with Staggered Treatment Timing: Example

Y_{it} ⟶ $\hat{\beta}_{OLS} = 6$

?

	$t = 1$	$t = 2$	$t = 3$	$t = 4$
ID 1	0	2	2	2
ID 2	0	0	0	10

Treated cells

Positive weights (in treatment group)

Diff-in-Diff with Staggered Treatment Timing: Example

	Y_{it}			
	$t=1$	$t=2$	$t=3$	$t=4$
ID 1	0	2	2	10
ID 2	0	0	0	2

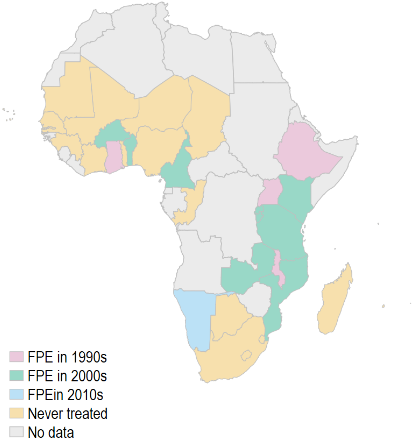
$Y_{it} \longrightarrow \hat{\beta}_{OLS} = -2$

Treated cells

Positive weights (in treatment group)

TWFE in Practice

Policy Context: Free Primary Education in Sub-Saharan Africa



Country	FPE Year
Benin	2006
Burkina Faso	2007
Burundi	2005
Cameroon	2000
Ethiopia	1995
Ghana	1996
Kenya	2003
Lesotho	2006
Malawi	1994
Mozambique	2005
Namibia	2013
Rwanda	2003
Tanzania	2001
Uganda	1997
Zambia	2002

TWFE Specification: Countries that Implemented Free Primary

TWFE specification: $Primary_{it} = \alpha_i + \gamma_t + \beta FPE_{it} + \varepsilon_{it}$

Stata code: `reg primary fpe i.cid i.year, cluster(cid)`

	(1)
	Enrollment
Free primary education	20.428
	(9.120)
	[0.042]
Country fixed effects	Yes
Year fixed effects	Yes
Never treated	No

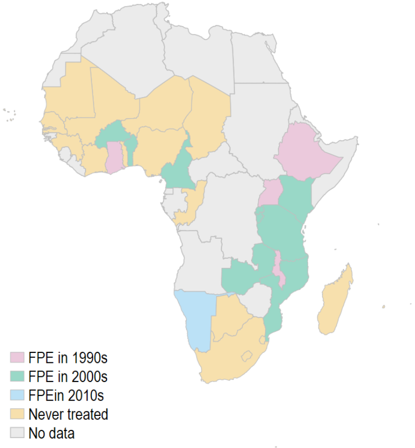
Dependent variable: gross enrollment ratio. Data on gross enrollment ratio in 15 countries comes from the World Development Indicators, years 1981 through 2015. Standard errors (clustered at the country level) in parentheses; p-values in square brackets.

TWFE Diagnostics

1. Are treated observations getting negative weight in my TWFE estimation?
   - Are treated observations (i.e. country-years) being weighted in a sensible way?
2. Are treatment effects (likely to be) heterogeneous? If yes, how?
   - Conceptually: do you expect the treatment effects to vary over time, across units, or both?
   - Do you see evidence contradicting the assumption of homogeneous treatment effects?
     - Event study specifications
     - Scatter plots of residuals
   - Are your estimated treatment effects robust across specifications?

Negative Weights

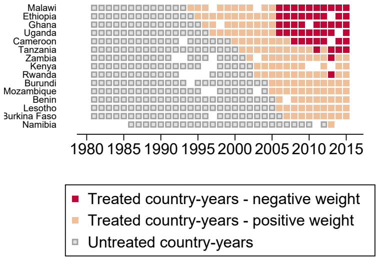
Including Never Treated Countries to Eliminate Negative Weights

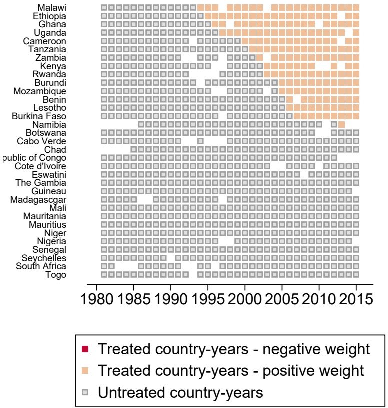


	(1) Enrollment	(2) Enrollment
Free primary education	20.428	18.951
	(9.120)	(6.733)
	[0.042]	[0.008]
Country fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Never treated	No	Yes

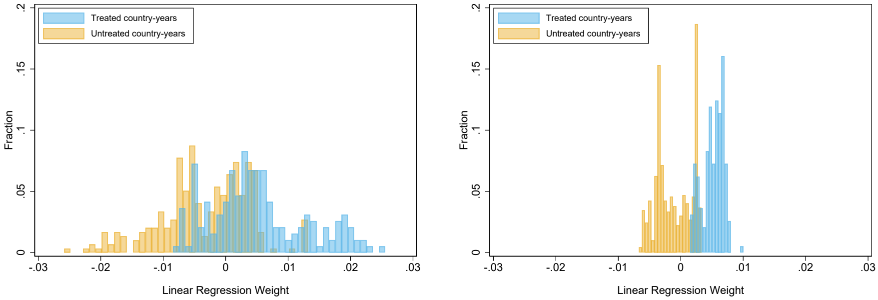
Dependent variable: gross enrollment ratio. Data on gross enrollment ratio in 33 countries comes from the World Development Indicators, years 1981 through 2015. Standard errors (clustered at the country level) in parentheses; p-values in square brackets.

Including Never Treated Countries to Eliminate Negative Weights

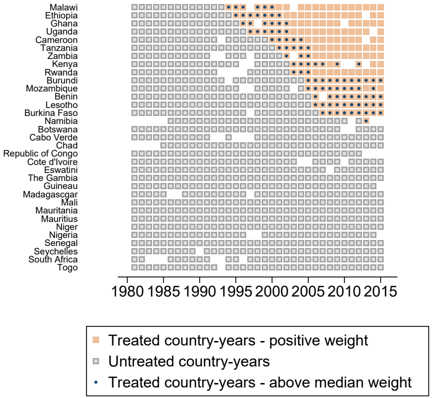




Including Never Treated Countries to Eliminate Negative Weights



Eliminating Negative Weights $\neq$ Weighting Country-Years Equally



Event Study Specifications

Negative weights are a major issue if treatment effects change over (relative) time

- Relative time is the number of years since treatment was implemented (in country t)
- We can also think of negative relative time as years until treatment starts (in country t)

An **event study** specification allows us to estimate treatment effects for every (relative) time

- Provides direct evidence on the stability of the treatment effect (over time)
- Also allows us to check for violations of common (pre)trends
- Because we are estimating many parameters instead of one, statistical power is an issue

Event Study Specifications

Let G_i indicate the time t when treatment starts in country i

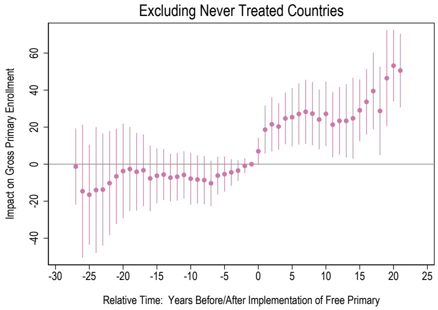
⇒ $R_{it} = t - G_i$ is relative time, and treatment starts when $R_{it} = 0$

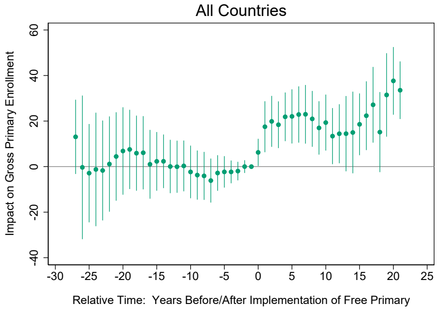
TWFE event study specification:

$$Primary_{it} = \alpha_i + \gamma_t + \sum_{r \leq 2} \beta_r 1\left[R_{it} = r\right] + \sum_{r \geq 0} \delta_r 1\left[R_{it} = r\right] + \varepsilon_{it}$$

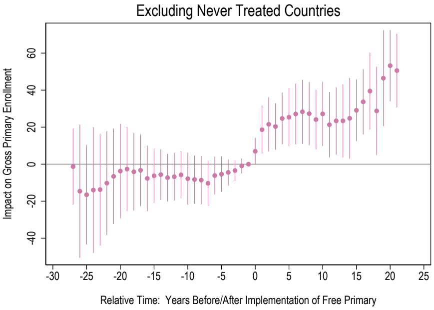
Impacts are defined relative to $R_{it} = -1$, the last period before treatment

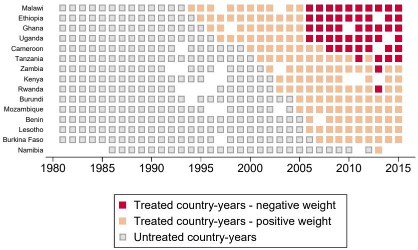
The Impacts of Free Primary: Event Study Specifications



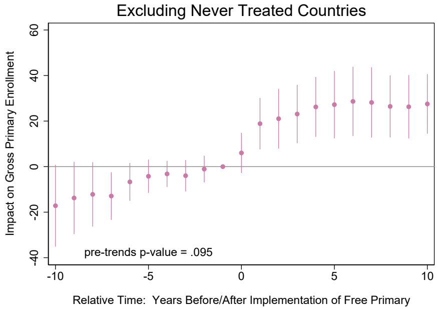


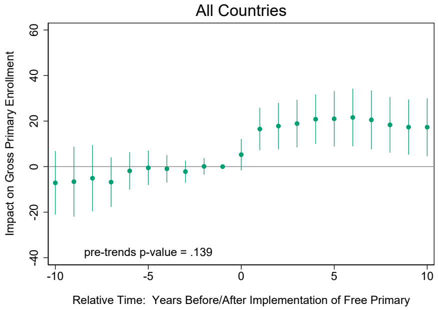
Not All Relative Times Are Observed for All Countries



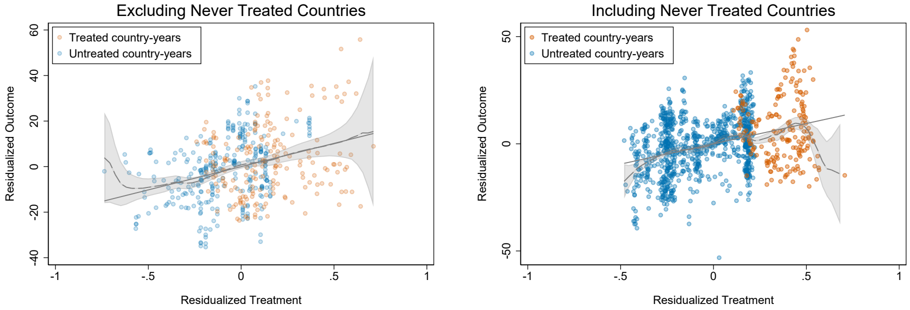


Balancing Event Time

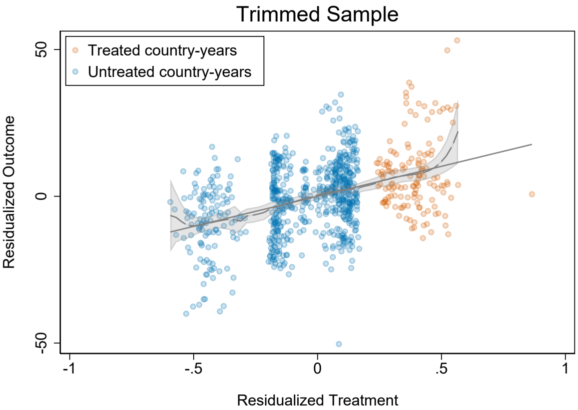




Plotting the Residuals



Trimming the Sample



TWFE Robustness Checks

Summarizing these and other approaches to assessing the robustness of TWFE estimates:

- Show results including additional never-treated units, when possible
- Trim sample so that it is balanced in event time
- Jackknife estimation: drop one country at a time to assess other types of heterogeneity
- Include country-specific trends as controls
- Imputation-based estimation: `did_imputation` in Stata
  - Two-step procedure: estimates country, year fixed effects using untreated country-years

With homogeneous treatment effects, these approaches should generate similar results

TWFE: Checklist

- Check for negative weights, and consider eliminating them
  - The most important thing is to know what you are estimating
- Assess the linearity of the residuals: is homogeneity a reasonable assumption?
- Implement an event study design, if feasible given sample/power
- Robustness checks, more robustness checks, and even more robustness checks