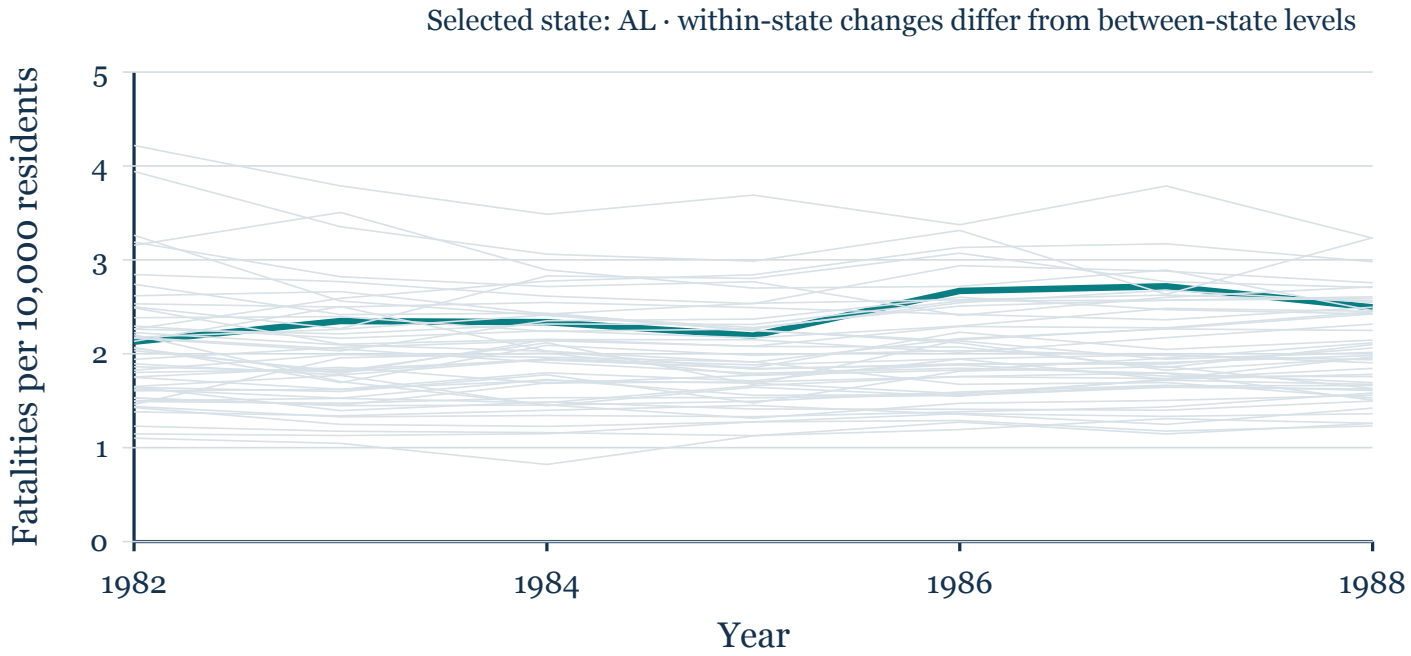
Can repeated observations improve a policy comparison?

Use changes within states to study alcohol taxes and traffic fatalities.



Observed state fatality rates, 1982–1988. The control selects a state; pale paths show the remaining states.

A panel has identities as well as dates

Repeated cross sections do not give the same within-unit comparison.

> One row is a state in a year.

> Before estimating, verify that the state-year pair uniquely identifies rows and inspect the coverage.

Write a stable omitted effect explicitly

An unobserved unit component may be related to policy.

> $Y_{it} = \beta X_{it} + \alpha_i + \lambda_t + u_{it}$

> α_i: stable state component; λ_t: common year shift.

> u_{it}: remaining state-year error.

Follow one unit across two dates

Subtracting the same unit's outcomes removes its stable intercept.

> $(X_{i1}, X_{i2}) = (2, 4), \quad (Y_{i1}, Y_{i2}) = (10, 14)$

> $\Delta X_i = 2, \quad \Delta Y_i = 4, \quad \alpha_i - \alpha_i = 0$

First differencing removes a stable unit component

Subtract the earlier equation from the later equation.

> $$\Delta Y_i = \beta \Delta X_i + \Delta \lambda + \Delta u_i$$

Differencing can create dependence even from independent level errors

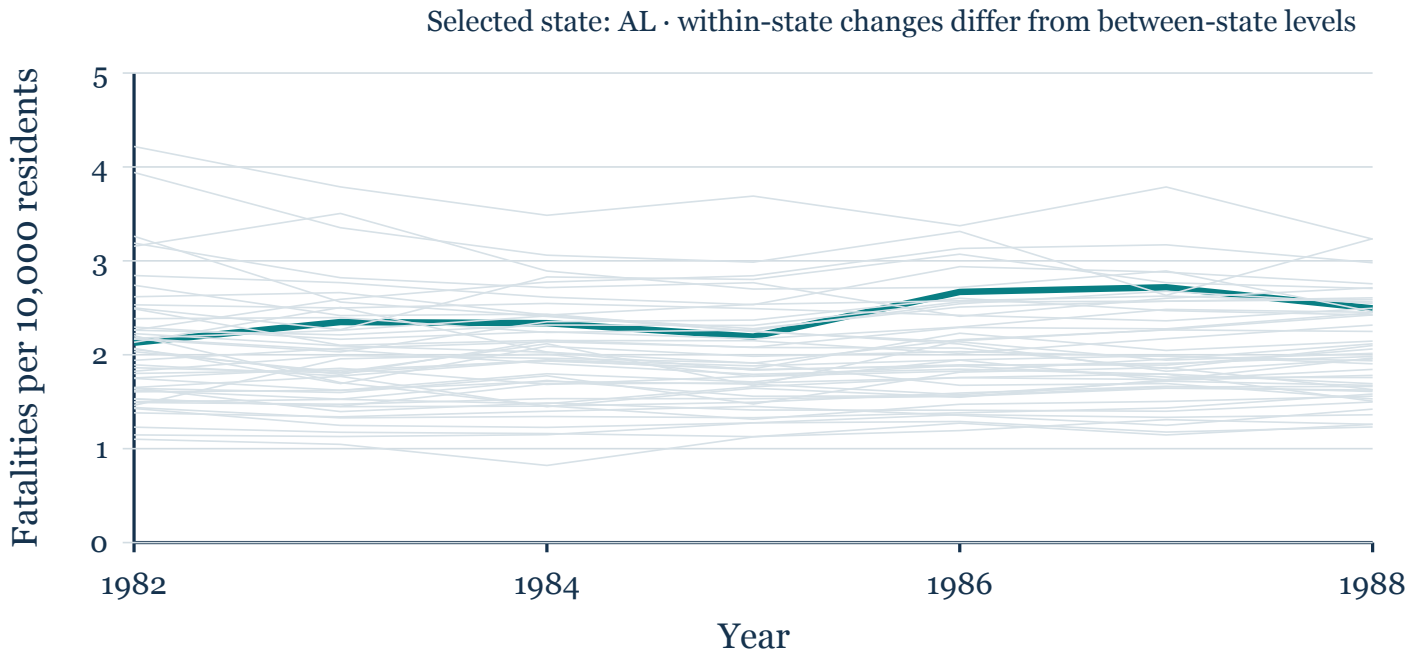
Adjacent differences share one error with opposite signs.

> $\Delta u_{it} = u_{it} - u_{i,t-1}$; if level errors are independent with variance σ^2, $\operatorname{Var}(\Delta u_{it}) = 2\sigma^2$.

> $\operatorname{Cov}(\Delta u_{it}, \Delta u_{i,t-1}) = -\sigma^2$; adjacent correlation $= -1/2$.

Demeaning uses each unit's own average

Fixed effects remove the same stable component over several periods.



Observed state fatality rates, 1982–1988. The control selects a state; pale paths show the remaining states.

$X_i = (2, 4) \Rightarrow X_i - \overline{X}_i = (-1, 1)$

$Y_i = (10, 14) \Rightarrow Y_i - \overline{Y}_i = (-2, 2)$

$Y_{it} - \overline{Y}_i = \beta(X_{it} - \overline{X}_i) + (\lambda_t - \overline{\lambda}) + (u_{it} - \overline{u}_i)$

Two-period fixed effects and first differences share a slope identity

The identity requires the same two-period sample and specification.

> With two observations per unit, demeaning produces one half of the change and its negative.

> Our application compares a two-period change from 1982 to 1988 with a fixed-effects fit using all seven years.

Read pooled and within-state associations

Removing stable differences can change the sign.

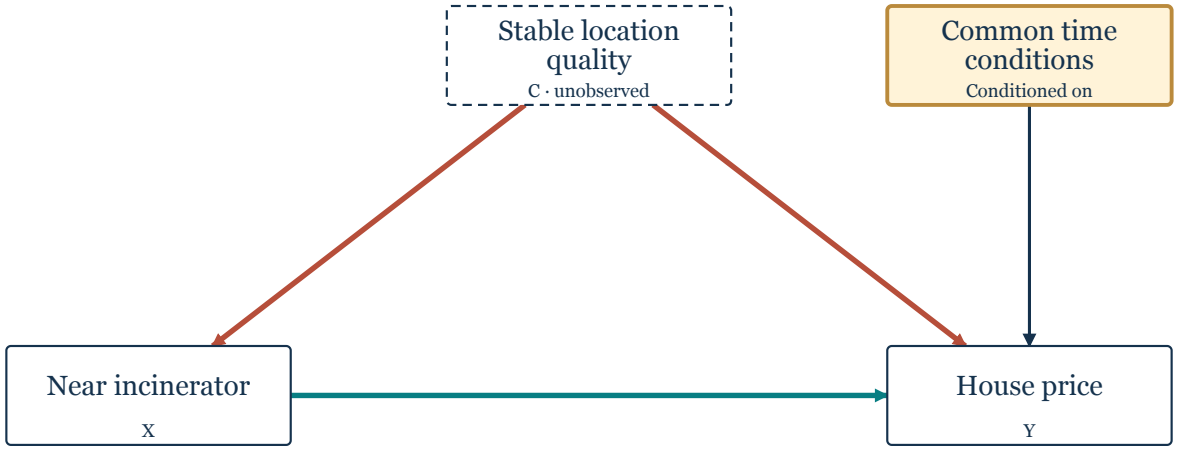
```
. xtreg fatality beertax i.year, fe vce(cluster state)

------------------------------------------------------------------------------
             |               Robust
    fatality | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     beertax |  -.6399799   .3570783    -1.79   0.080    -1.358329     .078369
             |
       _cons |    2.42847   .2016884    12.04   0.000     2.022725    2.834215
-------------+----------------------------------------------------------------
```

Calendar-indicator coefficient rows are omitted here. The output download contains the full numerical results.

Use year effects for common shocks

They do not remove state-specific changes.



1 open directed route · 1 open alternative · gold: conditioned on

Assumed causal structure · dashed boxes: unobserved · gold boxes: conditioned on

> Year effects: common date shifts. State effects: stable additive state differences.

> A location common cause is not removed by a common calendar control.

Check whether fixed effects absorb a regressor

A time-only variable cannot survive unrestricted time indicators.

> If $Tax_t = 1 + 1\{t = 2\}$, its column duplicates year-intercept variation.

> $Tax_t \times near_i$ can vary across groups within a year.

> A separate coefficient needs variation beyond the included controls.

State the fixed-effects exogeneity condition

Transforming variables does not make their remaining changes random.

> $E(u_{it} \mid X_{i1}, \ldots, X_{iT}, \alpha_i) = 0$ as a sufficient panel condition

Build a difference-in-differences comparison

Begin with the Head Start policy question from the original DiD lesson.

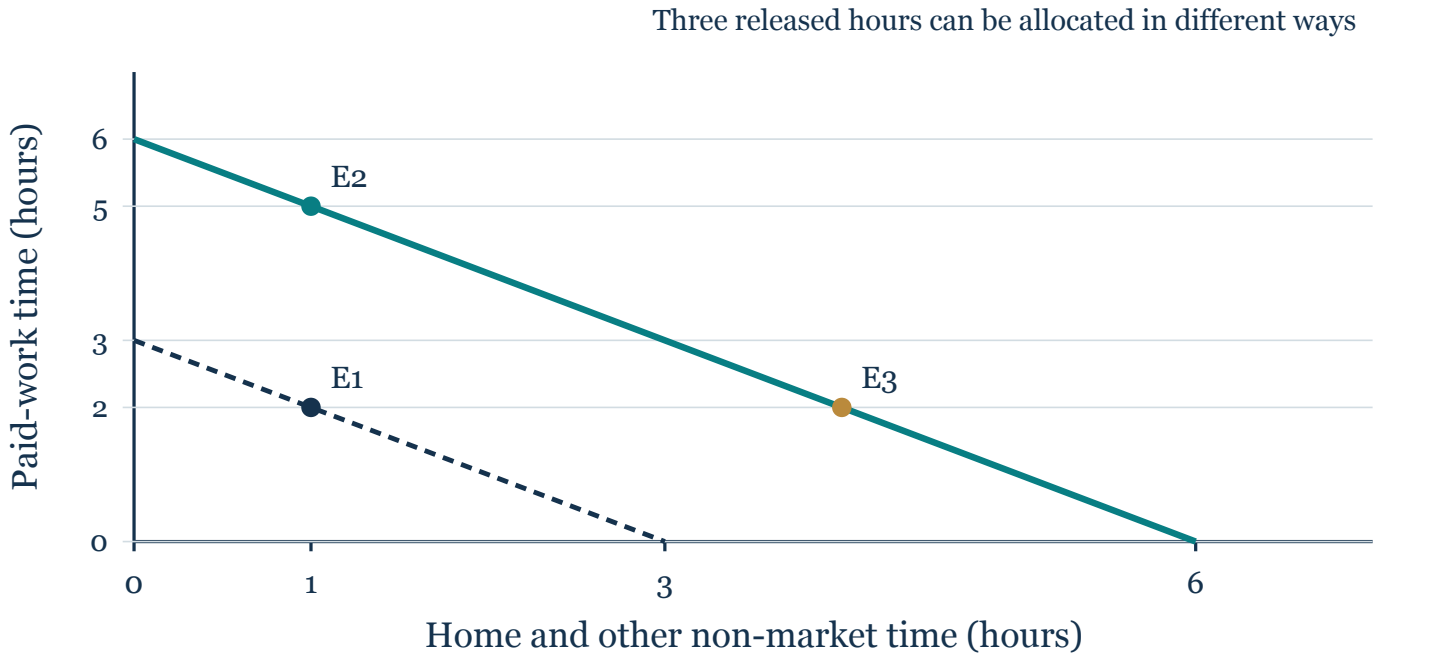
> Policy question: preschool access and maternal employment.

> Define eligibility before the expansion, the comparison population and the two dates.

> Reading: Wikle and Wilson (2023), Access to Head Start and Maternal Labor Supply.

Preschool access changes a household's feasible choices

Released care time can support paid work or another activity.

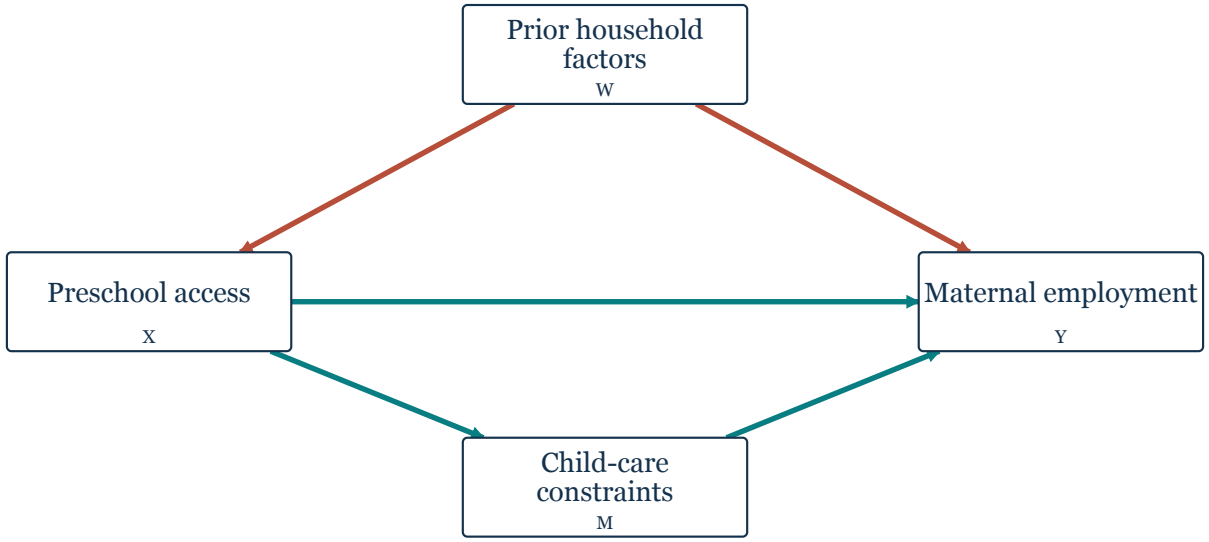


Constructed time-allocation diagram adapted from Qiu's Head Start lesson. Points illustrate feasible choices, not estimated responses.

- Axes: home/other time and paid-work time.
- Preschool can relax a care constraint without forcing a particular response.
- The diagram is a constructed time-allocation illustration.

Eligibility, access and participation are different variables

State the policy exposure whose effect is being studied.



2 open directed routes · 1 open alternative · gold: conditioned on

Assumed causal structure · dashed boxes: unobserved · gold boxes: conditioned on

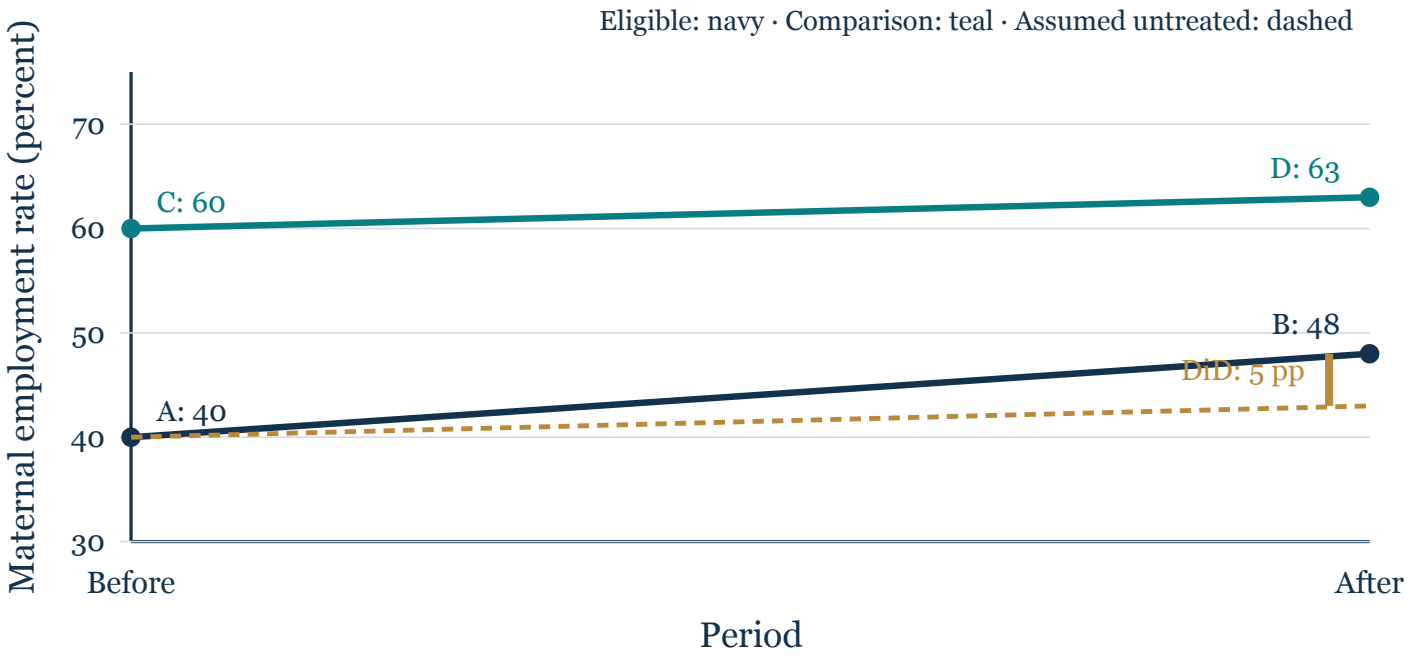
> Eligibility: policy conditions; access: an available place; participation: actual use.

> Define the expansion comparison before interpreting a coefficient as an attendance effect.

> A causal diagram does not establish parallel untreated trends.

Parallel trends is about unobserved untreated changes

It is not a requirement that the groups begin at the same level.



Constructed Head Start employment-rate example using the original A/B/C/D sequence. Dashed untreated paths are assumptions; these are not estimates from the reading.

- Observed DiD: $(Y_B - Y_A) - (Y_D - Y_C) = (48 - 40) - (63 - 60) = 5$ percentage points.
- Assumed untreated eligible post rate: $40 + (63 - 60) = 43\%$.
- Parallel trends concerns the missing no-expansion change, not equal starting levels.

No anticipation and stable comparison units also matter

Define when treatment can begin to affect behaviour.

> No anticipation: the policy has not already changed the baseline outcome.

> Check spillovers and the population represented at both dates.

> A clear policy definition is part of the causal model.

The interaction in a two-by-two DiD equation has a specific meaning

Its interpretation follows the four group-time means.

> $Y_{it} = \beta_0 + \beta_1 P_t + \beta_2 G_i + \beta_3 G_i P_t + u_{it}.$

> G_i: predefined eligible group; P_t: post period; Y_{it}: employment indicator.

> Constructed interaction: $\beta_3 = 0.05$, or five percentage points.

Different characteristics can imply different untreated changes

Unequal starting levels and unequal untreated trends are different issues.

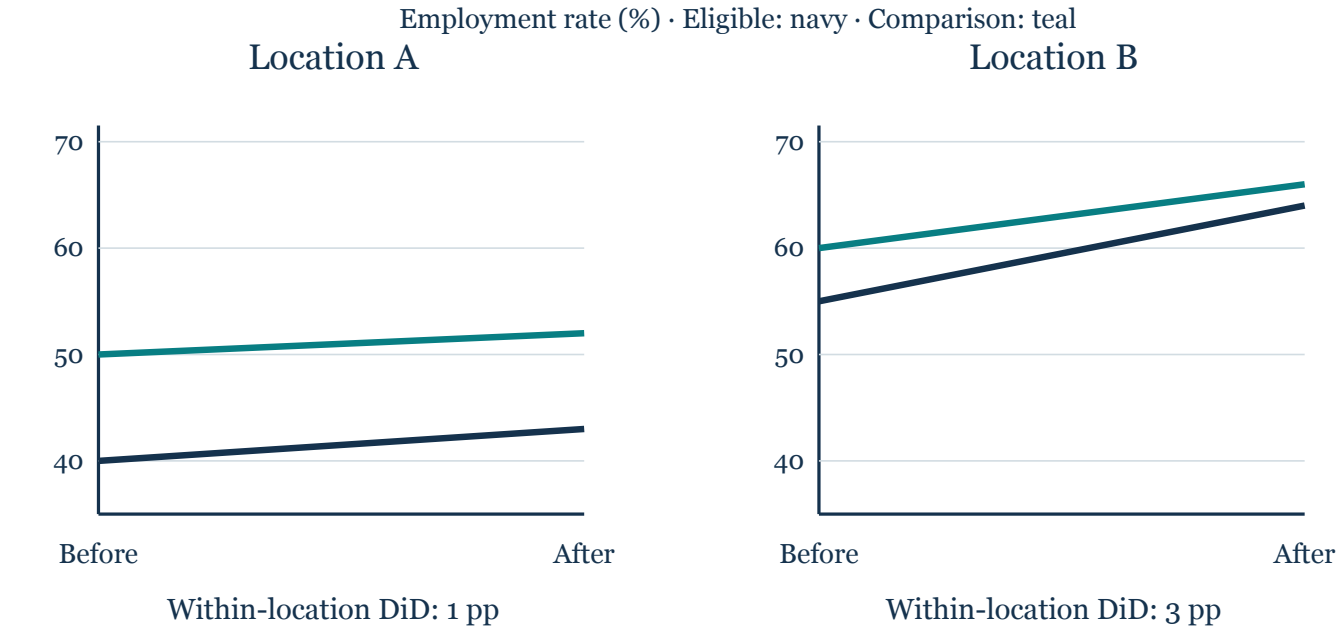
> W: characteristics fixed before intervention.

> Conditional trends: $E[\Delta Y(0) \mid G = 1, W] = E[\Delta Y(0) \mid G = 0, W]$.

> Here the change in Y(0) means the change without expansion.

Compare within location before averaging across locations

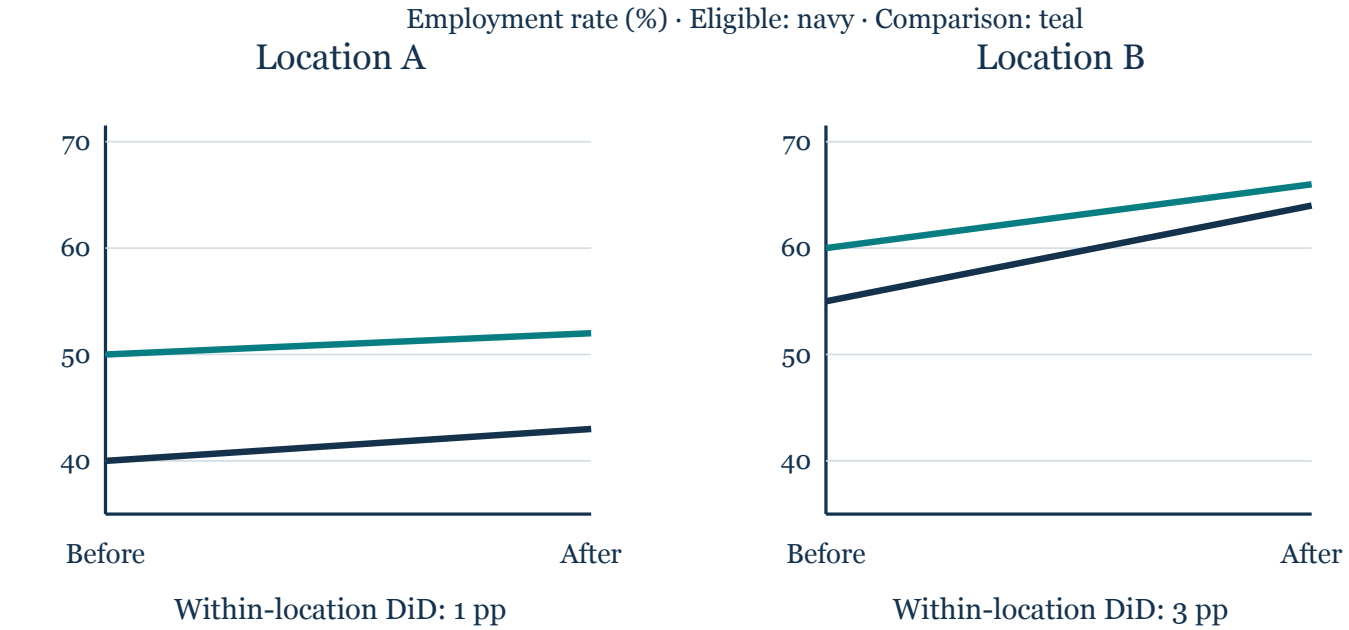
The original two-location idea separates changing opportunities from policy effects.



- Location A: group changes 3 and 2; within-location DiD 1 percentage point.
- Location B: group changes 9 and 6; within-location DiD 3 percentage points.
- Different location mixtures can reverse the unadjusted aggregate comparison.

Average conditional effects for the treated population

The target determines the averaging weights.



Constructed two-location comparison adapted from the Overleaf lesson. Within-location trends and effects are specified for the illustration.

- $\tau(W)$: the conditional effect among treated units at characteristic W.
- $ATT = E[\tau(W) \mid G = 1]$.
- Example: $0.75(1) + 0.25(3) = 1.5$ percentage points.

A covariate in levels need not adjust a trend

Decide whether the model represents untreated changes and effect heterogeneity.

> Time-invariant levels: $\Delta(\gamma W_i) = 0$.

> Model the untreated change as a function of pre-treatment W.

> Allow appropriate heterogeneity and average for the stated population.

Repeat the original zero-effect income-targeting experiment

Known simulation truth lets us diagnose a misleading comparison.

```
. regress dy group, vce(robust)

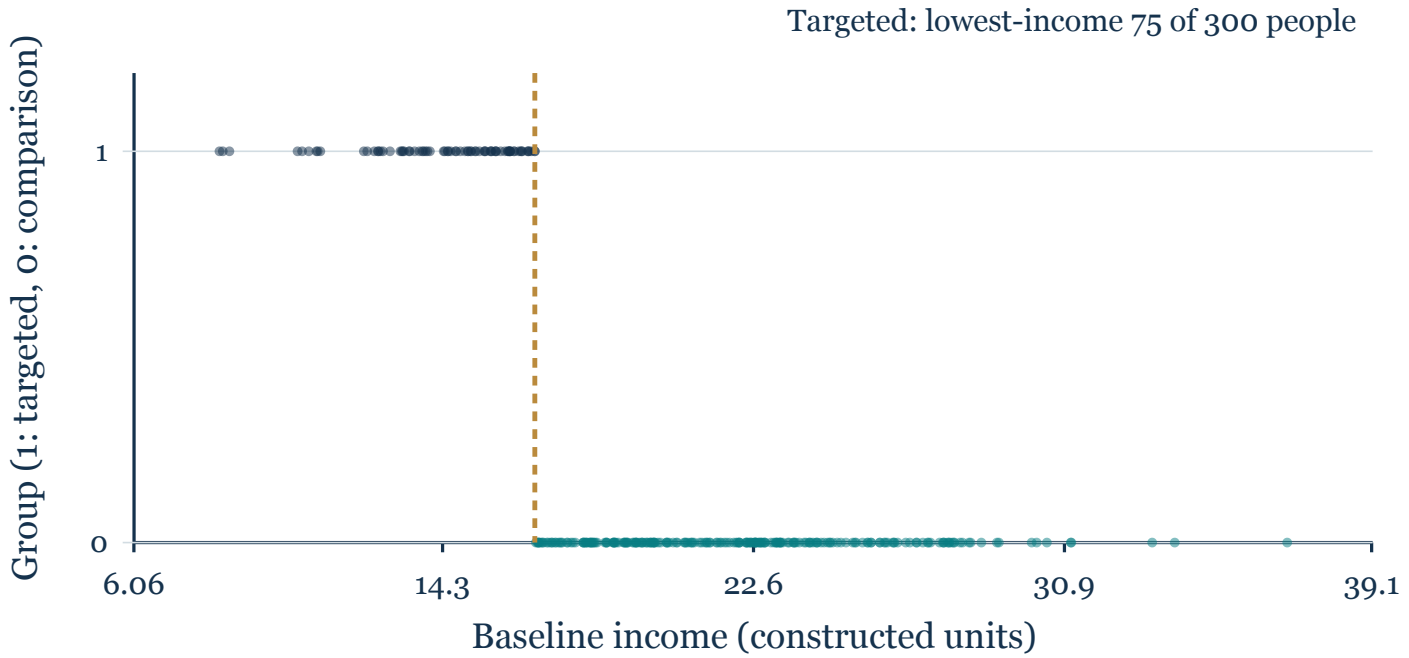
------------------------------------------------------------------------------
             |               Robust
          dy | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       group |  -3.222965   .1420669   -22.69   0.000    -3.502547   -2.943384
       _cons |   2.900887   .1026169    28.27   0.000     2.698941    3.102832
------------------------------------------------------------------------------
```

Actual StataMP output. Reproduce with did_overleaf.do; the full output is in the download.

- Constructed data: 300 people; 75 targeted; known effect zero.
- Untreated changes vary with the baseline income used for targeting.
- A correct linear change model and a defensible comparison are different requirements.

Similar-group comparisons need overlap

An estimator cannot create comparison observations absent from the data.



Actual generated observations from the zero-effect 300-person experiment. Lowest-quarter baseline-income targeting leaves no comparison observations below the cutoff.

- Overlap for treated characteristics requires comparison observations: $P(G = 0 \mid W) > 0$.
- Deterministic income targeting leaves no comparison observations below its cutoff.
- A fitted extrapolation should be justified and labelled as such.

Separate spillovers, treatment versions and anticipation

Use the original SUTVA discussion to define the policy comparison precisely.

> No interference: other units' exposure does not change this unit's potential outcome.

> Well-defined treatment: specify the preschool access or service bundle being compared.

> No anticipation: the baseline outcome is not already affected by the future intervention.

Cluster inference for repeated units

Errors within a state can remain related after fixed effects.

> State fixed effects remove a stable mean component; they do not force the remaining errors in different years to be independent.

> The policy assignment and shared shocks help determine the appropriate clustering unit.

Interpret a feasible tax difference

Use the fitted contrast and preserve the outcome units.

```
. lincom .10*beertax

------------------------------------------------------------------------------
    fatality | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   -.063998   .0357078    -1.79   0.080    -.1358329    .0078369
------------------------------------------------------------------------------
```

Inspect the 1982-to-1988 changes separately

A longer change can use different variation from annual fixed effects.

```
. regress dy dx, vce(robust)

------------------------------------------------------------------------------
             |               Robust
          dy | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          dx |  -1.040973   .3550061    -2.93   0.005    -1.755563   -.3263823
       _cons |  -.0720371   .0653552    -1.10   0.276    -.2035903    .0595161
------------------------------------------------------------------------------
```

Distinguish within-state and endpoint comparisons

Reproduce the evidence, then explain its economic meaning.

Practice

Change the endpoint first-difference comparison to 1983 and 1988. Use the same 48 states, calculate the differences yourself and compare the slope with the original endpoint window. Keep your commands, log, output and interpretation.

Distinguish within-state and endpoint comparisons: explanation

A reproducible result needs both a calculation and an interpretation.

> State FE slope: −0.639980; 1982–1988 endpoint slope: −1.040973.

> 1983–1988 endpoint slope: −0.845389, robust SE 0.332000.

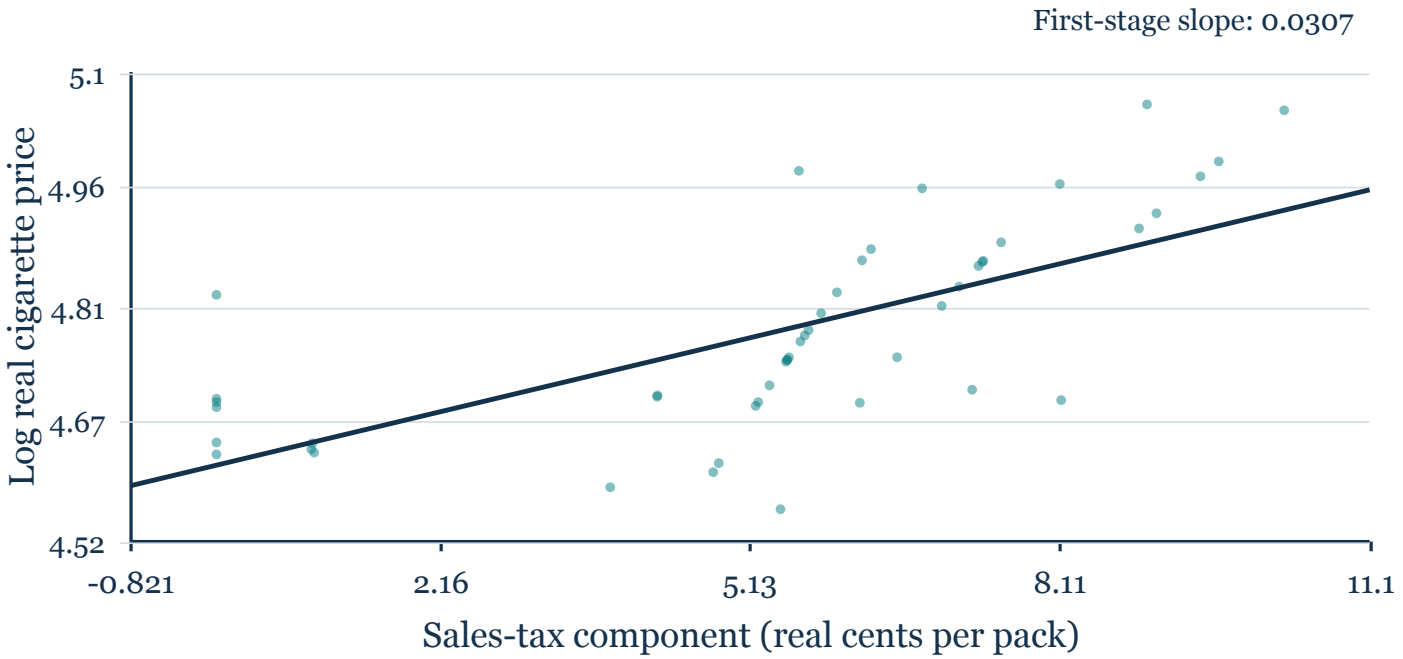
> A changed endpoint window changes the dated comparison.

From panel comparisons to instrumental variation

From panel comparisons to instrumental variation

Do higher cigarette prices reduce sales?

Price is chosen in a market, so its variation may be endogenous.



Actual 1995 state price and tax data. The first stage describes relevance; it cannot establish the exclusion restriction.

- Question: isolate a demand response from other state differences.
- IV: instrumental variables; the proposed source of price variation needs its own argument.

Define the demand equation and its units

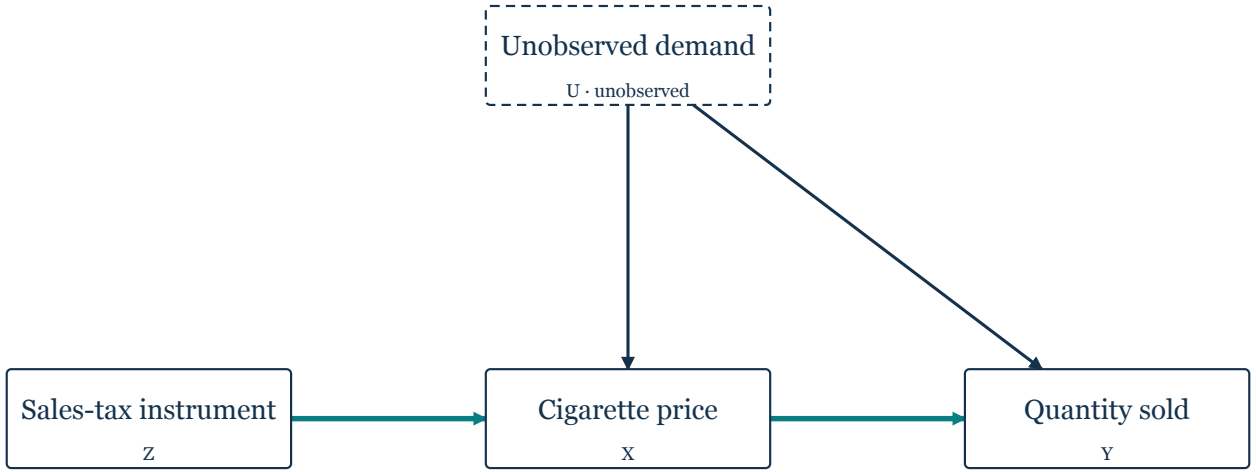
A log price slope is an elasticity under the model.

> $$\log Q_i = \alpha + \beta \log P_i + u_i$$

> Source deflator has base one: real cents per pack equal nominal cents divided by the recorded price index.

Relevance is the first instrument requirement

The instrument must move the endogenous regressor.



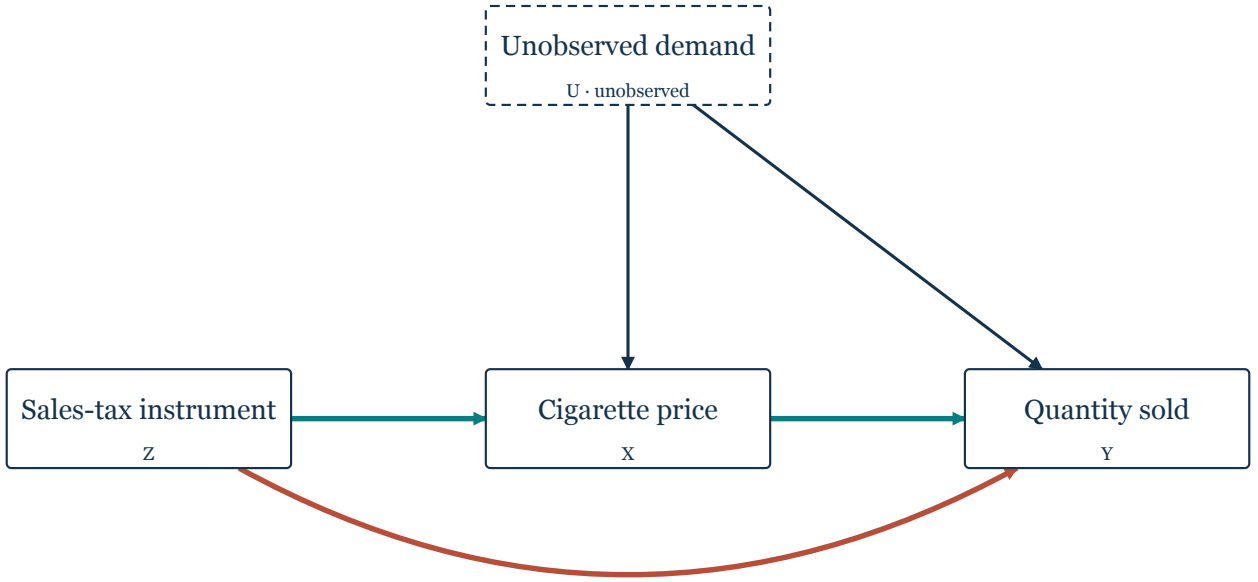
1 open directed route · 0 open alternatives · gold: conditioned on

Assumed causal structure · dashed boxes: unobserved · gold boxes: conditioned on

- First stage: $X = \pi_0 + \pi_1 Z + v$
- Relevance: $\pi_1 \neq 0$
- Z: real cents per pack; $X = \log P$

Exogeneity and exclusion justify the comparison

A strong first stage does not establish either condition.



- Exogeneity: no instrument relationship with omitted structural demand causes.
- Exclusion: no causal instrument-to-quantity channel bypassing price.
- Required IV moment: $\mathrm{Cov}(Z, u) = 0$

Derive the simple IV ratio

Compare how the instrument moves the outcome and the regressor.

> First stage: $X = \pi_0 + \pi_1 Z + v$

> Reduced form: $Y = c + \rho Z + \eta, \quad \rho = \beta \pi_1$

> $\widehat{\beta}_{IV} = \hat{\rho} / \widehat{\pi}_1$

Inspect the OLS demand comparison first

It is a useful benchmark, not the target by default.

```
. regress logq logp, vce(robust)

------------------------------------------------------------------------------
             |               Robust
        logq | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        logp |  -1.213057   .1945896    -6.23   0.000    -1.604746   -.8213685
       _cons |   10.33892   .9348202    11.06   0.000     8.457229    12.22062
------------------------------------------------------------------------------
```

Read the actual first stage

State the instrument units as recorded in the data.

```
. regress logp salestax, vce(robust)

------------------------------------------------------------------------------
             |               Robust
        logp | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    salestax |   .0307289   .0048354     6.35   0.000     .0209956    .0404621
       _cons |   4.616546   .0289177   159.64   0.000     4.558338    4.674755
------------------------------------------------------------------------------
```

Read the reduced form using the same controls and sample

Its scale is quantity change per instrument unit.

```
. regress logq salestax, vce(robust)

------------------------------------------------------------------------------
             |               Robust
        logq | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    salestax |  -.0332974   .0102541    -3.25   0.002    -.0539378    -.012657
       _cons |   4.717449   .0650586    72.51   0.000     4.586493    4.848405
------------------------------------------------------------------------------
```

What the two stages actually fit

Use the instrument-predicted part of price, with a dedicated IV variance calculation.

> C: included exogenous controls; γC: their coefficient-weighted sum.

> First stage: $\widehat{X} = \widehat{E}_{linear}(X \mid Z, C)$

> Structural residual: $\widehat{u} = Y - \widehat{\alpha} - \widehat{\beta} X - \widehat{\gamma} C$

> 2SLS uses instrument-predicted variation; IV inference uses the structural model.

Use the dedicated IV estimator for uncertainty

Manual fitted-regressor OLS is not enough for valid IV SEs.

```
ivregress 2sls logq (logp = salestax), vce(robust)
```

Compare the actual IV coefficient with its ratio

The do-file verifies equality before rounding.

```
. ivregress 2sls logq (logp=salestax), vce(robust)

------------------------------------------------------------------------------
             |               Robust
        logq | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        logp |  -1.083587   .3122036    -3.47   0.001    -1.695495    -.471679
       _cons |   9.719877   1.496143     6.50   0.000      6.78749    12.65226
------------------------------------------------------------------------------
```

Included exogenous controls belong in the IV specification

They help define the same conditional comparison in both stages.

```
. ivregress 2sls logq logincome (logp=salestax), vce(robust)

------------------------------------------------------------------------------
             |               Robust
        logq | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        logp |  -1.143375   .3604805    -3.17   0.002    -1.849904   -.4368463
   logincome |   .2145153   .3018477     0.71   0.477    -.3770953    .8061258
       _cons |   9.430658   1.219402     7.73   0.000     7.040675    11.82064
------------------------------------------------------------------------------
```

Hold income fixed in both supporting IV regressions

Compatible controls and observations make the controlled ratio meaningful.

```
. ivregress 2sls logq logincome (logp=salestax), vce(robust)

------------------------------------------------------------------------------
             |               Robust
        logq | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
        logp |  -1.143375   .3604805    -3.17   0.002    -1.849904   -.4368463
   logincome |   .2145153   .3018477     0.71   0.477    -.3770953    .8061258
       _cons |   9.430658   1.219402     7.73   0.000     7.040675    11.82064
------------------------------------------------------------------------------
```

Actual StataMP output; reproduce with iv_extension.do in the Stata download. Coefficient table shown; the download retains the full model output.

- Use the same exogenous income control and 48 states in every supporting equation.
- Controlled ratio: $\widehat{\beta} = \hat{r}_Z / \hat{\pi}_Z$.
- Read the dedicated IV result for its uncertainty.

Whose treatment changes when encouragement changes?

A binary encouragement example introduces the treatment-response groups.

> Compliers: attend only if offered. Always-takers/never-takers: unchanged attendance.

> Monotonicity: the offer does not reduce any individual's attendance.

> Constructed ratio: $0.90/(0.60 - 0.30) = 3$ score points.

With heterogeneous effects, say whose response IV identifies

A constant coefficient can hide a local treatment comparison.

> LATE: average treatment effect for compliers under the binary-IV assumptions.

> An IV population is determined by the instrument and the maintained model.

Weak instruments make the ratio fragile

A near-zero first stage magnifies uncertainty.

```
. estat firststage

  First-stage regression summary statistics
  ------------------------------------------------------------------------------
               |             Adjusted      Partial       Robust
      Variable |   R-sq.       R-sq.        R-sq.       F(1,45)   Prob > F
  -------------+----------------------------------------------------------------
          logp |  0.6389      0.6228       0.5009       44.7305    0.0000
  ------------------------------------------------------------------------------
```

> Weak first stage → denominator uncertainty can dominate the ratio.

> Strength and validity are different requirements.

A weak-instrument-robust test asks a stated coefficient question

Protecting inference from a weak first stage does not validate the instrument.

```
. estat weakrobust, betanull(-1)

Test robust to weak instruments
Model VCE: Robust

 ( 1)  logp = -1

Anderson–Rubin chi2(1) =   0.16
           Prob > chi2 = 0.6914

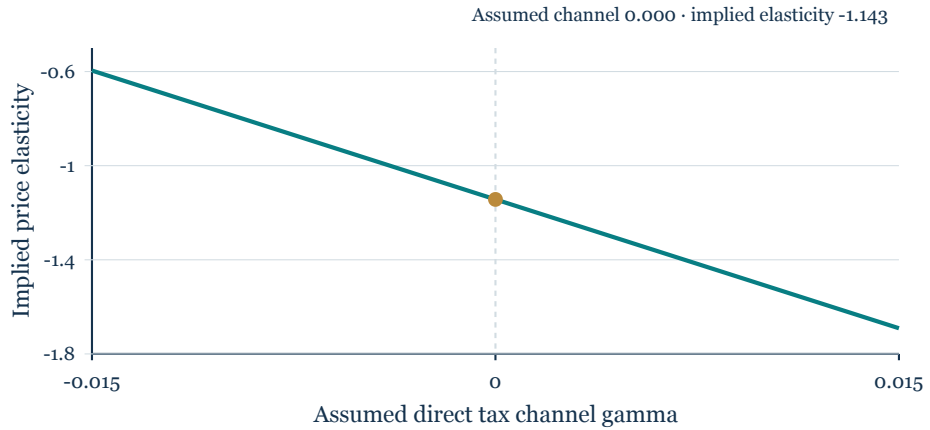
Note: Anderson–Rubin test reported by
      default because model is just
      identified.
```

Actual StataMP output; reproduce with iv_extension.do in the Stata download.

- For a proposed $\beta_0 = -1$, consider $\log Q - \beta_0 \log P = \log Q + \log P$.
- Anderson–Rubin: test the instrument restriction under the proposed coefficient.
- Validity still requires exogeneity, exclusion and justified sampling/inference conditions.

How sensitive is IV to an excluded tax channel?

Treat an assumed exclusion violation as a sensitivity scenario, not an estimate.



Supporting regression slopes are actual income-controlled estimates. The extra direct tax channel is assumed for sensitivity, not estimated.

> Extended reduced form: $r = \beta\pi + \gamma$.
>
> Assumption-based sensitivity: $\beta(\gamma) = (r - \gamma)/\pi$.
>
> The direct-channel value is assumed; the supporting regression slopes are actual estimates.

Practice

Is the direct-channel value gamma estimated by this exercise?

Explanation: No. It is an explicitly assumed sensitivity value.

Reconstruct IV and assess the instrument separately

Reproduce the evidence, then explain its economic meaning.

Practice

Add log real income to the first stage, reduced form and IV equation, using the same 48 states. Verify the controlled ratio and compare the elasticity with the original no-income result. Keep your commands, log, output and interpretation.

Reconstruct IV and assess the instrument separately: explanation

A reproducible result needs both a calculation and an interpretation.

> Original IV elasticity: −1.083587.

> Income-controlled IV: −1.143375, robust SE 0.360481.

> Both stages need the same controls and states; exclusion remains a substantive argument.

Next, change the outcome from a level to a probability

A binary response requires a different interpretation of the fitted function.

> Next outcome: a zero-or-one event.

> Compare models on the probability scale after defining their functions.

A two-period athlete example uses the same algebra

State what disappears and what the comparison still needs.

$$GPA_{it} = \beta\, season_{it} + \alpha_i + \lambda_t + u_{it}$$

$$\Delta GPA_i = \beta \Delta season_i + \Delta\lambda + \Delta u_i$$

GPA: grade point average; season: a zero-or-one indicator.

First differences

Try the comparison before opening the explanation.

Practice

In $GPA_{it} = \beta\, season_{it} + \alpha_i + \lambda_t + u_{it}$ for two semesters, write the differenced equation and say what disappears.

First differences: explanation

Follow the calculation and interpret its units.

> $\Delta GPA_i = \beta \Delta season_i + \Delta\lambda + \Delta u_i$. The time-invariant α_i disappears, as do coefficients on time-invariant regressors. Semester effects do not disappear; with two periods they become a common intercept. Time-varying confounding can remain.

What year effects absorb

Try the comparison before opening the explanation.

Practice

Tax_t is identical across all districts in each year. Can its coefficient be estimated separately in a regression with an intercept and a full set of year effects? Can $Tax_t \times near_i$ survive?

What year effects absorb: explanation

Follow the calculation and interpret its units.

> The common tax column is a linear combination of the year indicators, so its separate coefficient is not identified in that regression. The tax-by-group product can retain variation because it differs across groups within years. Check that it is not reproduced by all the included controls. It describes a differential group slope; it does not recover the absorbed common tax effect.

DiD arithmetic

Try the comparison before opening the explanation.

> Work in percentage points before converting to a zero-or-one outcome scale.

Practice

In the Head Start teaching diagram, eligible mothers' employment rises from 40% to 48%; comparison mothers' employment rises from 60% to 63%. Calculate DiD in percentage points and on the zero-or-one employment-regression scale. State the untreated-trend assumption.

DiD arithmetic: explanation

Follow the calculation and interpret its units.

> $(48 - 40) - (63 - 60) = 5$ percentage points.

> Binary-outcome regression contrast: 0.05.

> The untreated eligible post mean is assumed, not observed.

Doubly robust DiD combines two supporting models

Two chances for model correctness retain the design assumptions.

> Outcome model: $m_0(W) = E[\Delta Y \mid G = 0, W]$.

> Propensity score: $p(W) = P(G = 1 \mid W)$.

> A logistic membership curve stays between zero and one.

> Double robustness concerns supporting-model correctness; conditional trends and overlap remain assumptions.

Read the real-data example from the Overleaf lesson

Identify the sample and the reported target before interpreting ATET.

```
. drdid re age educ black married nodegree hisp re74, ///

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
ATET         |
experimental |
   (1 vs 0)  |  -901.2703   393.6127    -2.29   0.022    -1672.737   -129.8037
------------------------------------------------------------------------------
```

Actual StataMP output. Reproduce with did_overleaf.do; the full output is in the download.

- Retained sample: NSW experimental controls and CPS comparison observations.
- Data-source group membership is not actual training receipt.
- ATET and ATT name the same treated-group average target.

Choose the dependence unit for inference

The number of rows need not equal the number of independent sources.

> Cluster: a group within which errors may be related.

> Independent information follows the assignment/error structure, not just the row count.

Cluster choice

Try the comparison before opening the explanation.

Practice

A policy is assigned to states and repeated outcomes are observed for many districts over years. Why is clustering by district alone potentially insufficient?

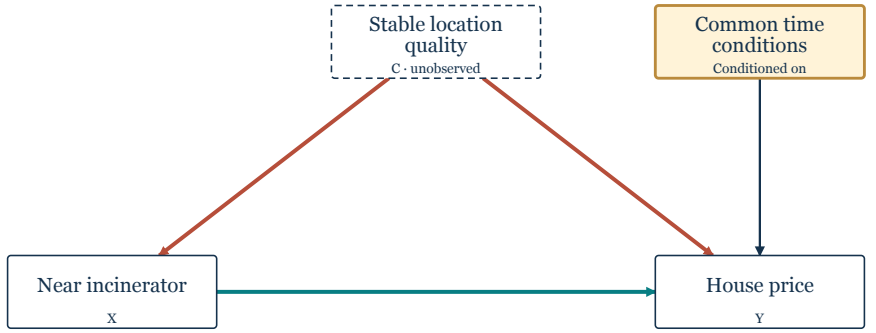
Cluster choice: explanation

Follow the calculation and interpret its units.

> Errors and policy exposure may be shared across districts within a state. Inference should reflect the assignment and dependence structure; state clustering is a starting point when states are independent sampling/assignment clusters. Few states require special care, and a universal cluster-count threshold is not a theorem.

DAG practice: pooling is not within-location identification

Trace the routes before choosing a control.



1 open directed route · 1 open alternative · gold: conditioned on

Assumed causal structure · dashed boxes: unobserved · gold boxes: conditioned on

Practice

In the incinerator example, common time conditions affect prices while stable location quality affects proximity and prices. Do year indicators close the location route? If proximity never changes, what happens to its level coefficient under unit fixed effects, and what extra comparison could support a DiD question?

DAG practice: pooling is not within-location identification: explanation

Explain what conditioning changes in this stated graph.

> Year indicators do not block $X \leftarrow C \rightarrow Y$ through stable location quality. Unit fixed effects can cancel an additive stable location component, but they also absorb a proximity indicator that never changes within a unit. Its separate level coefficient is then unavailable. A proximity-by-post-period interaction can vary across groups and dates. Interpreting that interaction as a DiD effect requires a well-defined intervention and parallel untreated trends, along with the other stated comparison assumptions. Neither pooling nor the causal diagram alone establishes those assumptions.

Wald ratio

Try the comparison before opening the explanation.

Practice

A proposed instrument increases treatment by 0.30 units; its reduced-form outcome coefficient is −0.09. Compute the simple IV ratio and name the conditions behind a causal interpretation.

Wald ratio: explanation

Follow the calculation and interpret its units.

> The ratio is −0.09 divided by 0.30, or −0.30 outcome units per treatment unit. Under the constant-slope structural model, relevance, instrument exogeneity and exclusion justify the causal interpretation. The reduced-form and first-stage regressions must use compatible controls and observations. A strong first stage alone does not establish instrument validity.

Two-stage least squares projects regressors onto the instrument space

The instrument matrix includes the exogenous controls and the excluded instruments.

> $P_Z = Z(Z\prime Z)^{-1}Z\prime$; $\widehat{X} = P_Z X$.

> $\widehat{\beta}_{2SLS} = (X\prime P_Z X)^{-1} X\prime P_Z y$.

> Population moments: $E[z_i u_i] = 0$, with the required rank condition.

Two stages and uncertainty

Try the comparison before opening the explanation.

Practice

Why not report ordinary OLS standard errors from regressing y on first-stage fitted x?

Two stages and uncertainty: explanation

Follow the calculation and interpret its units.

> The second-stage fitted regressor is generated, and the relevant structural residual and variance calculation differ from the naive second-stage OLS calculation. Use a proper IV estimator such as ivregress 2sls with a justified covariance option.

School availability illustrates the exclusion question

A plausible first stage can still leave another route to the outcome.

> Suppose nearby single-sex school availability predicts attendance at that school type, and the outcome is a student's score.

> Residential choice, local resources or school competition may create other relationships.

Unexpected first-stage sign

Try the comparison before opening the explanation.

Practice

School availability predicts less attendance at the intended school type. Is a negative first-stage coefficient alone proof that IV is invalid?

Unexpected first-stage sign: explanation

Follow the calculation and interpret its units.

> No. Relevance requires a nonzero first stage, not a prescribed positive sign. An unexpected sign should prompt checks of coding, selection, theory and strength. It does not by itself prove or disprove exclusion or exogeneity.

The simple IV precision formula shows the cost of a weak first stage

A formula under strong-instrument asymptotics is not a cure for weak identification.

> Simple homoskedastic strong-IV approximation: $\mathrm{Var}(\widehat{\beta}_{IV}) \approx \sigma_u^2 / [SST_X R^2_{X,Z}]$.

> Instrument validity, first-stage strength and the validity of an inference approximation are distinct.

A larger instrument set adds assumptions as well as information

More instruments are not automatically a better design.

> Overidentified: more excluded instruments than endogenous regressors.

> A compatibility test cannot certify all substantive instrument assumptions.

Instrument scaling does not change the simple ratio

Consistent numerator and denominator transformations cancel.

> $\dfrac{\hat{\rho}/c}{\hat{\pi}_1/c} = \dfrac{\hat{\rho}}{\hat{\pi}_1}$ for nonzero c

Write the IV argument as a chain of evidence

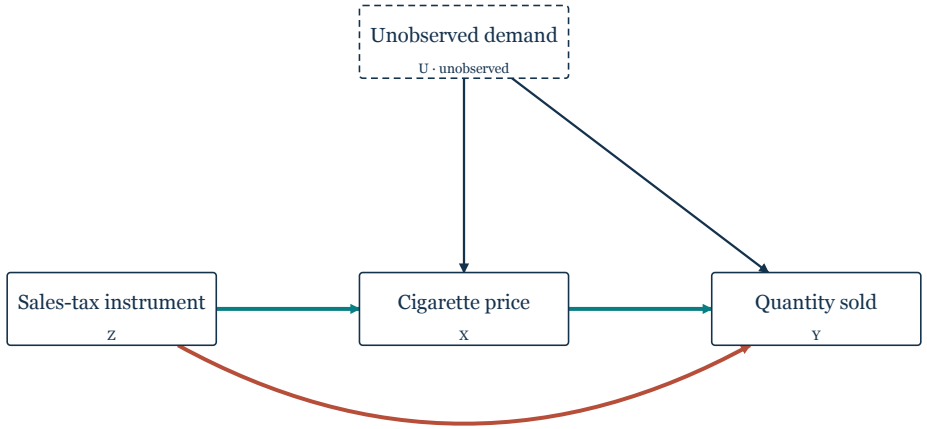
Separate relevance, exclusion, estimation and inference.

> Begin with the endogenous comparison and the omitted forces that threaten OLS.

> This sequence helps a policy reader assess the strategy rather than treat 2SLS as a label for causality.

DAG practice: challenge exclusion with one extra arrow

Trace the routes before choosing a control.



2 open directed routes · 0 open alternatives · gold: conditioned on

Assumed causal structure · dashed boxes: unobserved · gold boxes: conditioned on

Practice

The tax instrument predicts cigarette price. Add a direct tax-to-quantity arrow that bypasses price. Could the first stage stay strong? Why would dividing the reduced form by the first stage then fail to isolate the intended price elasticity?

DAG practice: challenge exclusion with one extra arrow: explanation

Explain what conditioning changes in this stated graph.

> The first stage can remain strong because the tax-to-price arrow has not disappeared. Relevance does not rule out a second route from the instrument to quantity that bypasses price. The reduced-form instrument association would combine the price channel and that additional direct channel. Dividing by the first stage no longer isolates the price response under the original exclusion argument. Use the graph to name a concrete institutional threat and assess it separately from the numerical strength of the first stage.