

# Formulae

A reference for the main formulae in the completed weeks. Each section links to the notes for the derivations and examples. See [Greek letters](#) for the symbols and [Python](#) for estimation code.

Expectations refer to population quantities. Hats denote sample estimates, and bars denote sample means. The meaning of a symbol follows the model in its section.

## Background: Population quantities and sample analogues

The **analogy principle** replaces population expectations with sample averages. For observations  $(X_i, Y_i)$ ,  $i = 1, \dots, n$ , define  $\mu_X = E[X_i]$  and  $\mu_Y = E[Y_i]$ . The sample column below uses  $1/n$  throughout.

| Quantity           | Population                                            | Sample analogue                                                                |
|--------------------|-------------------------------------------------------|--------------------------------------------------------------------------------|
| Mean               | $\mu_X = E[X_i]$                                      | $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$                                       |
| Variance           | $\sigma_X^2 = E[(X_i - \mu_X)^2]$                     | $\hat{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$                |
| Standard deviation | $\sigma_X = \sqrt{\sigma_X^2}$                        | $\hat{\sigma}_X = \sqrt{\hat{\sigma}_X^2}$                                     |
| Covariance         | $\sigma_{X,Y} = E[(X_i - \mu_X)(Y_i - \mu_Y)]$        | $\hat{\sigma}_{X,Y} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$ |
| Correlation        | $\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$ | $r_{X,Y} = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X \hat{\sigma}_Y}$           |

Variance and covariance require finite second moments. Correlation also requires positive variances for both variables.

### The correction from $n$ to $n - 1$

For independent, identically distributed observations and  $n > 1$ , the usual unbiased estimators of population variance and covariance are

$$s_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} \hat{\sigma}_X^2,$$

$$s_{X,Y} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \frac{n}{n-1} \hat{\sigma}_{X,Y}.$$

Estimating the means from the same observations makes the  $1/n$  variance estimator downward biased and multiplies the expected covariance by  $(n-1)/n$ . The correction above removes that bias. Taking a square root gives the usual sample standard deviation  $s_X$ , although  $s_X$  itself is generally biased for  $\sigma_X$ .

The common correction cancels in the covariance-to-variance ratio used for the OLS slope:

$$\frac{s_{X,Y}}{s_X^2} = \frac{\hat{\sigma}_{X,Y}}{\hat{\sigma}_X^2}.$$

It also cancels in the sample correlation when the same convention is used for covariance and both variances.

# Week 1: Selection bias and experiments

## Week 1 notes

### Potential outcomes and selection bias

For binary treatment  $D_i$ , the observed outcome is

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0.$$

The average treatment effect and the average treatment effect on the treated are

$$ATE = E[Y_i^1 - Y_i^0], \quad ATT = E[Y_i^1 - Y_i^0 | D_i = 1].$$

The observed difference in population means decomposes as

$$E[Y_i | D_i = 1] - E[Y_i | D_i = 0] = ATT + \underbrace{E[Y_i^0 | D_i = 1] - E[Y_i^0 | D_i = 0]}_{\text{selection bias}}.$$

Random assignment of treatment makes the selection-bias term zero and makes ATT equal ATE. This interpretation assumes well-defined treatments and no interference between individuals.

### OLS and omitted-variable bias

With an intercept and no additional controls, the population short-regression slope and its sample analogue are:

| Method    | Population                                                 | Sample estimator                                                                           |
|-----------|------------------------------------------------------------|--------------------------------------------------------------------------------------------|
| OLS slope | $\beta_1^S = \frac{\text{Cov}(D_i, Y_i)}{\text{Var}(D_i)}$ | $\hat{\beta}_1^S = \frac{\sum_i (D_i - \bar{D})(Y_i - \bar{Y})}{\sum_i (D_i - \bar{D})^2}$ |

Each denominator requires variation in the explanatory variable.

With an intercept and one explanatory variable,

$$\hat{\beta}_1 = \frac{\sum_i (D_i - \bar{D})(Y_i - \bar{Y})}{\sum_i (D_i - \bar{D})^2}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{D}.$$

For binary  $D_i$ , the slope is  $\bar{Y}_{D=1} - \bar{Y}_{D=0}$ .

Suppose the true population relationship and the auxiliary regression are

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 A_i + \varepsilon_i, \quad E[\varepsilon_i | D_i, A_i] = 0,$$

$$A_i = \gamma_0 + \gamma_1 D_i + v_i, \quad \gamma_1 = \frac{\text{Cov}(D_i, A_i)}{\text{Var}(D_i)}.$$

Omitting ability  $A_i$  gives the population short-regression slope

$$\beta_1^S = \beta_1 + \beta_2 \gamma_1.$$

The superscript  $S$  denotes the short regression. Its difference from the causal slope is  $\beta_2 \gamma_1$ .

## Power and sample size

For a specified alternative, power is  $1 - \beta$ , where  $\beta$  is the probability of a Type II error. With independent observations, common outcome variance  $\sigma^2$ , and a two-sided test at level  $\alpha$ , the normal approximation gives

$$\text{MDE} = (z_{1-\alpha/2} + z_{1-\beta}) \sqrt{\frac{\sigma^2}{n_T} + \frac{\sigma^2}{n_C}}.$$

Here  $z_p$  is the  $p$  quantile of the standard normal distribution. With equal treatment and control arms, the required size **per arm** to detect an effect  $\Delta$  is

$$n = 2 (z_{1-\alpha/2} + z_{1-\beta})^2 \frac{\sigma^2}{\Delta^2}.$$

## Week 2: Matching and weighting

---

### Week 2 notes

### Propensity scores

The propensity score is the probability of treatment conditional on observed pre-treatment characteristics:

$$p(X_i) = P(D_i = 1 \mid X_i).$$

For a single explanatory variable, the logit model is

$$p(X_i) = \frac{1}{1 + \exp(-(\gamma_0 + \gamma_1 X_i))}.$$

To identify ATT using untreated comparisons, we assume that conditional on  $X_i$ , untreated potential outcomes have the same mean for treated and untreated individuals. We also need controls with comparable characteristics throughout the support of the treated group.

### Exact matching

Within a covariate group  $x$ ,

$$\hat{\tau}(x) = \bar{Y}_{D=1, X=x} - \bar{Y}_{D=0, X=x}.$$

Weight these differences by the distribution of characteristics among treated observations:

$$\hat{\tau}_{\text{ATT}} = \sum_x \frac{N_{1x}}{N_1} \hat{\tau}(x).$$

Here  $N_{1x}$  is the number of treated observations in group  $x$  and  $N_1$  is the total number treated.

## Inverse-probability weighting for ATT

Using estimated propensity scores  $\hat{p}_i$ , assign weights

$$w_i = \begin{cases} 1 & D_i = 1, \\ \hat{p}_i / (1 - \hat{p}_i) & D_i = 0. \end{cases}$$

The normalised weighted difference in means is

$$\hat{\tau}_{\text{ATT}} = \frac{1}{N_1} \sum_{i:D_i=1} Y_i - \frac{\sum_{i:D_i=0} w_i Y_i}{\sum_{i:D_i=0} w_i}.$$

It is also the treatment coefficient from weighted least squares with an intercept and treatment indicator only:

$$\min_{b_0, b_1} \sum_i w_i (Y_i - b_0 - b_1 D_i)^2.$$

OLS uses the same objective with all weights equal to one. Very large ATT weights signal that a few controls may dominate the comparison.

## Weeks 3–4: Instrumental variables and LATE

Week 3 notes · Week 4 notes

### OLS bias: population slope and finite-sample expectation

In the omitted-ability model, the population short-regression slope differs from the causal effect by

$$\beta_1^S - \beta_1 = \beta_2 \gamma_1, \quad \gamma_1 = \frac{\text{Cov}(D_i, A_i)}{\text{Var}(D_i)}.$$

The **finite-sample bias** of the OLS estimator relative to the causal effect is

$$E[\hat{\beta}_1^S] - \beta_1.$$

To connect the two, [Week 3](#) writes

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 A_i + \nu_i$$

and conditions on all the observed schooling values,  $\mathbf{D} = (D_1, \dots, D_n)$ . Assume

$$E[\nu_i | \mathbf{D}] = 0, \quad E[A_i | \mathbf{D}] = \gamma_0 + \gamma_1 D_i.$$

For samples with variation in  $D_i$ , these assumptions imply

$$E[\hat{\beta}_1^S | \mathbf{D}] - \beta_1 = \beta_2 \gamma_1.$$

Averaging over possible schooling samples gives  $E[\hat{\beta}_1^S] - \beta_1 = \beta_2\gamma_1$ , provided the expectation exists. The linear conditional-mean assumption is what allows this population OVB expression to describe the finite-sample bias as well.

## Wald and the IV estimator

In  $Y_i = \beta_0 + \beta_1 D_i + u_i$ , instrument validity requires orthogonality,  $\text{Cov}(Z_i, u_i) = 0$ , and relevance,  $\text{Cov}(Z_i, D_i) \neq 0$ .

For a binary instrument, Wald divides the difference in conditional outcome means by the difference in conditional treatment means. The sample estimator replaces population expectations with sample means. In the constant-effect model, orthogonality and relevance identify the causal slope through the following ratios.

| Method                                | Population                                                                  | Sample estimator                                                                                                  |
|---------------------------------------|-----------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|
| Wald, binary instrument               | $\beta_1 = \frac{E[Y_i Z_i=1] - E[Y_i Z_i=0]}{E[D_i Z_i=1] - E[D_i Z_i=0]}$ | $\hat{\beta}_{1,\text{Wald}} = \frac{\bar{Y}_{Z=1} - \bar{Y}_{Z=0}}{\bar{D}_{Z=1} - \bar{D}_{Z=0}}$               |
| Exactly identified IV, one instrument | $\beta_1 = \frac{\text{Cov}(Z_i, Y_i)}{\text{Cov}(Z_i, D_i)}$               | $\hat{\beta}_{1,\text{IV}} = \frac{\sum_i (Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_i (Z_i - \bar{Z})(D_i - \bar{D})}$ |

Each denominator must be nonzero. With a binary instrument, the Wald and IV sample formulas give the same estimate. With heterogeneous effects and binary treatment, the assumptions in the [LATE section](#) give the population ratio a complier-effect interpretation.

## First stage, reduced form and 2SLS

With exogenous controls  $X_{ik}$ ,

$$D_i = \pi_0 + \pi_1 Z_i + \sum_{k=1}^K \pi_{k+1} X_{ik} + v_i,$$

$$Y_i = \delta_0 + \delta_1 Z_i + \sum_{k=1}^K \delta_{k+1} X_{ik} + e_i.$$

With one excluded instrument and one endogenous explanatory variable, the model is exactly identified:

$$\hat{\beta}_{1,\text{IV}} = \frac{\hat{\delta}_1}{\hat{\pi}_1}.$$

Use the same observations, controls and weights in both regressions. Two-stage least squares first predicts  $D_i$  from all the exogenous variables, then regresses  $Y_i$  on the predicted  $D_i$  and the controls. The [two-stage procedure below](#) writes out both stages explicitly and also applies when there is just one excluded instrument. Use an IV routine for the standard errors.

## Overidentified IV: the two-stage procedure

With one endogenous explanatory variable and more than one excluded instrument, the model is overidentified. Two-stage least squares proceeds as follows:

1. Regress  $D_i$  on all the excluded instruments and all the exogenous controls, including an intercept. Obtain  $\widehat{D}_i$ .
2. Regress  $Y_i$  on  $\widehat{D}_i$  and the same exogenous controls, including an intercept.

The fitted first stage with  $J$  excluded instruments and  $K$  exogenous controls is

$$\widehat{D}_i = \hat{\pi}_0 + \sum_{j=1}^J \hat{\pi}_j Z_{ij} + \sum_{k=1}^K \hat{\pi}_{J+k} X_{ik}.$$

The estimated **second-stage regression** is

$$Y_i = \hat{\beta}_{0,2SLS} + \hat{\beta}_{1,2SLS} \widehat{D}_i + \sum_{k=1}^K \hat{\beta}_{k+1,2SLS} X_{ik} + \hat{r}_i.$$

Here  $\hat{r}_i$  is the second-stage regression residual. For IV standard errors, the outcome-equation residual is computed using observed  $D_i$  in place of  $\widehat{D}_i$ .

Use an IV routine to estimate the model and obtain the appropriate standard errors. Report the excluded instruments and their joint first-stage test. The ratio of a single reduced-form coefficient to a single first-stage coefficient applies to the exactly identified case above.

## LATE

For binary treatment and instrument, let  $D_i^1$  and  $D_i^0$  be treatment status under the two instrument values. Under independence, exclusion, relevance and monotonicity, with well-defined treatments and no interference,

$$\frac{E[Y_i | Z_i = 1] - E[Y_i | Z_i = 0]}{E[D_i | Z_i = 1] - E[D_i | Z_i = 0]} = E[Y_i^1 - Y_i^0 | D_i^1 > D_i^0] = \tau_{\text{LATE}}.$$

The expectation on the right is the average treatment effect for compliers. If assignment is random only conditional on specified characteristics, the analysis must account for those characteristics.

## Leave-one-out leniency

For a simple illustration, let  $j(i)$  identify the official assigned to individual  $i$ . A leave-one-out treatment rate is

$$Z_i = \frac{\sum_{\ell \neq i: j(\ell)=j(i)} D_\ell}{n_{j(i)} - 1}.$$

Here  $n_j$  counts cases assigned to official  $j$ . The papers refine this construction to account for the assignment process and repeated cases. Leaving out the individual's own treatment decision avoids a mechanical relationship between their treatment and the instrument. Instrument validity also depends on the assignment process, exclusion and monotonicity.

## Week 5: Weak instruments

### Week 5 notes

This section follows BJB's notation:  $x_i$  is schooling,  $y_i$  is log earnings, and  $\epsilon_i$  is the outcome disturbance.

## Unbiasedness and consistency

Unbiasedness concerns the mean across repeated samples of a given size:

$$E[\hat{\beta}] = \beta.$$

Consistency means that for every fixed  $a > 0$ ,

$$P\left(\left|\hat{\beta} - \beta\right| > a\right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

We write  $\text{plim } \hat{\beta} = \beta$ . A probability limit can also be a value other than the true parameter.

## Consequences of endogeneity and imperfect instruments

For the simple model and under the moment and sampling conditions discussed in the notes,

$$\text{plim } \hat{\beta}_{OLS} - \beta = \frac{\sigma_{x,\epsilon}}{\sigma_x^2},$$

$$\text{plim } \hat{\beta}_{IV} - \beta = \frac{\sigma_{z,\epsilon}}{\sigma_{z,x}} = \frac{\rho_{z,\epsilon}}{\rho_{z,x}} \frac{\sigma_\epsilon}{\sigma_x}.$$

The IV expression assumes fixed, nonzero population relevance. Small correlation between the instrument and schooling can magnify a violation of orthogonality.

## First-stage diagnostics

For a first-stage regression with an intercept,

$$R^2 = 1 - \frac{\sum_i \hat{v}_i^2}{\sum_i (x_i - \bar{x})^2}.$$

To test  $q$  excluded instruments jointly, the conventional homoskedastic first-stage statistic is

$$F = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n - k_U)}.$$

$SSR_R$  is the residual sum of squares when the excluded instruments are omitted,  $SSR_U$  includes them, and  $k_U$  counts all coefficients in the unrestricted first stage, including the intercept. Both regressions contain the same controls. Robust or clustered tests require the corresponding variance calculation.

The approximation discussed in BJB connects relative finite-sample bias to first-stage strength:

$$\frac{\text{Bias}(\hat{\beta}_{IV})}{\text{Bias}(\hat{\beta}_{OLS})} \approx \frac{1}{F}.$$

This is an approximation under particular conditions. The reported first-stage  $F$  is a diagnostic of instrument strength, and does not establish the exclusion restriction.

## Week 7: Regression discontinuity

---

Week 7 notes

Let  $X_i$  be the running variable,  $x_0$  the cutoff, and  $R_i = X_i - x_0$ . This section uses  $\rho$  for the treatment effect at the cutoff.

## Sharp RD

When  $D_i = 1 (X_i \geq x_0)$  and both conditional potential-outcome means are continuous at the cutoff,

$$\begin{aligned}\rho &= E[Y_i^1 - Y_i^0 \mid X_i = x_0] \\ &= \lim_{x \downarrow x_0} E[Y_i \mid X_i = x] - \lim_{x \uparrow x_0} E[Y_i \mid X_i = x].\end{aligned}$$

The first limit approaches from above and the second from below.

A local linear specification with different slopes on the two sides is

$$Y_i = \beta_0 + \rho D_i + \beta_1 R_i + \gamma_1 D_i \cdot R_i + \epsilon_i.$$

At the cutoff  $R_i = 0$ , the jump is  $\rho$ . A bandwidth  $h$  determines how close observations must be to the cutoff. Triangular weights give greater weight to closer observations:

$$w_i = \max\left(0, 1 - \frac{|R_i|}{h}\right).$$

This expression uses a common bandwidth on the two sides. Software can select different bandwidths for the two sides and for bias correction.

## Fuzzy RD

When eligibility changes treatment probability at the cutoff, the fuzzy RD estimand is

$$\rho_{FRD} = \frac{\lim_{x \downarrow x_0} E[Y_i \mid X_i = x] - \lim_{x \uparrow x_0} E[Y_i \mid X_i = x]}{\lim_{x \downarrow x_0} E[D_i \mid X_i = x] - \lim_{x \uparrow x_0} E[D_i \mid X_i = x]}.$$

Under continuity, exclusion, monotonicity and a nonzero first-stage jump, this identifies the treatment effect for compliers at the cutoff. Retain the well-defined-treatment and no-interference assumptions.

## Estimating fuzzy RD with IV

Instrument actual treatment  $D_i$  with the cutoff indicator

$$Z_i = 1 (X_i \geq x_0) = 1 (R_i \geq 0).$$

The instrument is a **deterministic function of the running variable**. It records eligibility, while  $D_i$  records actual treatment. Identification comes from continuity at the cutoff and the discontinuous change in treatment probability induced by eligibility.

Using observations within the chosen bandwidth, the local linear first stage is

$$D_i = \pi_0 + \pi_1 Z_i + \pi_2 R_i + \pi_3 Z_i \cdot R_i + v_i.$$

Obtain predicted treatment from this regression:

$$\widehat{D}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i + \hat{\pi}_2 R_i + \hat{\pi}_3 Z_i \cdot R_i.$$

The estimated second stage is

$$Y_i = \hat{\beta}_0 + \hat{\rho} \widehat{D}_i + \hat{\beta}_1 R_i + \hat{\gamma}_1 Z_i \cdot R_i + \hat{r}_i.$$

The coefficient  $\hat{\rho}$  is the fuzzy RD estimate. Include  $R_i$  and  $Z_i \cdot R_i$  as controls in both stages to allow different slopes on the two sides. The cutoff indicator  $Z_i$  is the excluded instrument. Use the same observations and any kernel weights in both stages.

This specification is exactly identified. Its IV estimate equals the ratio of the estimated reduced-form jump to the estimated first-stage jump, using the same local specification. Use an IV routine for standard errors, based on the outcome-equation residual using observed  $D_i$ . Report the first-stage jump alongside the IV estimate.

## Density continuity

The manipulation test examines

$$H_0 : \lim_{x \uparrow x_0} f_X(x) = \lim_{x \downarrow x_0} f_X(x).$$

Here  $f_X$  is the density of the running variable. This is a diagnostic for sorting around the cutoff. The identifying continuity assumption concerns the conditional means of the potential outcomes.

## Week 8: Basic difference-in-differences

---

**Week 8 notes.** Let  $T_i$  identify the treatment group and  $Post_t$  the post-policy period.

### Parallel trends and the effect

$$E[Y_{i,post}^0 - Y_{i,pre}^0 | T_i = 1] = E[Y_{i,post}^0 - Y_{i,pre}^0 | T_i = 0].$$

With no anticipation, parallel trends identifies the post-policy ATT from the difference of changes. Its sample analogue is

$$\hat{\delta} = (\bar{Y}_{T,post} - \bar{Y}_{T,pre}) - (\bar{Y}_{C,post} - \bar{Y}_{C,pre}).$$

### Levels and differences

$$Y_{it} = \alpha + \gamma T_i + \tau Post_t + \delta(T_i \cdot Post_t) + u_{it}.$$

For a two-period panel, differencing gives

$$\Delta Y_i = \tau + \delta T_i + \Delta u_i.$$

### Baseline covariates

Allow baseline covariates  $X_{ik}$  to predict both levels and changes:

$$Y_{it} = \alpha + \gamma T_i + \tau Post_t + \delta(T_i \cdot Post_t) + \sum_{k=1}^K \beta_k X_{ik} + \sum_{k=1}^K \theta_k (X_{ik} \cdot Post_t) + u_{it}.$$

$$\Delta Y_i = \tau + \delta T_i + \sum_{k=1}^K \theta_k X_{ik} + \Delta u_i.$$

The  $\beta_k$  terms cancel because the baseline covariates are constant across the two observations. The  $\theta_k$  terms remain. Causal interpretation requires an appropriate conditional parallel-trends assumption and the regression specification to represent those conditional changes adequately.

## Week 9: TWFE and event studies

---

Week 9 notes.

### Two-way fixed effects

$$Y_{it} = \alpha_i + \lambda_t + \delta D_{it} + u_{it}.$$

$\alpha_i$  captures time-invariant unit differences,  $\lambda_t$  captures common calendar-year effects, and  $D_{it}$  indicates treatment in that unit-year.

### Calendar time and event time

If unit  $i$  adopts in calendar year  $G_i$ , its event time is  $k = t - G_i$ . Event time zero is the adoption year. Treatment remains on in subsequent years, while the event-time indicator changes each year.

For the common-adoption-date presentation in the notes,  $G_i = G$  for treated units:

$$Y_{it} = \alpha_i + \lambda_t + \sum_{\substack{k=-K \\ k \neq -1}}^L \beta_k (T_i \cdot \mathbf{1}(t - G = k)) + u_{it}.$$

Year  $-1$  is the reference period. The indicators cover the observed treated event times, with endpoint bins if specified.  $\lambda_t$  continues to refer to calendar time. With staggered adoption, replace  $G$  by  $G_i$  for treated units and set the indicators to zero for never-treated units. Interpreting the resulting TWFE coefficients requires care when effects differ across adoption cohorts or time since treatment.

## Week 10: Modern difference-in-differences

---

Week 10 notes.

### Effects by adoption cohort and calendar time

Let  $G_i = g$  mean that unit  $i$  first receives treatment in year  $g$ . Write  $G_i = \infty$  for never-treated units. For  $t \geq g$ ,

$$ATT(g, t) = E[Y_{it}^g - Y_{it}^\infty \mid G_i = g].$$

$Y_{it}^g$  is the outcome if treatment starts in year  $g$ , and  $Y_{it}^\infty$  is the outcome if the unit remains untreated. Thus  $ATT(2004, 2007)$  is the effect in 2007 for the cohort first treated in 2004.

## Identification and sample analogue

Using the last pre-treatment year  $g - 1$  as the baseline, parallel trends with never-treated controls requires

$$E[Y_{it}^\infty - Y_{i,g-1}^\infty | G_i = g] = E[Y_{it}^\infty - Y_{i,g-1}^\infty | G_i = \infty].$$

With no anticipation, this gives

$$ATT(g, t) = E[Y_{it} - Y_{i,g-1} | G_i = g] - E[Y_{it} - Y_{i,g-1} | G_i = \infty].$$

For a panel containing the same units in both periods, the sample analogue is

$$\widehat{ATT}(g, t) = (\bar{Y}_{g,t} - \bar{Y}_{g,g-1}) - (\bar{Y}_{\infty,t} - \bar{Y}_{\infty,g-1}).$$

The first subscript on each mean identifies the adoption cohort, and the second identifies calendar time. Not-yet-treated units with  $G_i > t$  can also supply the comparison under the corresponding parallel-trends and no-anticipation assumptions.

## Averaging at a common event time

At event time  $k = t - g \geq 0$ ,

$$\theta_k = \sum_{g \in \mathcal{G}_k} w_{g,k} ATT(g, g + k), \quad w_{g,k} \geq 0, \quad \sum_{g \in \mathcal{G}_k} w_{g,k} = 1.$$

$\mathcal{G}_k$  contains cohorts observed at event time  $k$  with a valid comparison group. With cohort-size weights, larger cohorts receive more weight. Changes across event times can reflect both changing effects and changes in which cohorts contribute.

## Week 11: Synthetic controls

[Week 11 notes](#) · [Python implementation](#).

### Counterfactual and estimated effect

Unit 1 receives treatment. Units  $j = 2, \dots, J + 1$  form the untreated donor pool. Treatment begins in period  $T_0 + 1$ .

$$\widehat{Y}_{1t}^0 = \sum_{j=2}^{J+1} w_j Y_{jt}, \quad w_j \geq 0, \quad \sum_{j=2}^{J+1} w_j = 1.$$

The donor weights are chosen using pre-treatment information and then held fixed across time. For post-treatment periods,

$$\hat{\tau}_{1t} = Y_{1t} - \widehat{Y}_{1t}^0.$$

Interpreting this gap as a causal effect requires the synthetic path to approximate the treated unit's outcome without the intervention, including the absence of spillovers affecting the donor outcomes.

### Pre- and post-treatment fit

Let  $g_{it}$  be unit  $i$ 's observed outcome minus its synthetic outcome. Apply the same construction to the treated unit and the placebo units. With  $T_1 = T - T_0$ ,

$$\text{MSPE}_i^{\text{pre}} = \frac{1}{T_0} \sum_{t=1}^{T_0} g_{it}^2, \quad \text{MSPE}_i^{\text{post}} = \frac{1}{T_1} \sum_{t=T_0+1}^T g_{it}^2.$$

$$R_i = \frac{\text{MSPE}_i^{\text{post}}}{\text{MSPE}_i^{\text{pre}}}.$$

A larger ratio indicates a greater deterioration in fit after treatment relative to before treatment. The ratio requires a positive pre-treatment MSPE. For the German exercise, the summary windows are 1960–1989 and 1991–2003, omitting 1990, with the denominators adjusted to the number of years in each window.

### Placebo rank

$$p = \frac{\text{number of units with } R_i \geq R_1}{J + 1}.$$

Include the treated unit in both counts. The smallest possible rank fraction is  $1/(J + 1)$ . An exact randomization interpretation requires an appropriate treatment-assignment mechanism. In the application, the rank summarizes how unusual the treated unit's result is within the chosen comparison set.