Doubly robust estimation with staggered adoption

1. Start with one cohort at one date

Prerequisites: Panel DiD and inference and cross-fitting. We now observe a panel at dates $1, \dots, \mathcal{T}$, with treatment beginning at different dates. Treatment is absorbing: once treated, a unit remains treated.

Let $G \in \{2, \dots, \mathcal{T}, \infty\}$ be the first treatment date. $G = \infty$ denotes never treated. A **cohort** is a group sharing the same first treatment date. Write $G_g = \mathbb{I}\{G = g\}$ for its indicator. Potential outcomes $Y_t(g)$ describe what would happen at date t if treatment began at g, and $Y_t(\infty)$ describes the never-treated path. The observation rule is $Y_t = Y_t(G)$.

The elementary target is

$$\mathrm{ATT}(g,t) = \mathbb{E}[Y_t(g) - Y_t(\infty) \mid G = g], \qquad t \geq g.$$

Fixing both indices matters. It permits effects to differ across cohorts and with time since adoption. No single regression coefficient is required to summarize those differences.

For a fixed pair (g,t), compare the outcome at t with the last untreated date $g-1$:

$$Z^{g,t} = Y_t - Y_{g-1}.$$

This is a **long difference**: it spans more than one period when $t > g$. A unit already treated at t cannot reveal an untreated comparison path merely because it belongs to another cohort. We must choose controls that are untreated at both endpoints.

We will use either never-treated controls, $C^{g,t} = \mathbb{I}\{G = \infty\}$, or not-yet-treated controls, $C^{g,t} = \mathbb{I}\{G > t\}$. For post-treatment cells $t \geq g$, both exclude cohort g.

The aim is to turn each valid pair into the two-period problem already solved, then aggregate its answer with clearly specified weights. This is the building-block approach of Callaway and Sant'Anna (2021).

2. Which assumptions justify this comparison?

No anticipation: $Y_s(g) = Y_s(\infty)$ for every $s < g$. In particular, the cohort's observed baseline satisfies

$$Y_{g-1} = Y_{g-1}(g) = Y_{g-1}(\infty) \quad \text{when } G = g.$$

Define the untreated long change $Z^{g,t}(\infty) = Y_t(\infty) - Y_{g-1}(\infty)$. On cohort g,

$$Z^{g,t} = Z^{g,t}(\infty) + \{Y_t(g) - Y_t(\infty)\}.$$

For selected controls, $G > t$ or $G = \infty$, both endpoints precede treatment. No anticipation and the observation rule therefore give $Z^{g,t} = Z^{g,t}(\infty)$ on $C^{g,t} = 1$.

Conditional parallel trends for this comparison:

$$\mathbb{E}[Z^{g,t}(\infty) \mid X, G = g] = \mathbb{E}[Z^{g,t}(\infty) \mid X, C^{g,t} = 1].$$

This is a separate identifying restriction for the chosen control group and long difference. Changing the comparison group changes the restriction. Covariates must be determined before treatment; we use a common baseline X throughout this note.

Define the population in which the propensity lives

Let $S^{g,t} = G_g + C^{g,t}$. Its value is one exactly for observations eligible for this comparison. Write

$$\rho = \mathbb{P}(S^{g,t} = 1), \qquad \mathbb{E}_Q[f] = \mathbb{E}[f \mid S^{g,t} = 1], \qquad \pi_Q = \mathbb{P}(G = g \mid S^{g,t} = 1).$$

Within this population define

$$p_{0,g,t}(X) = \mathbb{P}(G = g \mid X, S^{g,t} = 1), \qquad m_{0,g,t}(X) = \mathbb{E}[Z^{g,t} \mid X, C^{g,t} = 1].$$

Require $\rho > 0$, $\pi_Q > 0$, and $p_{0,g,t}(X) \leq 1 - \epsilon$ on the eligible population, with finite absolute outcome means. These conditions ensure a nonempty cohort and available controls on its support.

Never-treated comparisons have the same eligible population for every t within cohort g, so their propensity can be denoted $p_{0,g}$. Not-yet-treated populations change with t, so retaining both subscripts prevents an ambiguity.

3. Derive identification within the eligible population

For this page, suppress (g,t) on Z, m_0, and p_0. Within Q, set $D = G_g$; then $1 - D = C^{g,t}$. The assumptions on the previous page imply

$$\begin{aligned}\mathbb{E}[Z(\infty) \mid G = g] &= \mathbb{E}[\mathbb{E}[Z(\infty) \mid X, G = g] \mid G = g] \\ &= \mathbb{E}[\mathbb{E}[Z(\infty) \mid X, C^{g,t} = 1] \mid G = g] \\ &= \mathbb{E}[m_0(X) \mid G = g].\end{aligned}$$

Subtract this missing change from the cohort's observed change:

$$\mathrm{ATT}(g,t) = \mathbb{E}[Z - m_0(X) \mid G = g] = \frac{\mathbb{E}_Q[D(Z - m_0)]}{\pi_Q}.$$

For odds $o_0 = p_0/(1-p_0)$, conditioning inside Q gives

$$\mathbb{E}_Q[(1-D)o_0 Z \mid X] = (1-p_0)o_0 m_0 = p_0 m_0.$$

The weighted controls therefore recover the same imputed counterfactual. For candidates (m,p) the common-denominator DR construction is

$$T_Q(m,p) = \frac{\mathbb{E}_Q[\{D - (1-D)o\}(Z-m)]}{\pi_Q}, \qquad o = \frac{p}{1-p}.$$

The panel DiD calculation now yields

$$T_Q(m,p) - \mathrm{ATT}(g,t) = \frac{1}{\pi_Q}\mathbb{E}_Q\left[\frac{p_0 - p}{1-p}(m_0 - m)\right].$$

Every probability and expectation in this equation belongs to Q. Replacing π_Q by the full-population cohort share without changing the numerator would introduce a scale error.

Translate back to full-population notation

For any function on eligible observations, $\mathbb{E}[Sf] = \rho\mathbb{E}_Q[f]$. Also $q_g := \mathbb{P}(G = g) = \rho\pi_Q$. Thus

$$T_Q(m,p) = \frac{\mathbb{E}[\{G_g - C^{g,t}o\}(Z-m)]}{q_g}.$$

The full-population bias is $\mathbb{E}[S\{(p_0-p)/(1-p)\}(m_0-m)]/q_g$. The selection indicator is essential if the expectation uses the full population. Functions only defined inside Q can be assigned arbitrary values outside it; those values have zero contribution after selection.

4. Normalize each comparison and estimate it

Within Q, define $c_Q(p) = \mathbb{E}_Q[(1-D)o] > 0$. The normalized functional is

$$T_{H,Q}(m,p) = \mathbb{E}_Q\left[\left\{\frac{D}{\pi_Q} - \frac{(1-D)o}{c_Q(p)}\right\}(Z-m)\right].$$

Equivalently, in the full population,

$$\boxed{T_{H,Q}(m,p) = \mathbb{E}\left[\left\{\frac{G_g}{q_g} - \frac{C^{g,t}o}{\mathbb{E}[C^{g,t}o]}\right\}(Z-m)\right].}$$

The factors ρ cancel separately from each ratio.

If $m = m_0$, control residual means vanish. If $p = p_0$, $c_Q(p_0) = \pi_Q$ and conditional weight balance gives the correct IPW answer. These are the two robustness branches, applied inside this cell.

For completeness, its exact candidate bias follows the normalized panel formula:

$$a = \frac{p_0 - p}{1-p}, \quad e = m_0 - m, \quad b_Q = \frac{\mathbb{E}_Q[p_0 e]}{\pi_Q}, \qquad T_{H,Q} - \mathrm{ATT}(g,t) = \frac{\mathbb{E}_Q[a(e-b_Q)]}{c_Q(p)}.$$

This differs from the common-denominator product expression on the previous page.

The sample calculation

Select the eligible records. Estimate the propensity for cohort membership using all those records, and estimate m on their controls' long changes. Then compute

$$\widehat{\mathrm{ATT}}(g,t) = \frac{\sum_i G_{g,i}(Z_i^{g,t} - \hat{m}_i)}{\sum_i G_{g,i}} - \frac{\sum_i C_i^{g,t}\hat{o}_i(Z_i^{g,t} - \hat{m}_i)}{\sum_i C_i^{g,t}\hat{o}_i}.$$

Use held-out predictions when applying the cross-fitting inference theory. Split at the unit level so no unit's other dates enter its training sample. Use a common fold assignment across cells to retain a coherent vector of influence contributions.

For never-treated controls, the propensity fit can be reused across dates within cohort g; the regression still depends on the long difference. For not-yet-treated controls, changing t changes the propensity's conditioning population. Refitting or an explicitly justified model for those changing probabilities is required.

If there are no eligible controls, no treated records, or a zero weighted-control denominator, the cell is unavailable. Do not silently substitute a different comparison group or use already-treated units as untreated controls.

5. Work through the three-cohort example

Let X remain uniform on $\{0, 1, 2\}$ and let $G \in \{2, 3, \infty\}$:

x	$\mathbb{P}(G = 2 \mid x)$	$\mathbb{P}(G = 3 \mid x)$	$\mathbb{P}(G = \infty \mid x)$
0	1/8	1/2	3/8
1	1/4	1/2	1/4
2	3/8	1/2	1/8

Averaging columns gives cohort shares $(1/4, 1/2, 1/4)$. Bayes' rule gives $\mathbb{P}(X = x \mid G = 2) = (1/6, 1/3, 1/2)$; cohort 3 has uniform X.

Specify untreated outcomes by

$$Y_t(\infty) = a_G + (t-1)X^2 + U_t, \qquad \mathbb{E}[U_t \mid X, G] = 0,$$

where a_G permits permanent cohort differences. Let

$$Y_t(g) = Y_t(\infty) + \mathbb{1}\{t \geq g\}\, 2X.$$

This explicitly imposes no anticipation and gives every cohort the same conditional mean untreated trend, X^2, in every period. A restriction only on $\mathbb{E}[\Delta Y(\infty) \mid X]$ would not guarantee comparability across cohorts.

For any cell,

$$m_{0,g,t}(x) = (t - g + 1)x^2, \qquad \mathrm{ATT}(g, t) = \mathbb{E}[2X \mid G = g].$$

Consequently,

$$\mathrm{ATT}(2, 2) = \mathrm{ATT}(2, 3) = \frac{8}{3}, \qquad \mathrm{ATT}(3, 3) = 2.$$

For $(g, t) = (2, 2)$ with never-treated controls, the selection probability is $\mathbb{P}(S = 1 \mid x) = 1/2$ at every x. Thus X remains uniform inside Q, $\pi_Q = 1/2$, and

$$p^{\mathrm{nt}}_{0,2,2}(x) = \frac{\mathbb{P}(G = 2 \mid x)}{\mathbb{P}(G = 2 \mid x) + \mathbb{P}(G = \infty \mid x)} = (1/4, 1/2, 3/4).$$

This reproduces the panel example exactly. The cohort mean change is 5, the control mean change is 1, and adjustment gives $5 - 7/3 = 8/3$. For $(2, 3)$, the untreated long change doubles: the cohort's observed long mean is $14/3 + 8/3 = 22/3$, and its imputed untreated mean is $14/3$. Their difference is still $8/3$.

6. Change the controls and recompute the propensity

At $(2, 2)$, not-yet-treated controls include cohort 3 and never-treated units. Together with cohort 2 they form the whole population. Therefore

$$p^{\text{ny}}_{0,2,2}(x) = \mathbb{P}(G = 2 \mid x) = (1/8, 1/4, 3/8), \qquad o^{\text{ny}}_0 = (1/7, 1/3, 3/5).$$

The control share is $3/4$. Their covariate distribution is

$$\mathbb{P}(X = x \mid G > 2) = \frac{(1 - \mathbb{P}(G = 2 \mid x))/3}{3/4} = (7/18, 1/3, 5/18).$$

The weighted control mean long change is

$$\frac{(1/7) \cdot 0 \cdot (7/18) + (1/3) \cdot 1 \cdot (1/3) + (3/5) \cdot 4 \cdot (5/18)}{(1/7)(7/18) + (1/3)(1/3) + (3/5)(5/18)} = \frac{7/9}{1/3} = \frac{7}{3}.$$

The untreated counterfactual for cohort 2 is unchanged. Different controls and different odds recover the same target under their respective assumptions. Their ordinary control mean is $13/9$, so the unadjusted comparison is different.

At $(2, 3)$, cohort 3 has already adopted. Not-yet-treated controls now consist only of never-treated units, and the propensity returns to $(1/4, 1/2, 3/4)$. At $(3, 3)$ the eligible population is cohort 3 plus never treated, with

$$p_{0,3,3}(x) = \frac{1/2}{1/2 + \mathbb{P}(G = \infty \mid x)} = (4/7, 2/3, 4/5).$$

A propensity used for one cell is not automatically valid for another.

Feature	Never-treated controls	Not-yet-treated controls
Eligible controls	$G = \infty$	$G > t$
Composition over dates	Fixed	Changes as cohorts adopt
Propensity	One per cohort	Generally one per cell
Availability	Requires never-treated units	Requires units untreated at t

Adding eligible controls can improve information but does not guarantee smaller variance for every estimator. More importantly, it changes the parallel-trends assumption. In applications, justify that assumption for the chosen group rather than selecting controls solely to increase sample size.

7. Aggregate only after specifying the target weights

Let event time be $e = t - g$. For post-treatment event time $e \geq 0$, let $\mathcal{G}_e$ be the fixed set of cohorts with $g + e \leq \mathcal{T}$ and an identified cell. Define

$$H_e = \sum_{g \in \mathcal{G}_e} G_g, \qquad Q_e = \mathbb{E}[H_e], \qquad \omega_g(e) = \frac{q_g}{Q_e}.$$

The event-time ATT is

$$\boxed{\tau_e = \sum_{g \in \mathcal{G}_e} \omega_g(e) \mathrm{ATT}(g, g+e).}$$

The weights are nonnegative and sum to one. They describe the cohort composition of treated units eligible for this event time.

For the example, at $e = 0$ both cohorts contribute, with weights $1/3$ and $2/3$:

$$\tau_0 = \frac{1}{3}\frac{8}{3} + \frac{2}{3} \cdot 2 = \frac{20}{9}.$$

At $e = 1$, only cohort 2 is observed, so $\tau_1 = 8/3$. This increase does not show that effects grow with exposure: each cohort's effect is constant. The mix of cohorts changed. A balanced event-time summary uses a common eligible cohort set across the horizons being compared; here that would retain only cohort 2 for both $e = 0$ and $e = 1$.

An overall summary with equal weight on each unit's observed horizons

For each included cohort let $L_g = \mathcal{T} - g + 1$ and average its available post-treatment effects first:

$$\bar{\tau}_g = \frac{1}{L_g} \sum_{t=g}^{\mathcal{T}} \mathrm{ATT}(g, t).$$

Assuming all these cells are identified, define

$$\tau_{\text{overall}} = \sum_{g=2}^{\mathcal{T}} \frac{q_g}{\mathbb{P}(G < \infty)} \bar{\tau}_g.$$

This gives a cohort its treated-population share regardless of the number of post-treatment dates it contributes. Pooling all cells with equal weights would instead give early cohorts more influence. For the example, $\bar{\tau}_2 = 8/3$, $\bar{\tau}_3 = 2$, and $\tau_{\text{overall}} = 20/9$.

Double robustness in every contributing cell is sufficient for a consistent aggregate when the weights are consistently estimated. Biases can cancel across cells by accident, but aggregation supplies no general protection against a biased cell. Missing cells require an explicitly redefined target and eligible set.

8. Derive the influence contribution of estimated weights

Assume a fixed finite number of periods and cells, i.i.d. unit records, positive fixed cohort and selection shares, and the inference note's conditions within each cell. The normalized and common-denominator estimators then share the following full-population influence function when both nuisances are consistently estimated with negligible product remainder:

$$\phi_{g,t}(W) = \frac{G_g\{Z^{g,t} - m_{0,g,t}(X) - \mathrm{ATT}(g,t)\} - C^{g,t} o_{0,g,t}(X)\{Z^{g,t} - m_{0,g,t}(X)\}}{q_g}.$$

Indeed, the eligible-population influence function is scaled by π_Q. Multiplying it by S/ρ embeds it in the full population; the denominator becomes $\rho\pi_Q = q_g$. Its conditional mean inside Q is zero, so random changes in the eligible share introduce no additional first-order term.

Cohort shares are estimated by $\hat{q}_g = \mathbb{P}_n G_g$ and $\hat{Q}_e = \mathbb{P}_n H_e$. The ratio identity yields

$$\hat{\omega}_g - \omega_g = \frac{\mathbb{P}_n(G_g - \omega_g H_e)}{\hat{Q}_e}.$$

Thus the influence function of the weight is

$$\ell_g(W) = \frac{G_g - \omega_g H_e}{Q_e}.$$

Expand $\hat{\tau}_e = \sum_g \hat{\omega}_g \widehat{\mathrm{ATT}}(g, g+e)$:

$$\hat{\tau}_e - \tau_e = \sum_g \omega_g(\widehat{\mathrm{ATT}}_g - \mathrm{ATT}_g) + \sum_g \mathrm{ATT}_g(\hat{\omega}_g - \omega_g) + \sum_g (\hat{\omega}_g - \omega_g)(\widehat{\mathrm{ATT}}_g - \mathrm{ATT}_g).$$

Each product in the last sum is $O_p(n^{-1})$ under root-n convergence. For finitely many cells, that sum is $o_p(n^{-1/2})$. Hence

$$\boxed{\Phi_e(W) = \sum_{g \in \mathcal{G}_e} \omega_g \phi_{g,g+e}(W) + \frac{1}{Q_e} \sum_{g \in \mathcal{G}_e} G_g\{\mathrm{ATT}(g, g+e) - \tau_e\}.}$$

The second term is the effect of estimating the cohort mix. For the overall summary, replace each cell effect and influence function by its within-cohort time average, and use $H = \mathbb{I}\{G < \infty\}$ in the same calculation. Use unit-level sums of these contributions to retain the covariance across cells.

9. Inference for a path, and the limits of pre-trend checks

Estimate Φ_e by plugging held-out nuisance fits and sample shares into the previous page. With $\hat{V}_e = \mathbb{P}_n(\hat{\Phi}_e - \mathbb{P}_n\hat{\Phi}_e)^2$, the pointwise standard error is $\sqrt{\hat{V}_e/n}$. A separate 95% interval at each horizon does not give 95% simultaneous coverage of the whole path.

A **multiplier bootstrap** applies the same random multiplier to all influence contributions from one unit. For each repetition draw independent mean-zero, variance-one multipliers ξ_i and calculate

$$Z_e^* = \frac{1}{\sqrt{n}} \sum_i \xi_i(\hat{\Phi}_{e,i} - \mathbb{P}_n\hat{\Phi}_e), \qquad M^* = \max_e \frac{|Z_e^*|}{\sqrt{\hat{V}_e}}.$$

For nondegenerate variances, let $c_{.95}$ be the bootstrap 95th percentile of M^*. A simultaneous band is

$$\hat{\tau}_e \ \pm \ c_{.95}\sqrt{\hat{V}_e/n}.$$

Using one multiplier per unit preserves estimated dependence across horizons. Bootstrap validity requires joint asymptotic linearity and regularity conditions; Callaway and Sant'Anna establish the relevant results. For clustered sampling, use cluster influence sums and an appropriate cluster bootstrap theory.

What a pre-treatment comparison actually tests

For $s < g-1$, one can form $Y_s - Y_{g-1}$ and use controls untreated through $g-1$. No anticipation makes both dates untreated for cohort g. But the post-treatment parallel-trends assumptions in Section 2 do not restrict this pre-treatment long difference. A zero placebo additionally requires

$$\mathbb{E}[Y_s(\infty) - Y_{g-1}(\infty) \mid X, G = g] = \mathbb{E}[Y_s(\infty) - Y_{g-1}(\infty) \mid X, C = 1]$$

for the chosen pre-period controls. A rejection is evidence against the joint pre-period assumptions and fitted specification; it is not a direct observation of the treated group's missing post-treatment trend. Failure to reject does not establish post-treatment parallel trends.

Practical boundaries

Our asymptotics keep cohort shares positive as n grows. Very small cohorts can still produce poor finite-sample inference; an increasing number of periods or vanishing cohort shares needs additional theory. Check overlap per cell and avoid covariates affected by treatment. Pooling nuisance fits across cells adds modelling restrictions and must preserve the correct conditional population.

Further reading: Callaway and Sant'Anna. The supplementary TWFE and modern DiD slides discuss regression decompositions.