“Difference-in-Differences Estimators with Repeated Cross-Sections, Staggered Timing, and Heterogeneous Treatment Effects” by Partha Deb, Edward Norton, Jeffrey Wooldridge, and Jeffrey Zabel

Discussion by Irene Botosaru
McMaster University

October 2024

A quick summary

Context. Treatment at group level, staggered, heterogenous effects (group-time).

Contribution. Deb et al. (2024+) extends "extended TWFE" (Wooldridge, 2021+) to repeated cross-sections and time-varying covariates

Why. Expand range of applicability, maintaining ease of computation

- pooled OLS to estimate group-time ATT's and their standard errors
- aggregation yields *the* ATT, the Delta Method its standard error

Outline of talk

1. Deb et al. (2024+)
2. Discussion: benefits and potential challenges
3. An alternative / “robustness check”

Outline of talk

1. Deb et al. (2024): paper details
2. Discussion
3. An alternative

Deb et. al (2024+): FLEX

FLEX: **F**lexible **L**inear model **E**stimated by OLS with covariates **X**

- extended TWFE to **repeated cross-sections** and **time-varying covariates**
- inherits properties of Wooldridge's extended TWFE ▸ details
  - parallel trends for all time periods, and both never- and not-yet treated
  - simple computation and efficiency, under a linearity assumption with many interaction terms, and Gauss-Markov assumptions
- emphasis on time-varying nature of covariates when RCS
  - work on DiD with RCS, including Abadie (2005), Botosaru and Gutierrez (2018), Sant'Anna and Zhao (2020), Chang (2020), assumes no covariate changes across cross-sections from the same group
- motivated by ease of computation
  - RCS: Callaway and Sant'Anna (2021), **Borusyak et al. (2024)**
  - TV covariates: **Borusyak et al. (2024)**, Gardner et al. (2024+), Caetano et al. (2024+)

FLEX: Set-up

- time periods $t = 1, ..., T$,
- q: earliest implementation period, $q \in \{2, 3, ..., Q\}$, $Q \leq T$
- implemented at group level, groups indexed by
  $g = \underbrace{1, ..., \overline{g},}_{\text{eventually treated } g \in \mathcal{G}} \underbrace{\overline{g} + 1, ..., G}_{\text{untreated } g \in \mathcal{G}^{\infty}}$
- groups can receive treatment simultaneously, and form cohorts based on when they first receive it, $c \in \{q, q + 1, ..., Q\}$
  - $i \in! g$; if $g \in \mathcal{G}$, then $\forall i \in g$ treated; if g known, then c known
- R_g indicator for group membership
- $Y_t(g)$ potential outcome at t if i belonged to group g
- $Y_t(\infty)$ potential outcome at t if i belonged to $\mathcal{G}^{\infty}$
- X_t covariates that can change over time

FLEX: Parameter of interest and assumptions

Parameter of interest: Average TE at t for units in group g

$$ATT_{g,t} = E\left(Y_t(g) - Y_t(\infty) | R_g = 1 \right), t \geq q, g \in \mathcal{G}.$$

Proposed solution relies on the following additional assumptions:

(A1): **SUTVA**,

(A2) **no anticipation**: zero average TE for all eventually treated groups before implementation and for all subpopulations with $X_t = x$:

$$E\left(Y_t(g) - Y_t(\infty) | \{R_g\}_{g=1}^{G}, X_t = x \right) = 0, t \leq c-1, g \in \mathcal{G},$$

(A3) $\{X_t\}$ **evolve exogenously** from the treatment,

(A4) **conditional common trend and functional form** assumption on potential outcomes without the treatment [next]

FLEX: Parameter of interest and assumptions (continued)

Let $\{R_g\}_{g=1}^{G} \equiv \overrightarrow{R}$ (space considerations).

(A4) For all $t = 2, ..., T$, the CEF $E\left(Y_t(\infty)|\overrightarrow{R}, X_t\right)$ is

$$\underbrace{E\left(Y_1(\infty)|\overrightarrow{R}, X_t\right)}_{\text{"average initial outcome"}} + \underbrace{\left[E\left(Y_t(\infty)|\overrightarrow{R}, X_t\right) - E\left(Y_1(\infty)|\overrightarrow{R}, X_t\right)\right]}_{\text{"average trend in control state"}}$$

$$= E\left(Y_1(\infty)|\overrightarrow{R}, X_t\right) + \left[E\left(Y_t(\infty)|X_t\right) - E\left(Y_1(\infty)|X_t\right)\right]$$

$$= \underbrace{\left[X_t\xi + \sum_{g=1}^{G}\beta_g R_g + \sum_{g=1}^{G}(R_g X_t)\gamma_g\right]}_{\text{group-effects, } \times\text{covariates}} + \underbrace{\left[X_t\pi_t + \eta_t\right]}_{\text{time-effects, } \times\text{covariates}}$$

Average trend in control state in every period relative to initial period does not depend on treatment status is the "Conditional Common Trends, Staggered" in Wooldridge (2021+).

FLEX: Parameter of interest and assumptions (continued)

Let $\{R_g\}_{g=1}^{G} \equiv \vec{R}$.

(A4) For all $t = 2, ..., T$,

$$E\left(Y_t(\infty) | \vec{R}, X_t\right) = X_t\xi + \sum_{g=1}^{G} \beta_g R_g + \underbrace{\sum_{g=1}^{G} R_g X_t \gamma_g + X_t \pi_t}_{\text{group-time separable covariate effects}} + \eta_t$$

FLEX: Main result

Under A1-A4, estimating equation below recovers $ATT_{g,t}$ as $\{\tau_{gt}\}_{g\in\mathcal{G},t\geq q}$.
Let $P_{s,t(i)} = 1_{\{t(i)=s\}}$, $s = 2, ..., T$ (indicator for observations in period s):

$$
\begin{aligned}
& E\left(Y_{i,t(i)} \middle| \{R_{i,t(i),g}\}_{g=1}^{G}, \{P_{s,t(i)}\}_{s=2}^{T}, X_{i,t(i)} \right) \\
= \quad & X_{i,t(i)}\xi + \sum_{g=1}^{G} \beta_g R_{i,t(i),g} + \sum_{g=1}^{G} \left(R_{i,t(i),g} X_{i,t(i)}\right) \gamma_g \ \text{(\textbf{group effects}, } \times \text{covars)} \\
& + \sum_{s=2}^{T} \eta_s P_{s,t(i)} + \sum_{s=2}^{T} \left(P_{s,t(i)} X_{i,t(i)}\right) \pi_s \ \text{(\textbf{time effects}, } \times \text{covars)} \\
& + \sum_{s=q}^{Q} \sum_{g=1}^{G} \left(R_{i,t(i),g} P_{s,t(i)}\right) \gamma_{gs} \ \text{(\textbf{group time trend})} \\
& + \sum_{s=q}^{Q} \sum_{g=1}^{G} R_{i,t(i),g} P_{s,t(i)} \left(X_{i,t(i)} - \overline{X}_{g,t(i)}\right) \tau_{gs} \ \text{(treatment indicator } \times \text{covars)}
\end{aligned}
$$

covariates centered at group level when interacted with treatment indicators.

FLEX: Additional results

Paper also shows

- **aggregation** of τ_{gt} across g, t, or both
- **"leads and lags"** (here, only "lags")
- two-step imputation approach leads to same ATT estimate as pooling across all the data
- **two applications** with publicly available data:
  - punitive prenatal substance use policies and women's mental health
  - right-to-work laws and earnings

compared to estimates using Callaway-Sant'Anna (CS) and Cenzig et al.

Outline of talk

1. Deb et al. (2024+)
2. Discussion: benefits and potential challenges
3. An alternative and extension

Discussion: Potential benefits

1. Use of **standard software** for group-time ATTs and their standard errors
2. Incorporate time-varying (exogenous to treatment) covariates in straight-forward way
3. Explicit focus on **RCS and time-varying covariates**
4. **Efficiency** if linearity of CEF and Gauss-Markov assumptions hold
   - see Borusyak et al. (2024), Harmon (2024+)
5. Linearity and separability of covariate effect have **testable implications** if at least three pre-treatment periods and two different groups [next]
   - appears un(der)explored in imputation/regression DiD lit

Discussion: Potential benefits (continued)

(A4): For all $t = 2, ..., T$:

$$E\left(Y_t(\infty) | \{R_g\}_{g=1}^{G}, X_t\right) = X_t\xi + \sum_{g=1}^{G} \beta_g R_g + \sum_{g=1}^{G} (R_g X_t)\,\gamma_g + X_t \pi_t + \eta_t$$

Testable implication. The difference in average outcomes across groups over pre-treatment periods is **proportional** to the difference in covariates over the same pre-treatment periods:

$$\frac{\Delta Y_{g',g''}(t') - \Delta Y_{g',g''}(t'')}{\Delta Y_{g',g''}(t') - \Delta Y_{g',g''}(t''')} = \frac{X_{t'} - X_{t''}}{X_{t'} - X_{t'''}}, \text{for } g' \neq g'' \text{ and } (t', t'', t''') \leq c - 1$$

- Difference in average outcomes across g', g'' at same $t < c - 1$:

$$\Delta Y_{g',g''}(t) \equiv \overline{Y}_{g'}(t) - \overline{Y}_{g''}(t) = (\beta_{g'} - \beta_{g''}) + X_t (\gamma_{g'} - \gamma_{g''}),$$

  then take differences across pre-treatment (t', t'') and (t', t''').

Discussion: Potential challenges

Linearity in CEF.

- makes it difficult to compare to subgroup estimators (such as CS)
  - e.g., how would standard errors compare if we imposed linearity?
  - CS is efficient under serial correlation and short post-treatment horizon, Harmon (2024+) graph
- linearity, RCS, and TV in **Borusyak et al (2024)** (in all time periods)
  - if panel and no covariates: same point estimates as extended TWFE, and if Gauss-Markov, efficient
  - **how about for RCS with/out covariates**?
- **count outcomes**: consider extension to Generalized Linear Model
  - e.g., Poisson regression, see also Wooldridge (2021+)

Discussion: Potential challenges (continued)

Time change in covariates. Need to be careful – exogenous to treatment.

1. Possible thought experiment: **if panel data, would the variables be time-varying**?
   - If no, suggestive evidence that time-variation may be due to the sampling of different individuals across time, as long as treatment does not induce movement of individuals across groups over time.
2. Consider more conversation on choice of **covariates in applications**
   - e.g., effect of punitive prenatal substance laws use on mental health, with race and ethnicity and income as covariates. What do we know about migration patterns related to opioid use? Might the opioid crisis have led individuals to move out of states with punitive laws, to access healthcare and/or avoid criminal prosecution?
3. Consider **formalizing assumption** via the potential covariate framework
   - e.g., Zeldow and Hatfield (2021), Caetano et al. (2024+)

Discussion: Potential challenges (continued)

Many parameters to estimate.

- equality assumptions as in Wooldridge (2021+), Borusyak et al. (2024), e.g. $\tau_{gt} = \tau_g, \forall g \in \mathcal{G}, t \geq q$
  - bias if misspecified, inefficiency if not "enough" restrictions
- **data-driven approach** from high-dimensional lit, e.g., Faletto (2024+)
  - e.g., penalty **shrinks parameters to each other**: τ_{gt} "close" to $\tau_{gt'}$
  - efficiency preserved, oracle property

In applications, $\widehat{ATT} \to 0$ while its SE same or smaller as $\uparrow$ interactions with X_t

- maybe multicollinearity ("spurious") or outcome variation explained by $\times$covariates ("real")
  - **diagnostics**: condition number, compare nested models, k-fold CV
- consider simulations, also informative about power and coverage rates
- consider formalizing asymptotics and showing expression for SE

Outline of talk

1. Deb et al. (2024+)
2. Discussion
3. An alternative or "robustness check"

Forecasted Average Treatment (FAT) as robustness check

All individuals in a treated group are treated $\rightarrow$ universal treatment at group level
CPS data on hourly earnings: nonstationarity, serial correlation, skedasticity
Aggregate to state level, then FAT (Botosaru, Giacomini, and Weidner, 2023+)

What is FAT. For each treated group, regress pre-treatment outcomes on a polynomial in time, forecast counterfactual, then

$$\widehat{FAT}_{t+h} = \frac{1}{\overline{g}} \sum_{g=1}^{\overline{g}} \left(Y_{gt+h} - \widehat{Y}_{gt+h} \right), h \geq 1.$$

Benefits. $Y_{gt}(\infty)$ can be **nonstationary** (unit root), or mean stationary $(\mathrm{AR}(q_g))$, or both, with **group-specific time trends**.

- **no controls needed**, but can be used to get rid of potential common shock;
- effect in last period, once all treated; works with **very short** T

Why. Robustness check against parallel-trends, linearity, serial correlation, potential unit root.

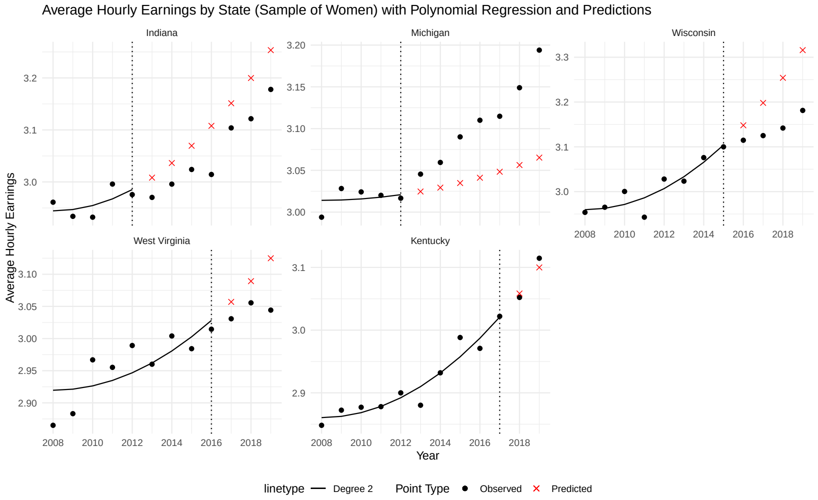
FAT (continued)

Effect of right-to-work laws on hourly earnings.

- CPS data on 5 treated states during 2008-2017 (24 control)
- set of covariates: age, gender, race and ethnicity, education, and marital status
- for illustration of FAT:
  - focus on 5 treated states, on sample of women
- also illustrate DFAT with 24 control groups
  - DFAT = (FAT on treated) - (FAT on control)
  - under a parallel trends assumption, eliminate possible common shock

FAT (continued)

State Average Hourly Earnings for Sample of Women and Predictions



FAT (continued)

Warning: Here, FAT uses state-time data, FLEX uses RCS data

	$\widehat{FLEX}_{\textbf{entire sample}}$
no X	-0.67 (0.11)
with X	-0.37 (0.08)

and

h	$\widehat{FAT}_{\textbf{entire sample}}$	$\widehat{DFAT}_{\textbf{entire sample}}$
1	-0.11 (0.25)	-0.28 (0.28)
2	-0.74 (0.21)	-0.66 (0.25)
3	-1.1 (0.37)	-1.28 (0.4)

	$\widehat{FAT}_{\textbf{sample of women}}$		
h	all	ages $[25, 45]$	ages $(45, 64]$
1	0.14 (0.32)	-0.37 (0.36)	0.79 (0.69)
2	-1.03 (0.33)	-0.84 (0.56)	-1.12 (0.45)
3	-1.15 (0.37)	-0.87 (0.35)	-1.28 (0.58)

- FAT relies on large N (here $\overline{g}$) for asymptotic normality
  - DFAT computed as if never treated were treated in 2012

Conclusion

- FLEX **extends applicability** of extended TWFE to time varying covariates and repeated cross-sections
  - broadens potential applications by leveraging publicly available datasets
- **Straight-forward implementation** via pooled OLS
- Run **diagnostics** to distinguish between “spurious” and “real” shrinkage of treatment effects as number of interactions increases
- If a concern:
  - linearity and separability of covariate effect have testable implications
  - decrease dimension of RHS variables via penalty function, and still preserve efficiency
  - extend to generalized linear models for count outcomes

Context: Heterogenous effects via TWFE

Objective. Recover the ATT via DiD, relying on no anticipation and parallel-trends.

Typical implementation. TWFE to facilitate estimation of standard errors

Challenge. TWFE does not recover ATT if multiple groups and periods and heterogenous treatment effects, de Chaisemartin and D'Haultfoeuille (2020), Goodman-Bacon (2021).

Solution. "Clean controls:" either never-treated, not-yet-treated, or both. Treated groups defined by adoption date; do not compare later- to earlier-treated.

Various methods for ensuring "clean controls."

Context: Methods

"Subgroup": selects subgroups of treated and control, 2x2 DiDs

- Cengiz, Dube, Lindner, and Zipperer (2019), de Chaisemartin and D'Haultfœuille (2020), Callaway and Sant'Anna (2021), Sun and Abraham (2021)

"Imputation"/regression: counterfactuals via a linear regression specification

- Borusyak, Jaravel, and Spiess (2024), Wooldridge (2021+), Gardner, Thakral, To, and Yap (2024+), Harmon (2024+), **Deb et al. (2024+)**

Differences. Which clean controls, which pre-treatment periods, covariates or not, and estimation

- e.g., de Chaisemartin and D'Haultfœuille (2023), Callaway (2023), Roth et al. (2023) on parallel-trends/clean groups,
- Harmon (2024+) on efficiency,
- Egerod and Hollenbach (2024+) on power and coverage rates details

Context: Methods compared

Harmon (2024+): focuses on efficiency

- "imputation": homoskedasticity and no serial correlation
- "subgroup": serial dependence and for "short" post-treatment period
- "step-wise" DiD: serial dependence and for "long" time horizons
  - imputation regression estimator for first-differenced regression model

Egerod and Hollenbach (2024+): focuses on power and coverage rates

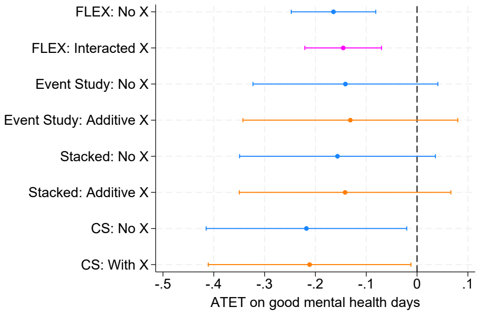
- even if point estimates are the same, standard errors vary due to different asymptotic arguments, reflecting distinct sources of uncertainty.

back

FLEX: Mental Health and Punitive Laws

Different Estimators

ATET: Lags and Leads Specifications



back

Wooldridge (2021+): Extended TWFE

Imputation/regression

- extended TWFE = **TWFE + large set of interaction terms**
  - uses intuition from the Mundlak device
- **linearity of CEF** of never-treated potential outcomes
  - which time periods: all
  - which clean controls: both never and not-yet treated
- **pooled OLS** for sequence of group-time ATTs, clustered standard errors
- aggregate across groups and time for ATT, standard error via Delta Method
- **efficient if Gauss-Markov** assumptions hold
  - homoskedasticity and no serial correlation in never-treated potential outcomes
- panel data and time-invariant covariates (with time varying effects)

back