

Rolling Estimation Methods in Stata for Robust Estimation of Treatment Effects with Staggered Interventions

2026 UK Stata Conference
September 3-4, London

Jeff Wooldridge

Department of Economics, Michigan State University

Collaborators: Elizabeth Hur

Department of Economics, Michigan State University

Soo Jeong Lee

Department of Economics, Southern Illinois University

1. Introduction
2. Notation and Definitions
3. Regression-Based Methods
4. Rolling (Lags Only) Doubly Robust Methods
5. Long-Differencing (Leads and Lags) Doubly Robust Methods
6. Unit-Specific Trends in Rolling Methods
7. Small Sample Sizes
8. Summary

1. Introduction

- Much interest in staggered policy interventions with heterogeneity.
- The linear, constant effect model is restrictive.
 - ▶ Estimation by two-way fixed effects (TWFE).
- de Chaisemartin and D'Haultfoeuille (2020, AER), Goodman-Bacon (2021, J of E), Borusyak, Jaravel, and Spiess (2024, REStud) provide analytical results and critiques.
- Many estimators have been proposed to overcome the “forbidden comparisons” problem.
 - ▶ Also include covariates to relax parallel trends.

2. Notation and Definitions

- T time periods with no units treated in $t = 1$.
- First unit is treated at $t = q \leq T$.
 - ▶ Special case is common timing.
- Initially, no reversibility: once a unit is subjected to the intervention, it stays in place.
- Treated units are added up through $t = T$.
- Is there a never treated group?
 - ▶ Assume that for this talk; easy to relax.

- Notation: for $g \in \{q, \dots, T\}$, $Y_t(g)$ is the outcome if the unit is first subjected to the intervention at time g .
 - ▶ In $Y_t(g)$, the number of treated periods decreases with g .
 - ▶ $Y_t(T)$ is the outcome when treated in only the final period.
 - ▶ $Y_t(\infty)$ is the outcome if a unit is never treated in $\{q, \dots, T\}$.

- Treatment effects of primary focus:

$$TE_{gt} = Y_t(g) - Y_t(\infty), g = q, \dots, T; t = g, \dots, T$$

- ▶ For any t , $Y_t(\infty)$ is the outcome in the control state.
- Exhaustive and mutually exclusive treatment cohort dummy variables:

$D_g = 1$ if unit is first subjected to intervention at $g \in \{q, \dots, T\}$

$$D_\infty = 1 - (D_q + D_{q+1} + \dots + D_T)$$

- ▶ $D_{i\infty} = 1$ means unit i is never treated (up through T).
- ▶ $D_{i\infty} = 0$ for all i means that all units are treated by time T .

- Goal is to estimate the average treatment effects on the treated:

$$\tau_{gt} \equiv E(TE_{gt}|D_g = 1) = E[Y_t(g) - Y_t(\infty)|D_g = 1], t = g, g + 1, \dots, T$$

- We have a row vector of time-constant covariates $\mathbf{X}$.
 - ▶ Assume no bad controls: $\mathbf{X} = \mathbf{X}(g) = \mathbf{X}(\infty)$.
 - ▶ Can allow time-varying covariates, but have to be careful. Are the affected by the intervention?

3. Regression-Based Methods

- To overcome the “bad comparisons” problems of constant effect models, or imposing common effects by exposure time, flexible regression-based methods have been proposed.
- Has led to many suggested solutions that allow for heterogeneous effects.
- Broadly speaking, the methods divide into two approaches.

1. “Lags only”: Use implications of parallel trends assumption throughout the entire time period.

- ▶ Effectively, uses all pre-treatment outcomes to construct a reference outcome.

- ▶ Borusyak, Jaravel, Spiess (2024, REStud) [BJS]: imputation with unit-specific fixed effects.

- ▶ Wooldridge (2025, Empirical Economics): imputation on cohort dummies, pooled estimation on cohort dummies; extended two-way fixed effects.

- ▶ Wooldridge (2023, Econometrics Journal): nonlinear models estimated by imputation on cohort dummies, pooled quasi-MLE.

2. “Leads and lags” (“event study”): choose a reference period and also estimate pre-treatment effects.

- ▶ Typically, the reference period is the one just before the intervention begins.
- ▶ Sun and Abraham (2021, J of Econometrics): full set of leads and lags; covariates are not included in a flexible way.
- ▶ Callaway and Sant’Anna (2021, J of Econometrics): apply regression adjustment to long-differences.
- ▶ Wooldridge (2025): a flexible version of Sun and Abraham with covariates.
- ▶ Wooldridge (2023): logit, fractional logit, Poisson versions of leads and lags on cohort dummies.

Lags Only Regression

- Wooldridge (2025) shows that no anticipation and parallel trends assumptions, conditional on $\mathbf{X}$, lead to estimation of a flexible equation.
- Let i denote a random draw.
- Let fs_t be the time-period dummies.

$$\begin{aligned}
E(Y_{it}|D_{iq}, \dots, D_{iT}, \mathbf{X}_i) = & \alpha + \sum_{g=q}^T \beta_g D_{ig} + \mathbf{X}_i \boldsymbol{\kappa} + \sum_{g=q}^T (D_{ig} \cdot \dot{\mathbf{X}}_{ig}) \boldsymbol{\xi}_g \\
& + \sum_{s=2}^T \gamma_s f s_t + \sum_{s=2}^T (f s_t \cdot \mathbf{X}_i) \boldsymbol{\pi}_s \\
& + \sum_{g=q}^T \sum_{s=g}^T \tau_{gs} (D_{ig} \cdot f s_t) \\
& + \sum_{g=q}^T \sum_{s=g}^T (D_{ig} \cdot f s_t \cdot \dot{\mathbf{X}}_{ig}) \boldsymbol{\rho}_{gs}
\end{aligned}$$

- Lots of heterogeneity.
 - ▶ The ATTs τ_{gt} are unrestricted.
 - ▶ Effects of covariates can change with treatment cohort and calendar time.
 - ▶ Treatment effects can change with $\mathbf{X}$: observed heterogeneity.
- To operationalize, use subsample averages:

$$\dot{\mathbf{X}}_{ig} = \mathbf{X}_i - \bar{\mathbf{X}}_g$$

Algebraic Equivalences

- Wooldridge (2025) shows the following (all with time fixed effects):

$$\begin{aligned}\text{Cohort-specific imputation} &= \text{POLS (cohort dummies)} \\ &= \text{RE (cohort dummies)} \\ &= \text{FE (unit dummies)} \\ &= \text{BJS imputation (unit dummies)}\end{aligned}$$

- ▶ With a balanced panel and time-constant controls, there is only one lags only estimator.

- Define the time-varying treatment indicator W_{it} (when treatment is absorbing):

$$W_{it} = D_{iq} \cdot (fq_t + \cdots + fT_t) + \cdots + D_{iT} \cdot fT_t$$

$$W_{it} \cdot D_{iq} \cdot fs_t = D_{iq} \cdot fs_t$$

- The following are equivalent Stata commands, assuming:
 - ▶ Time-constant covariates.
 - ▶ Balanced panel.
- Aggregate by exposure time and also a single weighted estimate.

```
xtset id year
xthdidregress twfe (y x1 ... xK) (w),
    group(id)
estat aggregation, dynamic
estat aggregation
```

- Fernando Rios-Avila [Wooldridge (2023,2025)]:

```
jwddid y x1 ... xK, ivar(id) tvar(year)
```

```
gvar(cohort)
```

```
estat event
```

```
estat simple
```

- Kirill Borusyak [BJS (2024)]:

```
did_imputation y id year cohort, timec(x)
```

```
allhorizons
```

```
did_imputation y id year cohort, timec(x)
```

Leads and Lags Regression

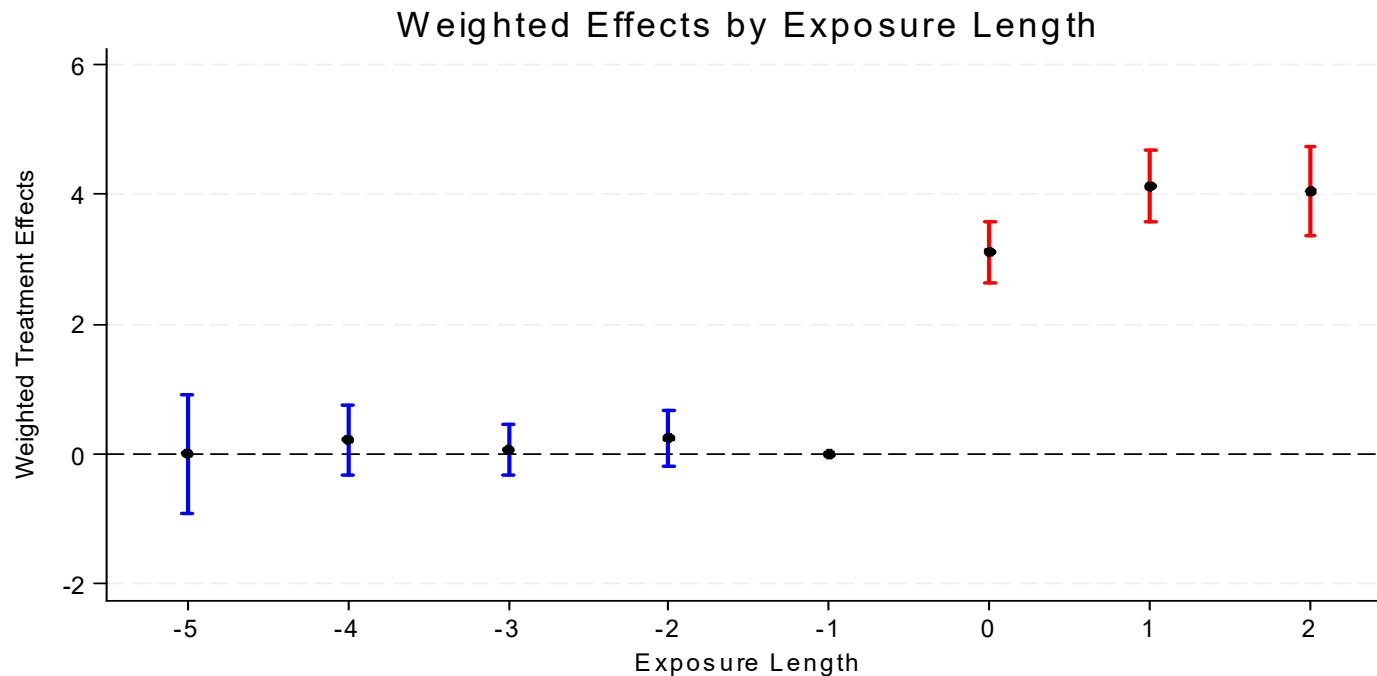
- Wooldridge (2025) proposes a leads and lags equation by adding pre-treatment indicators:

$$D_{ig} \cdot fs_t, D_{ig} \cdot fs_t \cdot \dot{\mathbf{X}}_{ig}, s \in \{1, \dots, g-2\}$$

- ▶ These indicators “lead” the intervention.
- ▶ Omitting $g-1$ for each cohort g forces $g-1$ to be the reference period.

$$\begin{aligned}
E(Y_{it}|D_{iq}, \dots, D_{iT}, \mathbf{X}_i) = & \alpha + \sum_{g=q}^T \beta_g D_{ig} + \mathbf{X}_i \boldsymbol{\kappa} + \sum_{g=q}^T (D_{ig} \cdot \dot{\mathbf{X}}_{ig}) \boldsymbol{\xi}_g \\
& + \sum_{s=2}^T \gamma_s f s_t + \sum_{s=2}^T (f s_t \cdot \mathbf{X}_i) \boldsymbol{\pi}_s \\
& + \sum_{g=q}^T \sum_{s \neq g-1}^T \tau_{gs} (D_{ig} \cdot f s_t) \\
& + \sum_{g=q}^T \sum_{s \neq g-1}^T (D_{ig} \cdot f s_t \cdot \dot{\mathbf{X}}_{ig}) \boldsymbol{\rho}_{gs}
\end{aligned}$$

- Aggregate using cohort shares to get a single event (exposure time) plot.
- Use coefficients on the pre-treatment dummies, $D_{ig} \cdot f_{st}$, $s \in \{1, \dots, g-2\}$, to test for pre-trends.



- Wooldridge (2025):

Cohort-specific imputation = POLS (cohort dummies)
= RE (cohort dummies)
= FE (unit dummies)
= BJS imputation (unit dummies)

- Another equivalence is linear regression adjustment using “long” differences as in Callaway and Sant’Anna (2021).

- For $s \neq g - 1$,

$Y_{is} - Y_{i,g-1}$ on $1, D_{ig}, \dot{\mathbf{X}}_{ig}, D_{ig} \cdot \dot{\mathbf{X}}_{ig}$ using $D_{ig} = 1$ or $D_{i\infty} = 1$

- Control units are the never treated units.

- Again, with balanced panel and time-constant covariates, several Stata commands are the same.

```
xthdidregress ra (y x1 ... xK) (w), group(id)
estat aggregation, dynamic
estat aggregation
jwddid y x1 ... xK, ivar(id) tvar(year)
      gvar(cohort) never
estat event
estat simple
```

```
csdid y x1 ... xK, ivar(id) time(year)
      gvar(cohort) long2 method(reg)
estat event
estat simple
```

- Also identical to Sun and Abraham (2020) without covariates.

4. Rolling (Lags Only) Doubly Robust Methods

- Why move beyond regression?
 - ▶ To achieve some resiliency to functional form.
 - ▶ Caution: Doubly robust estimators that combine regression with propensity score weighting do not relax conditional parallel trends.
- As before, the goal is to identify

$$\tau_{gt} \equiv E[Y_t(g) - Y_t(\infty) | D_g = 1], g = q, \dots, T; t = g, \dots, T.$$

Assumption NBC (No Bad Controls): $\mathbf{X} = \mathbf{X}(g) = \mathbf{X}(\infty)$. $\square$

- Covariates are not changing due to future assignment.

Assumption CNA (Conditional No Anticipation): For

$g \in \{q, \dots, T\}$, $t \in \{1, \dots, g-1\}$ and covariates $\mathbf{X}$,

$$E[Y_t(g)|D_g = 1, \mathbf{X}] = E[Y_t(\infty)|D_g = 1, \mathbf{X}]. \quad \square$$

- Pre-treatment “treatment” effects are zero.

Assumption CPT (Conditional Parallel Trends): For $t = 2, \dots, T$ and controls $\mathbf{X}$,

$$E[Y_t(\infty) - Y_1(\infty)|D_q, \dots, D_T, \mathbf{X}] = E[Y_t(\infty) - Y_1(\infty)|\mathbf{X}]. \quad \square$$

- Unconfoundedness with respect to the trends; allows confoundedness in levels.

A Transformation and Unconfoundedness

- Lee and Wooldridge (2026a, JBES):

$$\begin{aligned} Y_t(g) - Y_t(\infty) = & \left[Y_t(g) - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_s(g) \right] \\ & - \left[Y_t(\infty) - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_s(\infty) \right] \\ & + \frac{1}{(g-1)} \sum_{s=1}^{g-1} [Y_s(g) - Y_s(\infty)] \end{aligned}$$

- Under no anticipation, the last term has zero expectation, and so

$$\tau_{gt} = E[\dot{Y}_{tg}(g)|D_g = 1] - E[\dot{Y}_{tg}(\infty)|D_g = 1]$$

$$\dot{Y}_{tg}(g) \equiv Y_t(g) - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_s(g)$$

$$\dot{Y}_{tg}(\infty) = Y_t(\infty) - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_s(\infty)$$

- The averaging is over all pre-intervention periods for cohort g .
 - The lags only regression does something similar.
- First term is easy to estimate: average the transformed outcome over units in cohort g at time t .

- Cohorts $h = t + 1, \dots, T$ have not yet been treated.
- For estimating τ_{gt} for $t \geq g$, valid controls are

$$A_{t+1} \equiv D_{t+1} + D_{t+2} + \dots + D_T + D_\infty = 1$$

- Key is that, given $D_g + A_{t+1} = 1$, D_g is unconfounded for $\dot{Y}_{tg}(\infty)$ conditional on $\mathbf{X}$.
 - Cohorts already treated by period t are not used as controls.

Rolling Estimation (Lags Only) Methods

- Once the unit-specific demeaning has been done and the control group specified, any standard treatment effects estimator can be applied.

Lee and Wooldridge (2026a, JBES):

1. For the cohort g effects and for $t \in \{g, g + 1, \dots, T\}$, compute

$$\dot{Y}_{itg} \equiv Y_{it} - \frac{1}{(g-1)} \sum_{s=1}^{g-1} Y_{is} \equiv Y_{it} - \bar{Y}_{i,g-1}$$

2. Choose as the control group cohorts $h \in \{t + 1, \dots, T, \infty\}$ (not-yet-treated units).

3. Using the data with

$$D_{ig} + A_{i,t+1} = 1$$

apply standard TE methods (RA, IPW, AIPW, IPWRA, matching, machine learning) to

$$\{(\dot{Y}_{itg}, D_{ig}, \mathbf{X}_i)\}$$

► Can use a subset of the control cohorts, such as the NT group:

$$D_{ig} + D_{i\infty} = 1.$$

- User-written Stata command, `lwdid` (Soo Jeong Lee).

```
ssc install lwdid
```

```
lwdid y x1 ... xK, ivar(id) tvar(year)  
      gvar(cohort) rolling(demean) method(ipwra)
```

- ▶ Assumes `cohort` has been defined as with `jwdid`.
- ▶ Default is to use the not-yet-treated (including never treated) as controls.
- ▶ Add the `graph` option to obtain a “leads and lags” graph.
- ▶ The pre-trends are estimated using a “reverse” rolling method through $g - 2$ for cohort g .

5. Long-Differencing (Leads and Lags) Methods

- Recall the CS (2021) long-difference regressions using period $g - 1$ as the reference period and the NT group as the control group:

$$Y_{it} - Y_{i,g-1} \text{ on } 1, D_{ig}, \dot{\mathbf{X}}_{ig}, D_{ig} \cdot \dot{\mathbf{X}}_{ig} \text{ using } D_{ig} = 1 \text{ or } D_{i\infty} = 1$$

- The following sets of commands are identical.

```
csdid y x1 ... xK, ivar(id) time(year)
      gvar(cohort) method(reg) notyet long2
estat event
estat simple
```

- `lwdid` uses the not-yet-treated units as the controls.
 - Reports the estimates aggregated by exposure time as the default.
- To reproduce `csdid` with `method(reg)`:

```
lwdid y x1 ... xK, ivar(id) tvar(year)
      gvar(cohort) rolling(demean)
      method(ra) pre(1)
```

- A similar derivation of the CS approach is instructive.
- “Long” differencing:

$$Y_t(g) - Y_t(\infty) = [Y_t(g) - Y_{g-1}(g)] - [Y_t(\infty) - Y_{g-1}(\infty)] \\ + [Y_{g-1}(g) - Y_{g-1}(\infty)]$$

- By no anticipation,

$$E[Y_{g-1}(g) - Y_{g-1}(\infty) | D_g = 1] = 0$$

- Therefore, under NA,

$$E[Y_t(g) - Y_t(\infty)|D_g = 1] = E[Y_t(g) - Y_{g-1}(g)|D_g = 1] \\ - E[Y_t(\infty) - Y_{g-1}(\infty)|D_g = 1]$$

- First part is, as usual, easy:

$$E[Y_t(g) - Y_{g-1}(g)|D_g = 1] = E(Y_t - Y_{g-1}|D_g = 1)$$

- For the second, need CPT assumption to apply to $Y_t(\infty) - Y_{g-1}(\infty)$:

$$E[Y_t(\infty) - Y_{g-1}(\infty)|D_g = 1, \mathbf{X}] = E[Y_t(\infty) - Y_{g-1}(\infty)|D_\infty = 1, \mathbf{X}]$$

- ▶ The never treated group can be used as a control.
 - ▶ Also can use the not-yet treated under CNA.
 - ▶ Only need CPT starting in period $g - 1$.
 - ▶ But remember, this can lead to more bias if CPT fails into the treatment periods.
- The CPT assumption is unconfoundedness in the long difference.

Callaway and Sant'Anna (2021, Journal of Econometrics):

1. Create “long” differences

$$\mathring{Y}_{itg} \equiv Y_{it} - Y_{i,g-1}, t \geq g$$

2. Use as a control group those not yet subjected to the intervention at time t :

$$D_{ig} + D_{i,t+1} + D_{i,t+2} + \cdots + D_{iT} + D_{i\infty} \equiv D_{ig} + A_{i,t+1} = 1$$

► Or, only the never treated:

$$D_{ig} + D_{i\infty} = 1$$

3. In period $t \geq g$, use data $D_{ig} + A_{i,t+1} = 1$ and apply any cross-sectional treatment effect method (RA, IPW, AIPW, IPWRA, matching, machine learning) to

$$\left\{ \left(\overset{\circ}{Y}_{itg}, D_{ig}, \mathbf{X}_i \right) \right\}$$

- Can restrict attention to $D_{ig} + D_{i\infty} = 1$ (NT group as control).
 - ▶ Or any subset of NYT.
- Same strategy works for pre-treatment effects.
 - ▶ Obtain event-study estimates.
 - ▶ Test for pre-trends.

- The “rolling” method puts weight $1/(g - 1)$ on each of the pre-treatment outcomes $Y_{is}, s = 1, \dots, g - 1$.
- CS (2021) puts weight one on $Y_{i,g-1}$ and zero on $Y_{i1}, \dots, Y_{i,g-2}$.
- Optimal weights are available under certain assumptions.
 - Under system homoskedasticity – constant variances and covariances conditional on $(\mathbf{D}_i, \mathbf{X}_i)$ – can use unrestricted feasible GLS [Cui and Wooldridge (2026, WP)].

- In Stata 18, `xthdidregress aipw` (or `ipw` or `ra`).
 - ▶ These are different event study (leads and lags) estimators.
- User-contributed Stata command `csdid` (Fernando Rios-Avila).
- Doing the long differencing “by hand,” allows `ipwra` or `match` or `psmatch`.
 - ▶ `teffects` allows clustering (at higher level than unit).
 - ▶ With lots of controls, can apply machine learning.

```

. * Generated data.

. use did_staggered_6

. gen cohort = 0

. replace cohort = 2004 if d4
(714 real changes made)

. replace cohort = 2005 if d5
(498 real changes made)

. replace cohort = 2006 if d6
(216 real changes made)

. * The lwdid t statistics are the usual ones that divide the estimate by
. * its (cluster-robust) standard error. But the confidence intervals
. * are joint CIs across all aggregated effects.

```

. * Lee-Wooldridge (2026a) doubly robust estimator, not-yet-treated units as controls:

. lwdid y x, ivar(id) tvar(year) gvar(cohort) rolling(demean) method(ipwra)

 lwdid [large-N mode] rolling=demean method=ipwra

Aggregated WATT(r) estimates

effect	WATT	Std. err.	t	P> t	[95% CI/band]		N cells
Pre_avg	.0240158	.04802118	0.50	0.617	-.07010399	.11813558	
Post_avg	3.6520953	.17396758	20.99	0.000	3.3111251	3.9930655	
-5	.25500409	.32760249	0.78	0.436	-.59341698	1.1034252	
-4	-.11380955	.17354329	-0.66	0.512	-.56324993	.33563082	
-3	-.03454928	.12291173	-0.28	0.779	-.35286463	.28376607	
-2	.11655414	.13908531	0.84	0.402	-.24364736	.47675565	
-1	0	0	.	.	0	0	
0	3.0925723	.21316642	14.51	0.000	2.5405164	3.6446281	
1	4.0060068	.25488941	15.72	0.000	3.3458972	4.6661164	
2	4.1703841	.32386368	12.88	0.000	3.3316457	5.0091225	

Note: Pre_avg/Post_avg confidence intervals are pointwise normal.

Note: Event-study confidence bands are simultaneous (reps = 999)

; t statistics and p-values are pointwise normal.

. * In csdid, default is augmented IPW (AIPW) with covariate balancing
. * propensity score weights. The NT units are used as controls (the default).
. *
. * Use the long2 option to produce the typical ES estimates prior to the
. * intervention, where the reference period is the one just before the
. * intervention. The default is to use adjacent time periods. The choice
. * does not affect the estimated ATTs.

. * Because these data are generated with no serial correlation or heteroskedasticity
. * in the time-varying shocks, csdid is inefficient, and there is not guarantee
. * that including all NYT units is more efficient. jwdid without the
. * "never" option is efficient in this case.

```
. qui csdid y x, ivar(id) time(year) gvar(cohort) long2
```

```
. estat event
```

ATT by Periods Before and After treatment

Event Study:Dynamic effects

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Pre_avg	.0627463	.2244361	0.28	0.780	-.3771405	.5026331
Post_avg	3.753215	.2150635	17.45	0.000	3.331698	4.174731
Tm5	-.09348	.5180335	-0.18	0.857	-1.108807	.9218471
Tm4	.1791746	.2967618	0.60	0.546	-.4024679	.7608172
Tm3	-.0067565	.1976236	-0.03	0.973	-.3940917	.3805787
Tm2	.172047	.2224558	0.77	0.439	-.2639584	.6080524
Tp0	3.099835	.24568	12.62	0.000	2.618311	3.581359
Tp1	4.127951	.2798182	14.75	0.000	3.579517	4.676384
Tp2	4.031859	.3564757	11.31	0.000	3.333179	4.730538

```
. estat simple
```

Average Treatment Effect on Treated

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
ATT	3.669764	.2063346	17.79	0.000	3.265355	4.074172

Simulations

- $T = 6$, staggered entry at $q = 4$.
- Two covariates (gamma, Bernoulli).
- Parallel trends conditional on $\mathbf{X}$.
- Population $R^2 \approx 0.241$.
- Approximate cohort shares:

$$\rho_{\infty} = 0.358, \rho_4 = 0.291, \rho_5 = 0.225, \rho_6 = 0.126$$

- $N = 500$, Replications = 1,000.
- CS uses the default AIPW estimator and the NT units as controls.

Conditional Mean and Propensity Score Correctly Specified

	SATT	ETWFE		LW: IPWRA		CS: AIPW	
		Lags Only		Lags Only		Leads and Lags	
	Mean	Mean	SD	Mean	SD	Mean	SD
τ_{44}	3.99	3.99	0.288	3.99	0.288	3.99	0.362
τ_{45}	4.19	4.19	0.289	4.19	0.290	4.20	0.367
τ_{46}	4.59	4.60	0.317	4.60	0.321	4.60	0.372
τ_{55}	3.03	3.03	0.327	3.03	0.327	3.03	0.446
τ_{56}	3.62	3.63	0.358	3.63	0.369	3.63	0.430
τ_{66}	2.05	2.04	0.475	2.07	0.527	2.04	0.644

Conditional Mean Misspecified, PS Correctly Specified

	SATT	ETWFE		LW: IPWRA		CS: AIPW	
		Lags Only		Lags Only		Leads and Lags	
	Mean	Mean	SD	Mean	SD	Mean	SD
τ_{44}	4.75	4.75	0.606	4.72	0.606	4.75	0.809
τ_{45}	6.91	6.93	0.655	6.92	0.653	6.94	0.841
τ_{46}	9.53	9.54	0.741	9.54	0.744	9.54	0.896
τ_{55}	4.85	4.85	0.676	4.85	0.664	4.84	0.839
τ_{56}	5.39	5.45	0.684	5.45	0.693	5.44	0.835
τ_{66}	3.11	3.16	0.677	3.16	0.678	3.16	0.816

Conditional Mean and PS Misspecified

	SATT	ETWFE		LW: IPWRA		CS: AIPW	
		Lags Only		Lags Only		Leads and Lags	
	Mean	Mean	SD	Mean	SD	Mean	SD
τ_{44}	4.99	5.08	0.632	5.03	0.632	5.50	0.992
τ_{45}	7.15	7.48	0.711	7.32	0.709	8.06	1.008
τ_{46}	9.98	10.94	0.862	10.87	0.889	11.24	1.085
τ_{55}	5.37	5.68	0.762	5.53	0.740	5.77	1.007
τ_{56}	5.66	6.64	0.772	6.53	0.806	6.46	0.971
τ_{66}	3.39	4.32	0.776	4.16	0.802	3.81	0.984

6. Unit-Specific Trends in Rolling Methods

- Lee and Wooldridge (2026a, JBES): replace demeaning with detrending.
- Removing the pre-treatment average is the same as a very simple regression:

$$Y_{it} \text{ on } 1, t = 1, \dots, g - 1$$

$$\dot{Y}_{itg} \equiv Y_{it} - \bar{Y}_{i,pre(g)}, \text{ all } t$$

- Can extend this to unit-specific detrending:

$$Y_{it} \text{ on } 1, t, t = 1, \dots, g - 1$$

$$\ddot{Y}_{itg} = Y_{it} - \hat{Y}_{itg} = Y_{it} - (\hat{A}_{ig} + \hat{B}_{ig} \cdot t)$$

- The $\ddot{Y}_{itg}$ are out-of-sample residuals (prediction errors).
- For $t \geq g$, apply standard TE methods to

$$\{(\ddot{Y}_{itg}, D_{ig}, \mathbf{X}_i) : i = 1, \dots, N\}$$

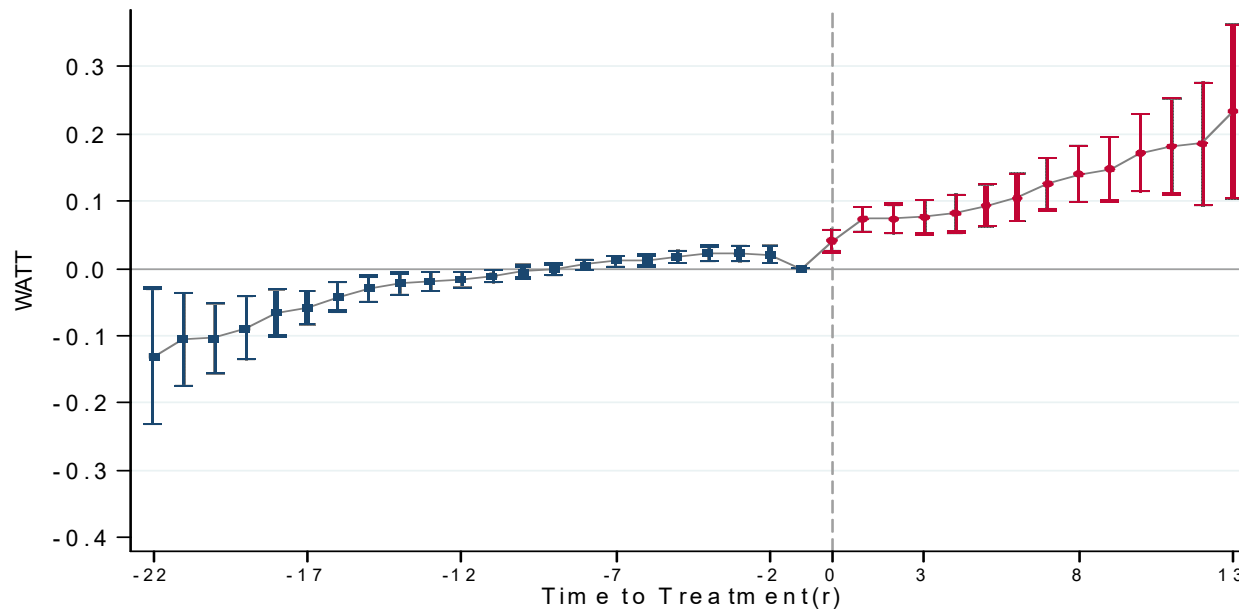
- ▶ Control group is

$$D_{i,t+1} + D_{i,t+2} + \dots + D_{iT} + D_{i\infty} = 1$$

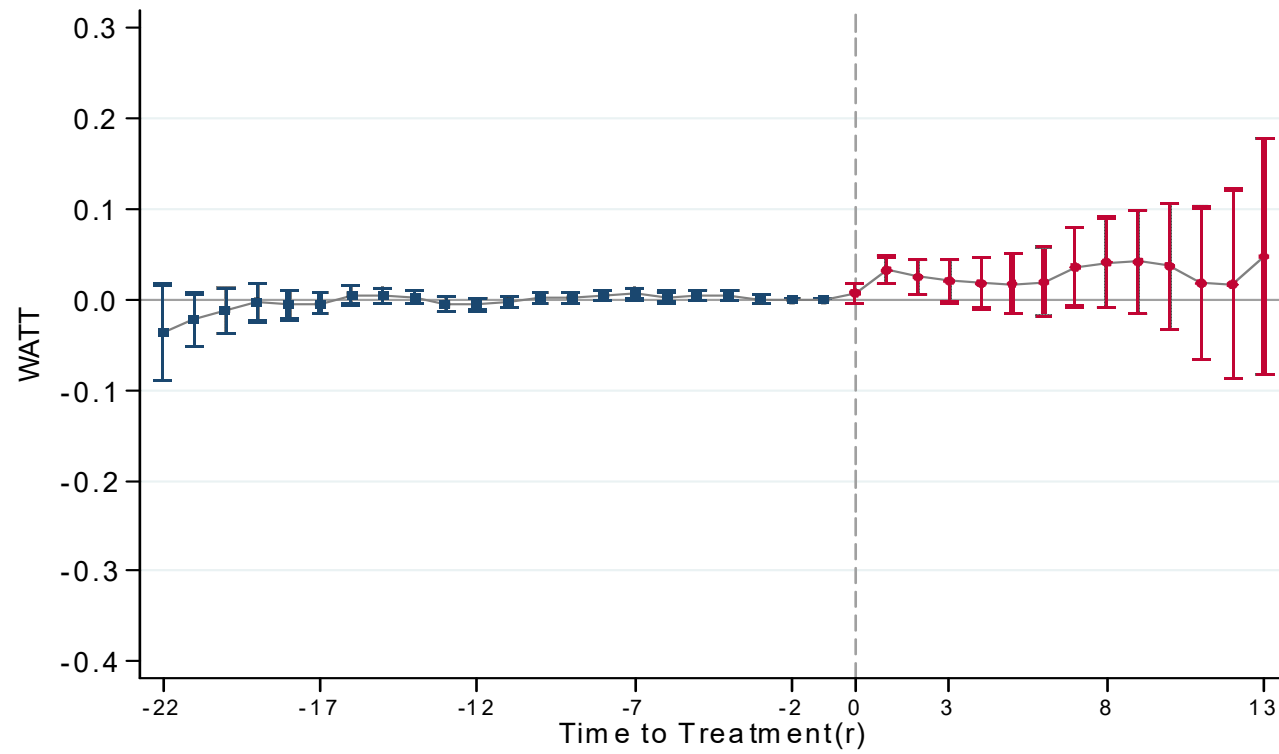
- ▶ Or a subset, such as never treated.
- Consistent when deviations from parallel trends are linear.
- Can be costly in terms of precision: the trends are collinear with the $D_{ig}f_s_t, s \geq g$.

- `lwdid` allows unit-specific detrending.
 - ▶ Use `rolling(detrend)`.
- Also allows deseasonalizing with quarterly or monthly data.
 - ▶ Use `rolling(dmeanq)` or `rolling(demeanm)` for removing unit-specific deseasonalizing.
 - ▶ Use `rolling(detrendq)` or `rolling(detrendm)` to detrend and deseasonalize.
- Can still apply IPWRA for a DR estimator after removing trends and/or seasonality.

- Applied to the Walmart sitings in U.S. counties.
 - ▶ Outcome is log of retail employment.
 - ▶ Confidence intervals are joint across all exposure-times.
- Demeaning followed by IPWRA:



- Detrending followed by IPWRA:



7. Small Sample Sizes

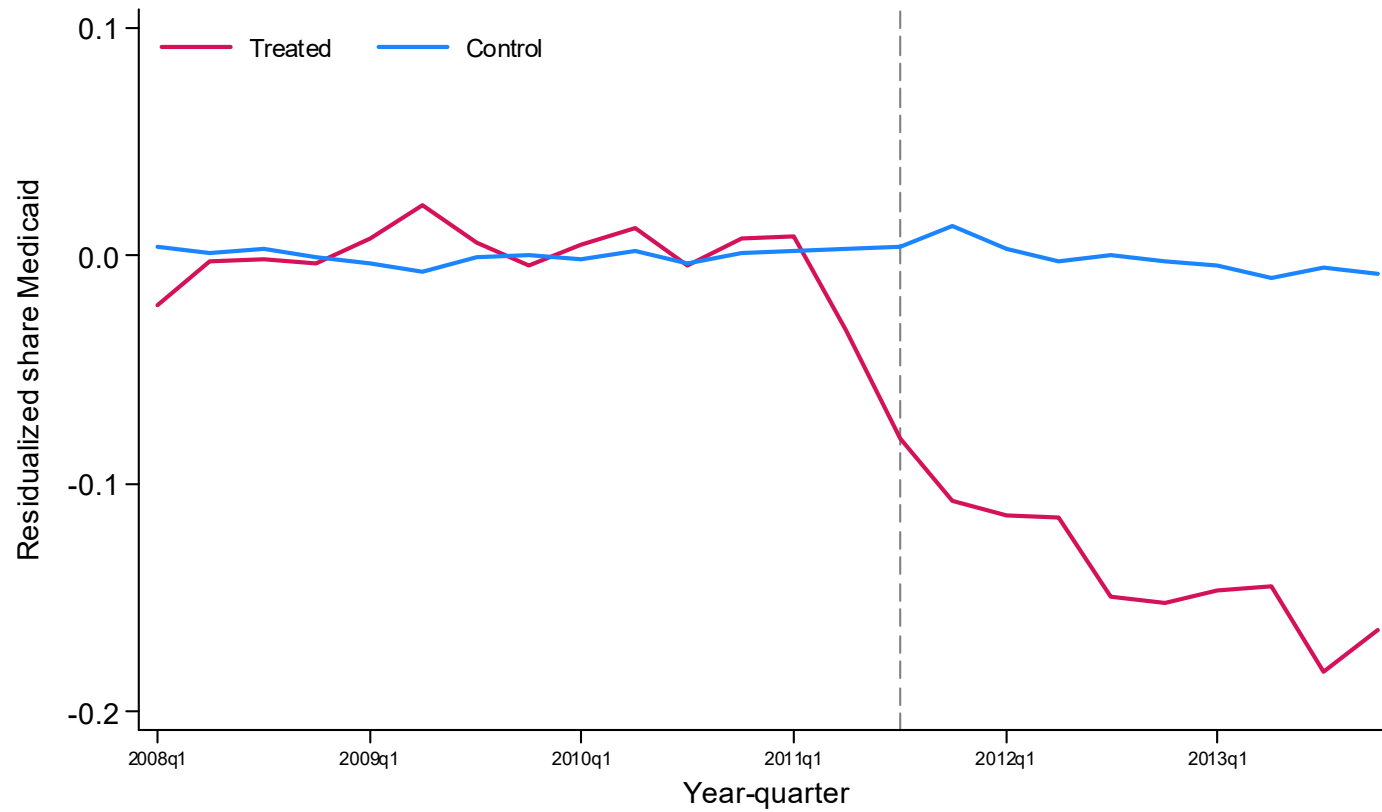
- Usual clustered standard errors cannot be used when some N_g (cluster sizes) are “small.”
- Common timing case: N_1 (treated), N_0 (control).
 - ▶ Sometimes have $N_1 = 1$.
- Synthetic Control, Synthetic DiD can be used.
 - ▶ But relies on enough “good” controls.
 - ▶ Inference relies on “large” T_0 and T_1 .

- Lee and Wooldridge (2026b) collapse the data to cross-sectional data; apply exact inference.
 - ▶ Can use usual DiD, unit-specific detrending, deasonalizing.
 - ▶ Only need $N_0 + N_1 \geq 3$, $T_0 \geq 1$, $T_1 \geq 1$.
 - ▶ With $T_0 \geq 2$, can use unit-specific detrending.
 - ▶ With quarterly or monthly data, can deasonalize.
 - ▶ Can combine deseasonalizing with detrending.

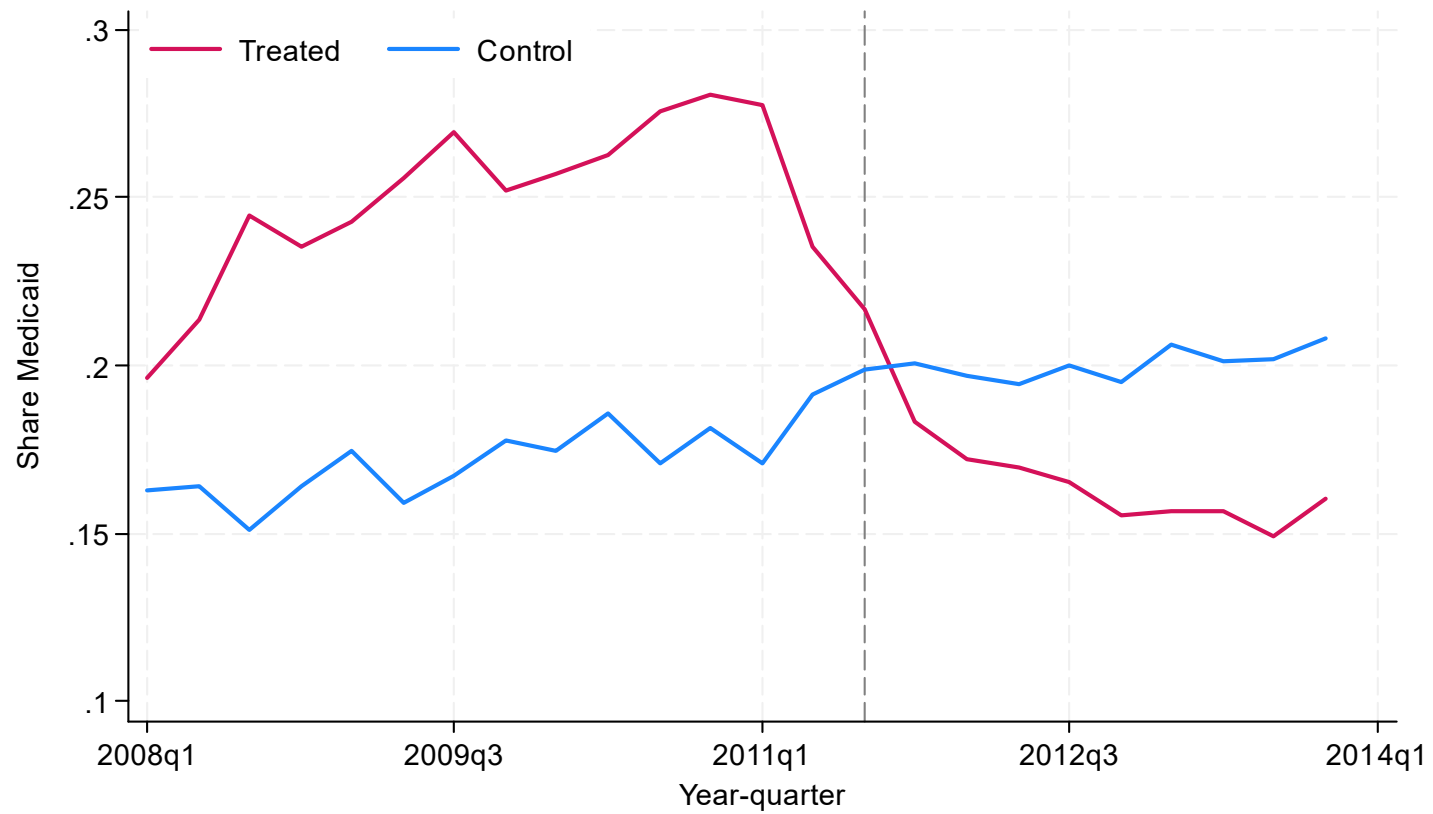
- Arizona Medicaid freeze: Hur, Lee, and Wooldridge (2026).
 - ▶ One treated state (AZ) ($N_1 = 1$) three control states (Arkansas, Maryland, and New Jersey) ($N_0 = 3$).
 - ▶ Quarterly data from 2008 to 2013.
 - ▶ Freeze implemented during 2011, Q3: $T_0 = 14$, $T_1 = 10$.
 - ▶ Account for different pre-trends and different seasonal patterns.
 - ▶ One outcome is the Medicaid participation rate.

```
lwdid y, ivar(state) tvar(year) gvar(cohort)
      rolling(detrendq) small
```

- State-specific detrending and deseasonalizing:



- Synthetic control:



- Estimated average effects on Medicaid coverage rate:
 - ▶ LW: -0.135 (0.011)
 - ▶ SC: -0.032 (0.043)
 - ▶ SDiD: -0.074 (0.016)

8. Summary

1. Time series transformations can be used with rolling methods (`lwdid`).

- ▶ Unit-specific demeaning, detrending, and even deseasonalizing using pre-treatment periods.

- ▶ Can apply standard TE estimators, including doubly robust AIPW and IPWRA, as well as matching.

- ▶ With high-dimensional covariates, can apply causal machine learning methods.

- ▶ With staggered entry, these do not use all pre-treatment information: only units not yet treated at time t contribute to the estimators.

2. Long-differencing methods proposed by Callaway and Sant'Anna (2021) are leads and lags (event study) estimators (`csdid`).

- ▶ Under “ideal” assumptions they are inefficient.
- ▶ Especially with strong, positive serial correlation, CS (2021) can be close to efficient (like first differencing).

3. `lwdid` can handle very small sample sizes and time periods if approximate normality is reasonable.

- ▶ With even moderate T_0, T_1 , the central limit theorem helps.