DiD extensions: repeated cross-sections and improved inference

1. Two extensions, and the extra assumptions they need

Prerequisites: Panel DiD and inference and cross-fitting. First we replace a panel by separate pre- and post-treatment samples. Then we choose particular nuisance fitting equations to obtain inference that remains valid when one working model is wrong. These are distinct extensions.

In repeated cross-sections, each record is $W = (Y, D, T, X)$, where $T = 0$ or 1 indicates its sampling period. We observe only one outcome per record, so we cannot form an individual $Y_1 - Y_0$. The group indicator D denotes membership in the group treated in period 1, including when sampled before treatment.

Assume independent draws from a mixture of two representative population samples, with $\lambda = \mathbb{P}(T = 1) \in (0, 1)$. Assume the distribution of (D, X) is stable across periods:

$$(D, X) \perp T.$$

This **stationarity** condition permits outcomes to change; it restricts group membership and covariate composition. Together with representative sampling, it allows the observed period-specific means to describe the same target population. Retain no anticipation, conditional parallel trends, $\pi = \mathbb{P}(D = 1) > 0$, and $p_0(X) = \mathbb{P}(D = 1 \mid X) \leq 1 - \epsilon$ from the panel note.

Write $r_{dt}(X) = \mathbb{E}[Y \mid D = d, T = t, X]$ for the true observed means and $m_{dt}(X)$ for candidate regressions. Identification gives

$$\tau = \mathbb{E}[(r_{11} - r_{10}) - (r_{01} - r_{00}) \mid D = 1].$$

To see this, parallel trends supplies $r_{01} - r_{00}$ as the treated group's untreated mean change within X. Its observed mean change is $r_{11} - r_{10}$. Subtract within X, then average over the stable treated composition.

Use $J_t = \mathbb{I}\{T = t\}$, $l_1 = \lambda$, $l_0 = 1 - \lambda$, and $s_1 = 1$, $s_0 = -1$ for period indicators, probabilities, and subtraction signs. Candidate odds are $o = p/(1 - p)$ and $c = \mathbb{E}[(1 - D)o] > 0$. Stationarity implies

$$\mathbb{E}[DJ_t] = \pi l_t, \qquad \mathbb{E}[(1 - D)J_t o] = cl_t.$$

These denominators will let us compare group-period means without pretending that individual changes are observed.

2. The basic repeated-cross-section DR formula

Define four normalized weights and a pooled treated weight:

$$w_{1t} = \frac{DJ_t}{\mathbb{E}[DJ_t]}, \qquad w_{0t} = \frac{(1-D)J_t o(X)}{\mathbb{E}[(1-D)J_t o(X)]}, \qquad w_1 = \frac{D}{\pi}.$$

Each of the four weights averages to one. Consider

$$\boxed{T_1 = \sum_{t=0}^{1} s_t \mathbb{E}[(w_{1t} - w_{0t})(Y - m_{0t}(X))].}$$

This subtracts the pre-period adjusted group gap from the post-period adjusted group gap. It needs two control regressions, one in each period.

Evaluate the formula by conditioning

Let $q(X) = (1 - p_0(X))o(X)$, so $c = \mathbb{E}[q]$, and define $e_t = r_{0t} - m_{0t}$, $e_\Delta = e_1 - e_0$. Stationarity and the conditional outcome means imply

$$\mathbb{E}[w_{1t}(Y - m_{0t})] = \frac{\mathbb{E}[p_0(r_{1t} - m_{0t})]}{\pi}, \qquad \mathbb{E}[w_{0t}(Y - m_{0t})] = \frac{\mathbb{E}[qe_t]}{c}.$$

For example, the treated numerator is $\mathbb{E}[DJ_t(Y - m_{0t})] = l_t\mathbb{E}[p_0(r_{1t} - m_{0t})]$; dividing by πl_t cancels the period sampling probability.

Since $r_{1t} - m_{0t} = (r_{1t} - r_{0t}) + e_t$, subtracting the two dates gives

$$T_1 = \tau + \frac{\mathbb{E}[p_0 e_\Delta]}{\pi} - \frac{\mathbb{E}[qe_\Delta]}{c}.$$

If the control change regression is correct, $e_\Delta = 0$, both extra terms vanish. Correctly specifying both control period means is a sufficient condition. If $p = p_0$, then $q = p_0$ and $c = \pi$, so the two terms cancel for any candidate control means. This proves both robustness branches.

For $a = (p_0 - p)/(1 - p)$ and $b = \mathbb{E}[p_0 e_\Delta]/\pi$, we also obtain

$$T_1 - \tau = \frac{\mathbb{E}[a(e_\Delta - b)]}{c}.$$

Indeed $q = p_0 - a$ and $c = \pi - \mathbb{E}[a]$, so this follows by exactly the normalized-bias calculation in the panel note. The formula specifies all terms; there is no unspecified period adjustment left over.

3. Add treated regressions without changing the population target

The two periods contain different treated people. Their empirical covariate compositions can differ even under population stationarity. We can use treated regressions to adjust those sample differences.

For $h_t(X) = m_{1t}(X) - m_{0t}(X)$, define

$$\boxed{T_2 = T_1 + \sum_{t=0}^{1} s_t \mathbb{E}[(w_1 - w_{1t})h_t(X)].}$$

The added term is explicit: the pooled treated mean of the modelled group gap minus its mean among treated observations sampled at that date. For any integrable function $h(X)$, stationarity gives

$$\mathbb{E}[w_{1t}h(X)] = \frac{l_t \mathbb{E}[Dh(X)]}{\pi l_t} = \mathbb{E}[w_1 h(X)].$$

Therefore every population adjustment is zero, even if the treated regressions are wrong. $T_2 = T_1$ for any admissible candidates, so T_2 inherits both robustness branches proved on the previous page.

An equivalent form that explains the extra models

For one period, combine the terms with w_{1t}:

$$w_{1t}(Y - m_{0t}) - w_{1t}(m_{1t} - m_{0t}) = w_{1t}(Y - m_{1t}).$$

Thus, writing $h = h_1 - h_0$,

$$T_2 = \mathbb{E}[w_1 h] + \sum_{t=0}^{1} s_t \mathbb{E}[w_{1t}(Y - m_{1t}) - w_{0t}(Y - m_{0t})].$$

The first term averages the modelled difference in changes over all treated observations. The remaining terms correct it using residuals in all four cells. With true period regressions, every residual has conditional mean zero.

Population equality of T_1 and T_2 does not imply that their sample analogues are equal. Their adjustments are sample differences between group means, which fluctuate. The adjustment changes the influence function and can use the information in stable treated composition more efficiently.

The two constructions correspond to the basic and augmented repeated-cross-section estimands in Sant'Anna and Zhao (2020), equations (2.8)–(2.9). We have used separate period notation to make every correction visible.

4. Sample formulas and a concrete check

Replace each expectation in the weights by its sample average. Define $\hat{w}_{1t,i} = D_i J_{t,i}/\mathbb{P}_n(DJ_t)$, $\hat{w}_{0t,i} = (1-D_i)J_{t,i}\hat{o}_i/\mathbb{P}_n((1-D)J_t\hat{o})$, and $\hat{w}_{1,i} = D_i/\mathbb{P}_n D$. Then

$$\hat{\tau}_1 = \sum_{t=0}^{1} s_t \mathbb{P}_n[(\hat{w}_{1t} - \hat{w}_{0t})(Y - \hat{m}_{0t})],$$

$$\hat{\tau}_2 = \hat{\tau}_1 + \sum_{t=0}^{1} s_t \mathbb{P}_n[(\hat{w}_1 - \hat{w}_{1t})(\hat{m}_{1t} - \hat{m}_{0t})].$$

Equivalently, evaluate the expanded expression from Section 3. These formulas require positive treated and weighted-control totals in both periods. They use the observed totals separately; replacing them by products of marginal sample shares defines a different finite-sample estimator.

Use the panel example as a population sampled twice

Let T be independent of (D, X) with $\lambda = 1/2$. At each sampled date, use the population from the panel note:

$$r_{00}(x) = 0, \quad r_{01}(x) = x^2, \quad r_{10}(x) = 5, \quad r_{11}(x) = 5 + x^2 + 2x.$$

Thus the adjusted conditional gap is 5 before treatment and $5 + 2x$ after. Their difference is $2x$. Averaging over treated shares $(1/6, 1/3, 1/2)$ gives

$$\tau = 0 \cdot \frac{1}{6} + 2 \cdot \frac{1}{3} + 4 \cdot \frac{1}{2} = \frac{8}{3}.$$

With true control regressions, the weighted control residuals are zero in both periods. The treated residual mean is 5 before and $5 + 8/3 = 23/3$ after. Hence $T_1 = 23/3 - 5 = 8/3$, regardless of the candidate propensity. With all four regressions true, T_2 instead expresses the answer as the treated mean of $h(X) = 2X$ plus zero residual corrections.

If $m_{00} = 0$, $m_{01} = -1/5 + 9X/5$, and $p = 1/3 + X/6$, the control change error is exactly the wrong-regression error in the panel example. Both normalized functionals give $151/55$, with bias $13/165$. Stationarity makes the population augmentation zero even under these wrong models.

This calculation uses separate representative samples of the same population; it never pairs pre- and post-period records to fabricate individual changes.

5. Derive influence functions for fixed nuisance candidates

The key tool is the derivative of a ratio of sample means. For fixed functions $a(W)$ and $b(W)$ with $\mathbb{E}[b] > 0$, let $R = \mathbb{E}[a]/\mathbb{E}[b]$. Then

$$\frac{\mathbb{P}_n a}{\mathbb{P}_n b} - R = \frac{\mathbb{P}_n(a - Rb)}{\mathbb{P}_n b}.$$

The first-order contribution is $(a-Rb)/\mathbb{E}[b]$. In particular, the influence of a normalized weighted residual mean is its weight times its residual minus its own weighted mean. This automatically accounts for estimated denominators.

For fixed candidates define

$$u_{dt}(W) = Y - m_{dt}(X), \qquad A_t = \mathbb{E}[w_{1t}(Y - m_{0t})], \quad B_t = \mathbb{E}[w_{0t}u_{0t}].$$

Applying the ratio rule separately to all four terms of T_1 gives

$$\phi_1^{\text{fixed}} = \sum_t s_t\{w_{1t}(Y - m_{0t} - A_t) - w_{0t}(u_{0t} - B_t)\}.$$

For T_2, let $H = \mathbb{E}[w_1 h]$ and $F_t = \mathbb{E}[w_{1t}u_{1t}]$. The expanded representation from Section 3 gives

$$\phi_2^{\text{fixed}} = w_1(h - H) + \sum_t s_t\{w_{1t}(u_{1t} - F_t) - w_{0t}(u_{0t} - B_t)\}.$$

Each term has mean zero by its definition. These formulas apply to fixed candidates. Estimated nuisance functions can contribute extra first-order terms; the next sections show how particular fitting equations eliminate them.

At the true period regressions and propensity

All residual means are zero and $H = \tau$. Therefore

$$\phi_{\text{rc}} = \frac{D}{\pi}(h_0(X) - \tau) + \sum_t s_t \left\{ \frac{DJ_t}{\pi l_t}(Y - r_{1t}) - \frac{(1-D)J_t o_0}{\pi l_t}(Y - r_{0t}) \right\},$$

where $h_0 = (r_{11} - r_{10}) - (r_{01} - r_{00})$. Sant'Anna and Zhao's efficiency theorem identifies this as the efficient influence function under their stationary repeated-cross-section model. The theorem is an external result; the ratio calculation derives the displayed estimator contribution. Using true control means alone in T_1 does not generally attain this bound.

6. Improved fitting: first see the idea in a panel

Return briefly to a panel with $Z = \Delta Y$. Let $z(X)$ be a fixed finite vector of regressors including an intercept. Specify logistic propensity and linear control-change candidates:

$$p_\gamma(X) = \frac{\exp(z'\gamma)}{1+\exp(z'\gamma)}, \qquad m_\beta(X) = z'\beta, \qquad o_\gamma(X) = \exp(z'\gamma).$$

These are working models; their probability limits may be wrong functions. We deliberately choose fitting criteria whose normal equations match the nuisance derivatives of the DR estimating equation.

Inverse probability tilting (IPT) minimizes

$$\mathbb{P}_n[(1-D)\exp(z'\gamma) - Dz'\gamma].$$

Differentiating with respect to γ gives the fitting equation

$$\mathbb{P}_n[(1-D)o_{\hat\gamma}z - Dz] = 0.$$

Thus odds-weighted controls match the treated regressor totals. Because z includes an intercept, they also match the treated count exactly.

Next fit the control regression by **weighted least squares (WLS)**:

$$\hat\beta = \arg\min_\beta \mathbb{P}_n[(1-D)o_{\hat\gamma}(Z - z'\beta)^2].$$

Differentiating and dividing by -2 yields

$$\mathbb{P}_n[(1-D)o_{\hat\gamma}z(Z - z'\hat\beta)] = 0.$$

The intercept component says the weighted control residual total is zero. The resulting normalized DR estimate therefore equals the treated mean residual. IPT balance also makes the linear regression correction cancel in its augmented-IPW form. These equalities concern the specially chosen joint fits, not arbitrary OR and IPW estimators.

Assume finite interior solutions, nonsingular population derivative matrices, sufficient finite moments, and regular root-n convergence of the fixed-dimensional fits to (γ^*, β^*). For example, bounded regressors and outcomes, compact parameter neighborhoods, overlap and nonsingular design matrices provide simple sufficient local conditions. Separation or singular regressor matrices can prevent a fit; exact balance is not guaranteed to be feasible in every sample.

7. Why these panel fits remove first-order nuisance effects

The common-denominator estimating equation is

$$H(W;\tau,\beta,\gamma)=\{D-(1-D)o_\gamma\}(Z-z'\beta)-D\tau.$$

At the probability limits of the fits, their population equations are

$$\mathbb{E}[(D-(1-D)o_{\gamma^*})z]=0,\qquad \mathbb{E}[(1-D)o_{\gamma^*}z(Z-z'\beta^*)]=0.$$

Differentiate the population ATT equation with the distribution held fixed:

$$\partial_\beta \mathbb{E}H=-\mathbb{E}[\{D-(1-D)o_\gamma\}z],$$

$$\partial_\gamma \mathbb{E}H=-\mathbb{E}[(1-D)o_\gamma z(Z-z'\beta)].$$

Both derivatives vanish at (β^*,γ^*) by the fitting equations, even if one or both working models are wrong.

Let τ^* be the probability limit of the resulting DR estimate. A first-order expansion of its sample equation gives

$$0=\mathbb{P}_nH(W;\tau^*,\beta^*,\gamma^*)-\pi(\hat\tau-\tau^*)+0'(\hat\beta-\beta^*)+0'(\hat\gamma-\gamma^*)+o_p(n^{-1/2}).$$

Here root-n regularity and smooth finite-dimensional equations make the remaining terms negligible. Solving yields the influence contribution

$$\boxed{\phi_{\mathrm{imp,p}}(W)=\frac{\{D-(1-D)o_{\gamma^*}\}(Z-z'\beta^*)-D\tau^*}{\pi}.}$$

If either the propensity model or the control-change model is correctly specified, the corresponding population fit recovers that true function under the stated identification of its parameters. Double robustness then gives $\tau^*=\tau$. The same variance formula is valid under either branch. This is the meaning of **double robustness for inference** in this construction.

IPT's intercept equation equates the two sample weight totals, so the normalized and common-denominator versions agree exactly for these fits. The WLS intercept also makes the control residual mean zero; hence the ratio-derived influence function has precisely the displayed form. Estimate its variance by its empirical second moment and divide by n for the estimator variance.

Ordinary logit maximum likelihood instead solves $\mathbb{P}_n[(D-p_\gamma)z]=0$; ordinary control OLS omits the odds weights. Their equations do not generally cancel the derivatives above. They can still support valid inference if their nuisance-estimation contributions are included, for example in a joint sandwich variance. Ignoring those contributions is the error to avoid.

8. Extend the cancellation to repeated cross-sections

Use the same pooled IPT propensity fit as in Section 6, now on records (Y, D, T, X). Fit separate linear control regressions $m_{0t} = z'\beta_{0t}$ by WLS within each period:

$$\mathbb{P}_n[(1-D)J_t o_{\hat{\gamma}} z(Y - z'\hat{\beta}_{0t})] = 0, \qquad t = 0, 1.$$

For the augmented estimator, fit treated regressions $m_{1t} = z'\beta_{1t}$ by ordinary least squares in each treated-period sample:

$$\mathbb{P}_n[DJ_t z(Y - z'\hat{\beta}_{1t})] = 0.$$

Assume the same regularity conditions as for the panel, with all four cell shares positive. In what follows, evaluate weights and regressions at their probability limits; stars are suppressed to make the derivatives readable.

Stationarity and the population IPT equations imply

$$\mathbb{E}[w_{1t}z] = \mathbb{E}[w_1 z] = \frac{\mathbb{E}[Dz]}{\pi}, \qquad \mathbb{E}[w_{0t}z] = \frac{\mathbb{E}[(1-D)oz]}{c} = \frac{\mathbb{E}[Dz]}{\pi}.$$

The last step uses both the regressor-balance equations and their intercept component $c = \pi$. Period balance follows in the population from stationarity; it need not hold exactly in each finite-sample period.

For the basic functional, $\partial_{\beta_{0t}} T_1 = -s_t \mathbb{E}[(w_{1t} - w_{0t})z] = 0$. For the augmented functional in Section 3,

$$\partial_{\beta_{0t}} T_2 = -s_t \mathbb{E}[(w_1 - w_{0t})z] = 0, \qquad \partial_{\beta_{1t}} T_2 = s_t \mathbb{E}[(w_1 - w_{1t})z] = 0.$$

The first derivative uses balance; the second uses stationarity alone.

For a normalized control residual mean $B_t = \mathbb{E}[w_{0t}u_{0t}]$, the quotient rule and $\partial_\gamma o = oz$ give

$$\partial_\gamma B_t = \mathbb{E}[w_{0t} z u_{0t}] - B_t \mathbb{E}[w_{0t} z].$$

WLS sets the first term to zero; its intercept sets $B_t = 0$. Thus the propensity derivative of both DR functionals is also zero. These calculations show exactly which equations remove each nuisance contribution at possibly wrong limits. The treated OLS fits are convenient, but it is stationarity, not OLS correctness, that removes their first-order estimation effects.

9. The resulting repeated-cross-section standard errors

The derivatives on the previous page vanish. Under regular root-n convergence, the fitted estimators therefore have the fixed-candidate influence functions from Section 5, evaluated at the limiting candidates.

For the basic estimator, WLS makes $B_t = 0$. Its influence function becomes

$$\phi_{1,\text{imp}} = \sum_t s_t\{w_{1t}(Y - m_{0t} - A_t) - w_{0t}(Y - m_{0t})\},$$

where $A_t = \mathbb{E}[w_{1t}(Y - m_{0t})]$ and $\tau^* = A_1 - A_0$. Each treated-period ratio must be centered by its own A_t. Subtracting a single pooled $D\tau^*/\pi$ would not account for all four estimated denominators.

For the augmented estimator, the intercept equations from treated OLS and control WLS give $F_t = B_t = 0$. Thus $H = \tau^*$ and

$$\boxed{\phi_{2,\text{imp}} = w_1(h - \tau^*) + \sum_t s_t\{w_{1t}(Y - m_{1t}) - w_{0t}(Y - m_{0t})\}.}$$

These are influence functions at the working-model limits, which may differ from the true period means. If either the propensity model is correct or both control period models are correct, the population DR proof gives $\tau^* = \tau$. Correctness of both controls is a simple sufficient OR branch; Section 2's identification result also permits some errors that cancel in their difference. Correct treated models are not needed for DR consistency or the cancellation of their estimation effects.

Calculate the standard error

For the basic version, replace A_t by its treated-period sample residual mean. For the augmented version, replace h by the fitted difference of differences and τ^* by $\hat{\tau}_2$. In both, use the actual sample weight normalizers and fitted functions to obtain $\hat{\phi}_{j,i}$. Define

$$\hat{V}_j = \frac{1}{n}\sum_i(\hat{\phi}_{j,i} - \bar{\hat{\phi}}_j)^2, \qquad \widehat{\text{se}}(\hat{\tau}_j) = \sqrt{\hat{V}_j/n}.$$

The empirical mean is zero in exact arithmetic for the stated fits; centering also accommodates numerical fitting tolerance. Under the regularity and consistency branch just stated, normal intervals based on these standard errors have asymptotic validity. If all period regressions and propensity are correct, the augmented influence function reduces to ϕ_{rc} in Section 5 and attains the associated efficiency bound. The basic estimator need not.

10. How the two extensions fit into the series

The repeated-cross-section problem changes what is observed. We replace an individual change by four group-period means, state population stationarity, and normalize each observed group-period contribution. The extra treated-regression adjustment exploits stationarity while keeping its population expectation zero.

The improved-inference problem changes how the auxiliary functions are fitted. IPT balances regressor totals, WLS makes weighted control residual moments zero, and those equations cancel the derivatives that would otherwise transmit nuisance estimation error at first order. They supply a specific parametric route to valid inference under either correctness branch.

Keep the inference routes distinct

Route	Fits and assumptions	Result justified
Cross-fitting in the inference note	Both functions consistent; negligible product rate; moment and overlap conditions	Normal inference with the true-model influence function
Improved fitting here	Fixed-dimensional IPT and WLS; regular fits; one correctness branch	Normal inference using the displayed working-limit influence function
Generic parametric fitting	Regular joint estimating equations with nuisance effects retained	Sandwich inference accounting for the fitted nuisance parameters

Do not substitute arbitrary machine-learning predictions into the improved parametric formulas and assume the cancellations still hold. Conversely, sample splitting alone does not remove first-order error under a wrong limiting nuisance. For flexible repeated-cross-section estimation, a separate cross-fitting proof must handle the full score and its period regressions.

What to inspect in an application

Confirm that the samples represent a stable distribution of (D, X) across time; changing survey composition can violate this restriction. Ensure all four groups are observed with usable overlap. Check that the numerical fits satisfy their balance and residual equations to tolerance. A failed or nonunique fit does not justify silently switching estimation methods while retaining its variance formula.

When both working models are wrong, these procedures may still estimate their own population limit regularly, but that limit need not be the ATT. Neither a small standard error nor exact sample regressor balance establishes parallel trends, stationarity, or adequate overlap in the population.

Source and further reading: Sant'Anna and Zhao (2020), Doubly robust difference-in-differences estimators, Sections 2 and 3, especially equations (2.8)–(2.9) and Theorems 2–3. Their efficiency and general asymptotic theorems supply the external results identified above; the conditional-expectation, ratio, and derivative calculations here explain the constructions step by step.