

jwdid: Flexible Estimation of staggered DID designs

Fernando Rios-Avila

February 17, 2025

Introduction

- **Differences in Differences** (DiD) design is one of the most popular methods in applied microeconomics, because it requires relatively few assumptions to identify treatment effects.
 - No anticipation,
 - Parallel trends,
 - No spillovers
- This approach is so simple that canonical DiD, a 2x2 design, simply compares means of the outcome variable (before:after × treated:non-treated) to identify treatment effects.
 - Thus it can be used even if outcome is a limited dependent variable (binary, count, etc)

Canonical DiD: 2x2 design

- Consider the case for panel data (easiest scenario)
 - There are two groups, Treatment ($D=1$) and Control ($D=0$)
 - and all units are observed for two periods, a before and after ($T=0,1$)
- Under the assumption of full heterogeneity, the potential and realized outcomes for each unit are:

$$y_{it}(W) = \mu_i + \lambda_i t + \theta_i W \times t$$
$$y_{it} = D_i y_{it}(1) + (1 - D_i) y_{it}(0)$$

Canonical DiD: 2x2 design

In this framework, the TE for i is defined as:

$$\theta_i = y_{i,1}(1) - y_{i,1}(0)$$

Which may not be useful. Instead, we settle focus on something different. “average treatment effect” on treated (ATET or ATT)

$$\begin{aligned} ATT &= E[\theta_i | D_i = 1] = E[y_{i,1}(1) - y_{i,1}(0) | D_i = 1] \\ &= E_1[\theta_i] = E_1[y_{i,1}(1)] - E_1[y_{i,1}(0)] \end{aligned}$$

But is still not identified, because $E_1[y_{i,1}(0)]$ is not observed.

Canonical DiD: 2x2 design

How to calculate $E(y_{i,1}(0)|D = 1)$?

First, Decompose it:

$$E_1(y_{i,1}(0)) = E_1(y_{i,0}(0)) + E_1(\lambda_i)$$

- Under No anticipation: $E_1(y_{i,0}(0)) = E_1(y_{i,0}(1)) = E_1(y_{i,0})$
- Under Parallel trends: $E_1(\lambda_i) = E_0(\lambda_i) = E_0(y_{i,1} - y_{i,0})$
- Thus putting it all together:

$$\begin{aligned} ATT &= E_1[y_{i,1}] - [E_1(y_{i,0}) + (E_0[y_{i,1}] - E_0[y_{i,0}])] \\ &= [E_1[y_{i,1}] - E_1(y_{i,0})] - [E_0[y_{i,1}] - E_0[y_{i,0}]] \end{aligned}$$

How is this estimated?

- Setting aside estimation of Standard errors, the ATT can be estimated by simply comparing average outcomes for the treated and control groups before and after treatment.
- or using the following regression:

$$y_{it} = \alpha + \beta D_i + \gamma t + \theta(D_i \times t) + \epsilon_{it}$$

Where θ is the ATT.

- This is equivalent to comparing predictions of the outcome for the treated group, with and without treatment, in the post-treatment period.

But life its complicated

But life its complicated

- Canonical DiD is straight forward, but its also very restrictive.
 - What if we can identify and follow individuals or specific groups over time? (Individual/group fixed effects)
 - What if we have multiple periods? (Time fixed effects)
 - What if we want to control of $X' s$?
 - What if the treatment is staggered?
 - what if the outcome is a limited dependent variable?

And some times we make things even harder

- What if we are interested in X heterogeneity of ATT's?
 - Is the effect stronger for some groups?
- What if we have a dose treatment?
 - Is there a change in the effect based on the intensity of the treatment?
- What if we have multiple treatments?
 - What if we have multiple treatments (Cash vs In-kind transfers)?

ETWFE: A flexible approach to DiD

Answering some of these tough questions!

Panel data with multiple periods

- Having many individuals that can be followed across time can be easily adapted in the regression based DID approach.
- Specifically, instead of controlling for “treatment” class, we can control for individual fixed effects.

$$y_{it} = \gamma t + \theta(D_i \times t) + \beta_i + \epsilon_{it}$$

- In cases of balanced panels controlling for individual fixed effects gives you the same results as controlling for “ D_i ”

Panel data with multiple periods

- For multiple periods, the same logic applies. Instead of controlling for a before/after t , we can control for time fixed effect.

$$y_{it} = \theta(W_{it}) + \beta_i + \gamma_t + \epsilon_{it}$$

- $W_{it} = \mathbf{1}$ only for treated units in post-treatment periods, and $W_{it} = \mathbf{0}$ otherwise.
- θ would be the ATT for all post-treatment periods.
- Still possible and easy to estimate using regression.
- This is the [in]famous TWFE-DID estimator (generalized).

Controlling for X 's in DID

- Canonical DID assumes PTA holds:
 - Implicit assumption: TE is homogenous across X characteristics, or
 - Treated and control groups are similar.
- But what if that isn't the case?
 - We may want to control for those characteristics X

$$y_{it} = \theta(W_{it}) + \beta_i + \gamma_t + \delta X_i + \epsilon_{it}$$

- This is a simple extension of the DID regression,
 - But imposes the strong assumption that the TE is constant across subgroups.

Addressing X 's

- There are two solutions to the problem
 1. Make the assumption that the treatment effect is constant with respect to X .
 2. Relax the PTA and use CPTA

$$\text{PTA} : E_1(y_{i,1} - y_{i,0}) = E_0(y_{i,1} - y_{i,0})$$
$$\text{CPTA} : E_1(y_{i,1} - y_{i,0} | X) = E_0(y_{i,1} - y_{i,0} | X)$$

- In other words, one could simple estimate ATTs for all sub-groups:

$$ATT(X) = E_1[y_{i,1} | X] - E_1[y_{i,0} | X] - [E_0[y_{i,1} | X] - E_0[y_{i,0} | X]]$$

- Then simply aggregate $ATT = E_1[ATT(X)]$
- Most approaches usually provide this aggregate ATT (hidding the heterogeneity)

How is it done Empirically? (2x2 case)

- Sant'Anna and Zhao (2020) proposes few solutions:
 - Outcome Regression (Separate models for treated and control/pre and post periods with prediction)
 - Inverse Probability Weighting (IPW), to balance covariates
 - Combination of both (Doubly Robust)
- Jeffrey M. Wooldridge (2021) implicitly suggest an outcome regression with interactions

$$\begin{aligned} y_{it} = & \alpha & + \beta D & + \gamma t & + \theta(D \times t) \\ & + \lambda X & + \lambda_D X \times D & + \lambda_D X \times t & + \lambda_{DT} \tilde{X}(D \times t) \\ & + \epsilon_{it} & & & \end{aligned}$$

- Where $\tilde{X} = X - E[X|D, t]$ (but its not necessary)
- This is a flexible approach for embracing X heterogeneity and CPTA

What controls can be included?

- Tough question!
 1. In these models, more controls means more interactions, thus more parameters to be estimated, and more unstable estimates. Unless you have large panel or RC, you probably want to use only few controls.
 2. With Panel Data, aim to control for pre-treatment characteristics, that are time fixed. (Avoids bad controls)
 3. Use time varying variables with caution. In RC, the typical assumption is that time-varying variables are stationary, or that any changes are exogenous.

Note: `csdid[2]` forces you to use time fixed characteristics with panel data, But not for RC data

Note: `jwdid` does not impose this restriction.

Staggered treatment: The Fall

- What if the treatment is not applied at the same time to all units?
 - This is the case of staggered treatment.
 - And we usually impose the assumption of “once-treated, always-treated”
- The traditional approach: “Generalized TWFE-DID”

$$y_{it} = \beta_i + \gamma_t + \theta W_{it} + \epsilon_{it}$$

- However, this approach only works if the ATT is constant over time for all units.
 - But...there is no guarantee!
- When this is not the case, results are biased (even very biased)

Why?

- Standard OLS estimates identifies θ by comparing Pre and Post outcomes using various 2x2 designs.
 - contrasting units that changed treatment status (not-treated → treated)
 - ↻ Lets ignore cases of reversal (treated → not-treated)
 - With those that did not (Never-treated, Not-yet treated and **already treated**).
- This has been interpreted in two ways:
 1. Negative weight (Applied to Already Treated units)
 2. Incorrect control group

The solution(s)

- There are many solutions to the problem
 1. Avoid the problem using only pre-treatment data for estimation of counterfactuals (**Borusyak, Jaravel, and Spiess 2024; Gardner 2022**) (`did2s` and `did_imputation`)
 2. Use only good controls when estimating the ATT (**Callaway and Sant'Anna 2021**) (`csdid[2]`)
 3. Modify model specification to achieve 1 or 2. (**Jeffrey M. Wooldridge 2023; Jeffrey M. Wooldridge 2021; Deb et al. 2024; Yotov, Nagengast, and Rios-Avila 2024**) (`jwdid`)

How ETWFE addresses Staggered Treatment (`jwdid`)

- TWFE was not wrong. It was just too simple.

Wrong:

$$y_{it} = \beta_i + \lambda_t + \theta(W_{it}) + \epsilon_{it}$$

Right:

$$y_{it} = \beta_i + \lambda_t + \sum_{g=g_0}^G \sum_{t=g}^{t=T} \theta_{gt} \mathbf{1}(g, t) + \epsilon_{it}$$

- Allows ATT (θ_{gt}) to vary by group and time, using all not-yet treated units as controls.

```
1 jwdid y, tvar(tvar) ivar(ivar) gvar(gvar)
```

Testing for PTA

- The previous approach does not allow you to test for PTA.
 - Assumes that PTA holds!
- But that can be relaxed:

$$y_{it} = \beta_i + \lambda_t + \sum_{g=g_0}^G \sum_{t=t_0}^{t=g-2} \theta_{gt} \mathbf{1}(g, t) + \sum_{g=g_0}^G \sum_{t=g}^T \theta_{gt} \mathbf{1}(g, t) + \epsilon_{it}$$

- The control group are the Never treated (only)
 - Callaway and Sant'Anna (2021) `csdid[2]` with out controls and Balance panel
 - Sun and Abraham (2021) (but using absolute rather than relative time)

```
1 jwdid y, tvar(tvar) ivar(ivar) gvar(gvar) never
```

What if you only have Repeated Cross Section data?

- With repeated crosssection, not much changes.
 - You cannot estimate individual fixed effects
 - Instead you can use Group (G) fixed effects

$$y_{it} = \beta_g + \lambda_t + \sum_{g=g_0}^G \sum_{t=0}^{t=g-2} \theta_{gt} 1(g, t) + \sum_{g=g_0}^G \sum_{t=g}^{t=T} \theta_{gt} 1(g, t) + \epsilon_{it}$$

```
1 jwdid y, tvar(tvar) gvar(gvar) [never]
```

Small Problem, simple solution

`jwdid` may produce too many **ATT(G,T)** to analyze → aggregate!

- Simple: $ATT = \frac{\sum_g \sum_t \theta_{gt} \omega(g,t) 1(t \geq g)}{\sum_g \sum_t \omega(g,t) 1(t \geq g)}$ `estat simple`
- Group: $ATT(g) = \frac{\sum_t \theta_{gt} \omega(g,t) 1(t \geq g)}{\sum_t \omega(g,t) 1(t \geq g)}$ `estat group`
- Time: $ATT(t) = \frac{\sum_g \theta_{gt} \omega(g,t) 1(t \geq g)}{\sum_g \omega(g,t) 1(t \geq g)}$ `estat calendar`
- Event: $ATT(e) = \frac{\sum_g \sum_t \theta_{gt} \omega(g,t) 1(t-g=e)}{\sum_g \sum_t \omega(g,t) 1(t-g=e)}$ `estat event`
- where $\omega(g, t)$ is the weight (total number of units in group g observed at time t)

Controlling for X heterogeneity

- Allowing for X heterogeneity is simple. Simply consider a flexible model with interactions!

$$\begin{aligned}
 y_{it} = & \beta_0 + \beta_i + \lambda_t + \sum_{g=g_0}^G \sum_{t=g}^T \theta_{gt} \mathbf{1}(g, t) + \\
 & \delta X_{it} + \sum_g \delta_g X_{it} \mathbf{1}(g) + \sum_t \delta_t X_{it} \mathbf{1}(t) + \sum_{g=g_0}^G \sum_{t=g}^T \delta_{gt} X_{it} \mathbf{1}(g, t) + \\
 & \epsilon_{it}
 \end{aligned}$$

- Considerations:
 - Some of these interactions will be collinear with i/g and t fixed effects (Thus dropped)
 - X_{it} could be left as is (θ_{gt} is not ATT(G,T))
 - or substitute for \tilde{X}_{it} (where $\tilde{X}_{it} = X_{it} - E[X|g, t]$). (θ_{gt} is ATT(G,T))
 - δ_{gt} Is the impact of X on the ATT(g,t)

Controlling for X heterogeneity

In Stata, `jwdid` can also estimate both types of models:

```
1 * Demeaning Data ( $\theta$  is ATT(G,T))
2 jwdid y x1 i.x2, tvar(tvar) ivar(ivar) gvar(gvar) [never]
3
4 * Using X as is ( $\theta$  is not ATT(G,T))
5 * May be faster
6 jwdid y x1 i.x2, tvar(tvar) ivar(ivar) gvar(gvar) [never] xasis
```

- Both approaches will produce the same results once aggregations are obtained.

Estimating ATTs when using *xaxis*?

- When using X axis, θ_{gt} is not ATT(G,T).
- In such cases, we need a different approach to estimate the ATT(G,T):

$$\theta_{gt} = E[\hat{y}_{i,t}(X_{it}, 1(g, t) = 1) - \hat{y}_{i,t}(X_{it}, 1(g, t) = 0) | g, t]$$

- $\hat{y}_{i,t}(X_{it}, 1(g, t) = 1)$ Model prediction as if treated (as observed)
- $\hat{y}_{i,t}(X_{it}, 1(g, t) = 0)$ Model prediction as if **NOT** treated (counterfactual)

Limited Dependent Variables

- As with OLS, one Big advantage of using OLS is that it can be used for any type of outcomes.
 - Minor inconvenience: Heteroskedasticity and non-sensical predictions
- Using OLS, however, assumes that **[C]PTA** holds for the observed outcome (At levels).
What if that is not the case?
 - Wages grow at the same rate for treated and control (PTA holds)
 - But for high earners, wages grow faster in absolute terms (PTA does not hold)
- Jeffrey M. Wooldridge (2023) suggest focusing on the latent variable, not the outcome!

Limited Dependent Variables

There are few changes to consider:

1. We change the model and focus from outcome to latent variable (think GLM)

$$y_{it}^* = \beta_g + \lambda_t + \sum_{g=g_0}^G \sum_{t=g}^{t=T} \theta_{gt} \mathbf{1}(g, t)$$

$$E(y_{it}) = G(y_{it}^*)$$

Thus we need a model appropriate for $G()$ (logit, poisson, tobit, etc)

2. Most non-linear models have issues adding fixed effects (incidental parameters problem)
 - Use Cohort/Group fixed effects (instead of individual)
 - Use Correlated Random Effects (CRE) as an alternative to fixed effects.
3. θ_{gt} is not ATT(G,T) on outcome, but on latent variable.

Limited Dependent Variables `jwdid`

- `jwdid` can estimate these type of models simply by using `method()`.
 - Almost any model available in Stata can be called. Few cases may require tweaks (mostly two-word commands and clustered SE calls)
 - ⇒ when doing so, `jwdid` will use cohort FE instead of individual FE
 - Only `ppmlhdfc` (so far) would still add individual fixed effects.
- the CRE correction can be called using `cre` option.
 - Output not be displayed, for space.

```
1 ** uses cohort FE
2 jwdid y, tvar(tvar) ivar(ivar) gvar(gvar) [never] method(logit)
3 ** uses CRE correction
4 jwdid y, tvar(tvar) ivar(ivar) gvar(gvar) [never] method(logit) cre
```

Limited Dependent Variables `jwdid`

Note that θ_{gt} is not ATT(G,T) on outcome, but on latent variable.

however, one can request aggregations of the latent or outcome variable afterwards

```
1 estat [simple|group|calendar|event] /// Default uses "method()" margins options
2 estat [simple|group|calendar|event], predict(xb) /// but one can use other outcomes
```

Estimating ATT based on X Heterogeneity

- Some times you may want to estimate ATT for different sub-groups.
- `jwdid` already incorporates this by using flexible specifications, but produces **Average** ATT's.
- Thanks to `margins`, it is possible to estimate ATTs for different *discrete* sub-groups (not continuous):

```
1 ** setup
2 jwdid y i.x1 x2, tvar(tvar) ivar(ivar) gvar(gvar) never
3 ** ATTs for specific groups
4 estat simple // average ATT
5 estat simple, over(x1) // ATT estimated for each group of x1
6 // For observations where x2 is between 0 and 1
7 estat [simple|calendar|group|event], ores( x2>0 | x2<1 )
8 // For observations where x1 is 0
9 estat [simple|calendar|group|event], ores( x1==0 )
```

Treatment Intensity and Heterogeneity

- Another common scenario is when treatment varies across groups.
 - Case 1: Treatment is the same, but dose/intensity (0-1) varies
 - Case 2: Each treatment is different (Treatment A, B or C)
- The `jwdid` framework can be adapted to these scenarios, albeit with limitations.

Case 1: Treatment Intensity

- The Treatment intensity should be defined between 0-1.
- Impose the assumption that $ATT(G,T)(TI)$ is linear on the treatment intensity (TI).
 - This could be relaxed.
- It requires you to have a variable `trtvar` that defines treatment intensity.

```
1 ** setup
2 jwddid y i.x1 x2, tvar(tvar) ivar(ivar) gvar(gvar) trtvar(trtvar) [never]
3 ** ATT aggregation
4
5 *** Estimates Treatment effect, assuming Full Intensity (T=1)
6 estat [simple|group|calendar|event]
7
8 *** Estimates Treatment effect, assuming intensity as observed
9 estat [simple|group|calendar|event] , asis
```

- This gives a single ATT, but could be combined with `ores()` or `over()` to estimate ATTs for different sub-groups.

Case 2: Treatment Heterogeneity (including On and Off)

- This is a case where you could have different treatments implemented at the same time. (**H**)
- In principle, this could be done simply using different models for each treatment type. (all groups share the same control group)
- This entails a very flexible model. (Interaction **T x [G x H]**)
- Alternatively, we can assume that Heterogeneity only affects the ATT(G,T,H)

$$y_{it} = \beta_g + \lambda_t + \sum_{g=g_0}^G \sum_{t=g}^{t=T} \sum_h \theta_{gth} 1(g, t, h) + \epsilon_{it}$$

Case 2: Treatment Heterogeneity (including On and Off)

Estimation with `jwdid` is very simple:

```
1 ** setup
2 jwdid y i.x1 x2, tvar(tvar) ivar(ivar) gvar(gvar) xattvar( trt_l trt_h) [never]
3 ** Assume trt_m is the base treatment. trt_l and trt_h are potential treatments.
4 ** The base-line treatment will be dropped.
```

- Aggregation methods can be done as before
 - `estat [simple|group|calendar|event]` provide a single ATT
 - combined with `over()` or `ores()`, one could estimate ATT Heterogeneity

Statistical Inference:

- One of the main advantages of `jwdid` is that we do not need to be concerned with the estimation of standard errors.
 - They are automatically calculated using the package of your choice.
- Some notes, however
 - If using Panel data, SE are automatically clustered at the individual level.
 - if using RC, SE are not clustered. (you need to add that to the command)
 - By default, they will ignore uncertainty of covariates
- Still, that can be easily fixed adding `vce(unconditional)` to aggregation commands
 - `estat [simple|group|calendar|event], vce(unconditional)`
- Caveat: This does not work if using `reghdfe` or `ppmlhdfe` as the estimation method.
 - Use `regress cre` or `poisson cre` instead (Stata does this)

Other Options of interest: Gravity Models and Beyond

- `jwdid` has other “advanced” options that could further help model specification.
- `fevar()`: Allows introducing FE other than Panel (only with `reghdfe` or `ppm1hdfe`)

$$y_{it} = \beta_g + \lambda_t + \sum_{g=g_0}^G \sum_{t=g}^{t=T} \theta_{gt} \mathbf{1}(g, t) + \omega_j + \epsilon_{it}$$

- `exovar()`: Variables not interacted with treatment G nor time T , nor both.

$$y_{it} = \beta_g + \lambda_t + \sum_{g=g_0}^G \sum_{t=g}^{t=T} \theta_{gt} \mathbf{1}(g, t) + \phi X_{it} + \epsilon_{it}$$

Other Options of interest: Gravity Models and Beyond

- `xtvar()` and `xgvar()`: variables that will only interact the time or group fixed effects.

$$y_{it} = \beta_g + \lambda_t + \sum_{g=g_0}^G \sum_{t=g}^{t=T} \theta_{gt} \mathbf{1}(g, t) + \sum_{g=g_0}^G \gamma_g \mathbf{1}(g) X_{it} + \epsilon_{it}$$

$$y_{it} = \beta_g + \lambda_t + \sum_{g=g_0}^G \sum_{t=g}^{t=T} \theta_{gt} \mathbf{1}(g, t) + \sum_{t=t_0}^T \gamma_t \mathbf{1}(g) X_{it} + \epsilon_{it}$$

- `anticipation(#)`: Allows to set a different period as baseline for the treatment. (default is 1) (g-1)
- `hettype()`: Allows to impose some restrictions on heterogeneity type. Default its `timecohort`
 - `time`, `cohort`, `event`, `twfe`, `eventcohort`
- And For Event aggregation:
 - `window(#1 #2)` as is window
 - `cwindow(#1 #2)` Censored window

Conclusion

- The battle of methods trying to address the problems of the traditional TWFE-DID has been long and hard (2020-2023?)
 - I can probably say that it has settled down, with few options available: `csdid`, `did_imputation`, `did_multiple_dyn` and `jwdid`
- Under some assumptions, they are all equivalent. But some are more flexible than others.
- I wrote `jwdid` to be as flexible as possible
 - some of the extensions and options go beyond Wooldridge's work
 - some were developed independently
- I hope you find it useful.

Thank you!

If you are interested, you can install the latest version of `jwdid` using

```
net install jwdid, from(https://raw.githubusercontent.com/friosavila/stpackages/main)
```

You can find me on

- Twitter: [friosavila](#)
- GitHub: [friosavila](#)
- Email: friosa@gmail.com

References

- Borusyak, Kirill, Xavier Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *Review of Economic Studies* 91 (6): 3253–85. <https://doi.org/10.1093/restud/rdae007>.
- Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230. <https://doi.org/10.1016/j.jeconom.2020.12.001>.
- Deb, Partha, Edward Norton, Jeffrey Wooldridge, and Jeffrey Zabel. 2024. "A Flexible, Heterogeneous Treatment Effects Difference-in-Differences Estimator for Repeated Cross-Sections." <https://doi.org/10.3386/w33026>.
- Gardner, John. 2022. "Two-Stage Differences in Differences." <https://doi.org/10.48550/ARXIV.2207.05943>.
- Sant'Anna, Pedro H. C., and Jun Zhao. 2020. "Doubly Robust Difference-in-Differences Estimators." *Journal of Econometrics* 219 (1): 101–22. <https://doi.org/10.1016/j.jeconom.2020.06.003>.
- Wooldridge, Jeffrey M. 2023. "Simple Approaches to Nonlinear Difference-in-Differences with Panel Data." *The Econometrics Journal* 26 (3): C31–66. <https://doi.org/10.1093/ectj/utad016>.
- Wooldridge, Jeffrey M. 2021. "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3906345>.
- Yotov, Yoto, Arne Nagengast, and Fernando Rios-Avila. 2024. "The European Single Market and Intra-Eu Trade: An Assessment with Heterogeneity-Robust Difference-in-Differences Methods." <http://dx.doi.org/10.2139/ssrn.4848975>.