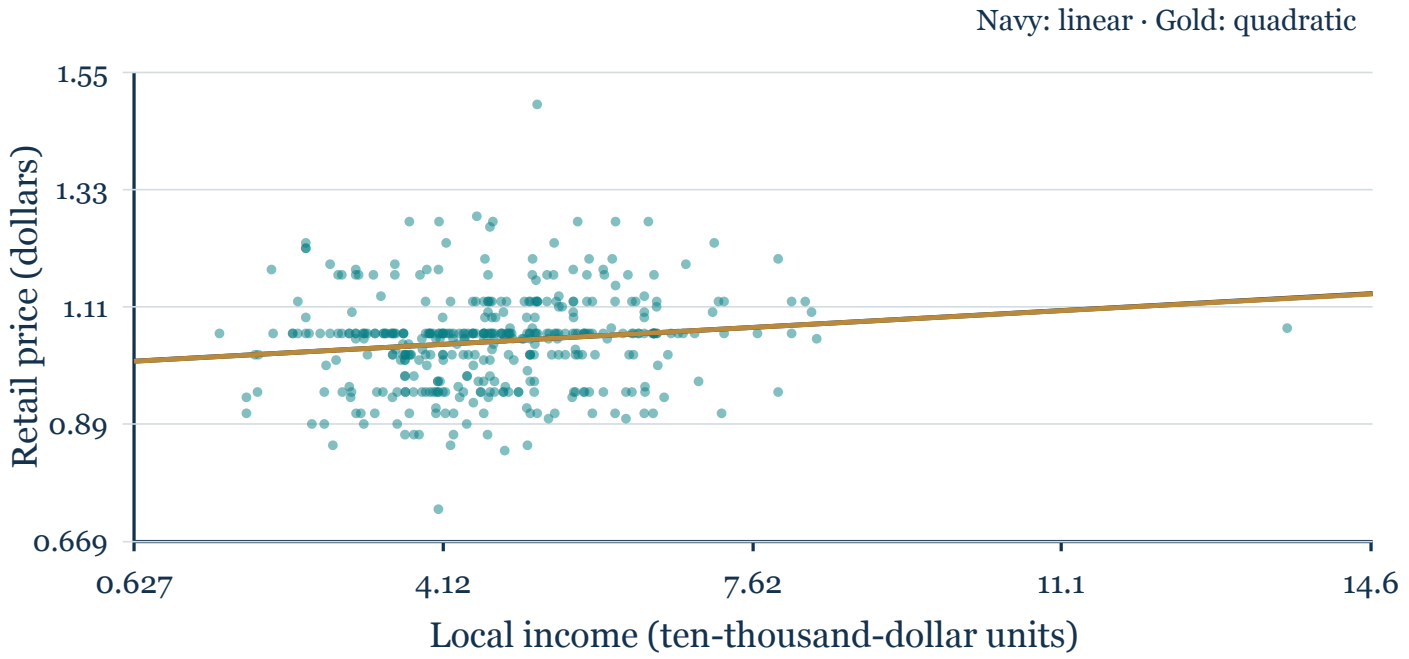
Do prices differ by the same amount or the same percent?

Choose a scale that answers the economic question.



401 observed positive-price and positive-income records. Lines use the actual level and quadratic regressions; no geographic sampling claim is made.

Define the observation and the usable sample

Inspect the data before interpreting a transformed variable.

> $X = \text{income}/10{,}000$

> Outcome: medium soda price, in dollars

A level–level model has a constant unit slope

A one-unit income change produces the same fitted dollar difference.

```
. regress sprice inc10, vce(robust)

------------------------------------------------------------------------------
             |               Robust
      sprice | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       inc10 |   .0090733   .0031523     2.88   0.004     .0028761    .0152705
       _cons |   1.002219   .0157154    63.77   0.000     .9713236    1.033114
------------------------------------------------------------------------------
```

$$\widehat{P} = a + bX$$

$$\Delta\widehat{P} = b\Delta X$$

A logarithm expresses a proportional scale

Small log differences approximate percentage differences.

> $\log(1) = 0, \quad \log(2) \approx 0.693, \quad \exp(0.693) \approx 2$

> $\log(2.20) - \log(2) = \log(1.10) \approx 0.09531$

> $100\{\exp(0.09531) - 1\} \approx 10\%$

A log–level model gives a proportional outcome contrast

Keep the regressor units visible.

> $\widehat{\log P} = a + bX$

> $100\{\exp(b\Delta X) - 1\}$

Which average does a log fit describe?

An average log is not the log of an arithmetic average.

> Arithmetic mean: $(1 + 9)/2 = 5$

> Exponentiated mean log: $\exp\{(\log 1 + \log 9)/2\} = 3$

> State which prediction quantity is being compared.

A level–log model gives a level outcome contrast

A percent change in X becomes a small change in log X.

$$\widehat{Y} = a + b \log X$$

$$\Delta \widehat{Y} = b \log(X_1 / X_0)$$

A log–log slope is an elasticity

It relates proportional changes in both variables.

$$\widehat{\log Y} = a + b \log X$$

$$\%\Delta \widehat{Y} = 100\{(X_1/X_0)^b - 1\}$$

Read the actual price–income elasticity

A unit-free slope still needs an empirical interpretation.

```
. regress lp li, vce(robust)

------------------------------------------------------------------------------
             |               Robust
          lp | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
          li |   .0374749   .0145083     2.58   0.010     .0089527    .0659971
       _cons |  -.3613775   .1559216    -2.32   0.021     -.667908   -.0548471
------------------------------------------------------------------------------
```

A quadratic lets the slope change with X

Define the squared term before interpreting it.

> $\widehat{Y} = 500 + 8X - 0.2X^2$

> Local slope: $d\widehat{Y}/dX = 8 - 0.4X$

> General slope: $b_1 + 2b_2X$

Use a finite quadratic difference for a finite change

The tangent slope is a local description.

> $\widehat{Y}(10) = 560, \quad \widehat{Y}(15) = 575, \quad \Delta\widehat{Y} = 15$

> $\Delta\widehat{Y} = b_1\Delta + b_2(2X\Delta + \Delta^2)$

> Local approximation: $4 \times 5 = 20$, not the finite difference.

A turning point must be inside a meaningful range

Do not report an unsupported optimum.

> $X^* = -\dfrac{b_1}{2b_2}$ when $b_2 \neq 0$

Inspect the actual curvature rather than assume it

A flexible model may provide little evidence of bending.

```
. regress sprice c.inc10##c.inc10, vce(robust)

------------------------------------------------------------------------------
             |               Robust
      sprice | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       inc10 |   .0091573   .0103187     0.89   0.375    -.0111286    .0294432
             |
c.inc10#c.inc10 |  -7.86e-06   .0008332    -0.01   0.992    -.0016458    .0016301
             |
       _cons |   1.002011   .0304353    32.92   0.000     .9421772    1.061845
------------------------------------------------------------------------------
```

Ask Stata for slopes at stated values

margins makes the comparison explicit.

```
. margins, dydx(inc10) at(inc10=(2 4 6))

Conditional marginal effects                                    Number of obs = 401
Model VCE: Robust

Expression: Linear prediction, predict()
dy/dx wrt:  inc10
1._at: inc10 = 2
2._at: inc10 = 4
3._at: inc10 = 6

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
inc10        |
         _at |
          1  |   .0091259   .0071961     1.27   0.205    -.0050213    .0232731
          2  |   .0090945    .004394     2.07   0.039      .000456    .0177329
          3  |    .009063   .0030074     3.01   0.003     .0031507    .0149754
------------------------------------------------------------------------------
```

Retransforming a log fit needs care

The exponential of an average log is not the average level.

> $\log Y = m(X) + u$

> $E(Y \mid X) = \exp\{m(X)\}\, E\{\exp(u) \mid X\}$

> The retransformation factor need not be one or constant across X.

Zeros and negative values are substantive information

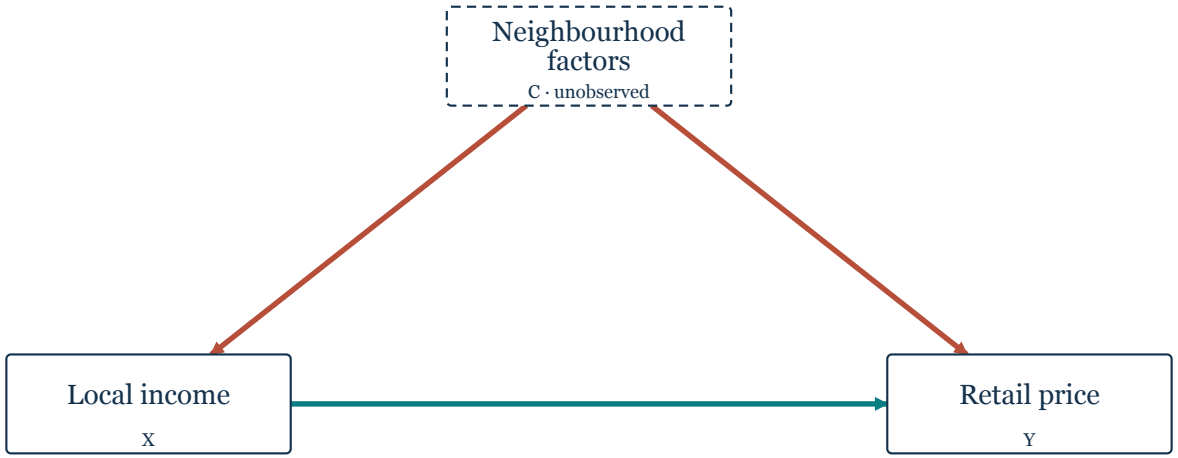
Do not hide them through an automatic log transformation.

> The natural log is defined only for positive values.

> In this retail application, the analysis explicitly requires positive observed price and income.

Transforming the variables does not close a backdoor

A better scale and a credible comparison answer different questions.



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

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

> Functional form chooses the scale and shape.

> Identification addresses the alternative causal explanations.

Translate retail coefficients into a useful contrast

Reproduce the evidence, then explain its economic meaning.

Practice

Change the income comparison from 30,000 to 50,000 dollars. Report the level-model dollar contrast and the exact exponentiated-log percentage contrast. Keep your commands, log, output and interpretation.

Translate retail coefficients into a useful contrast: explanation

A reproducible result needs both a calculation and an interpretation.

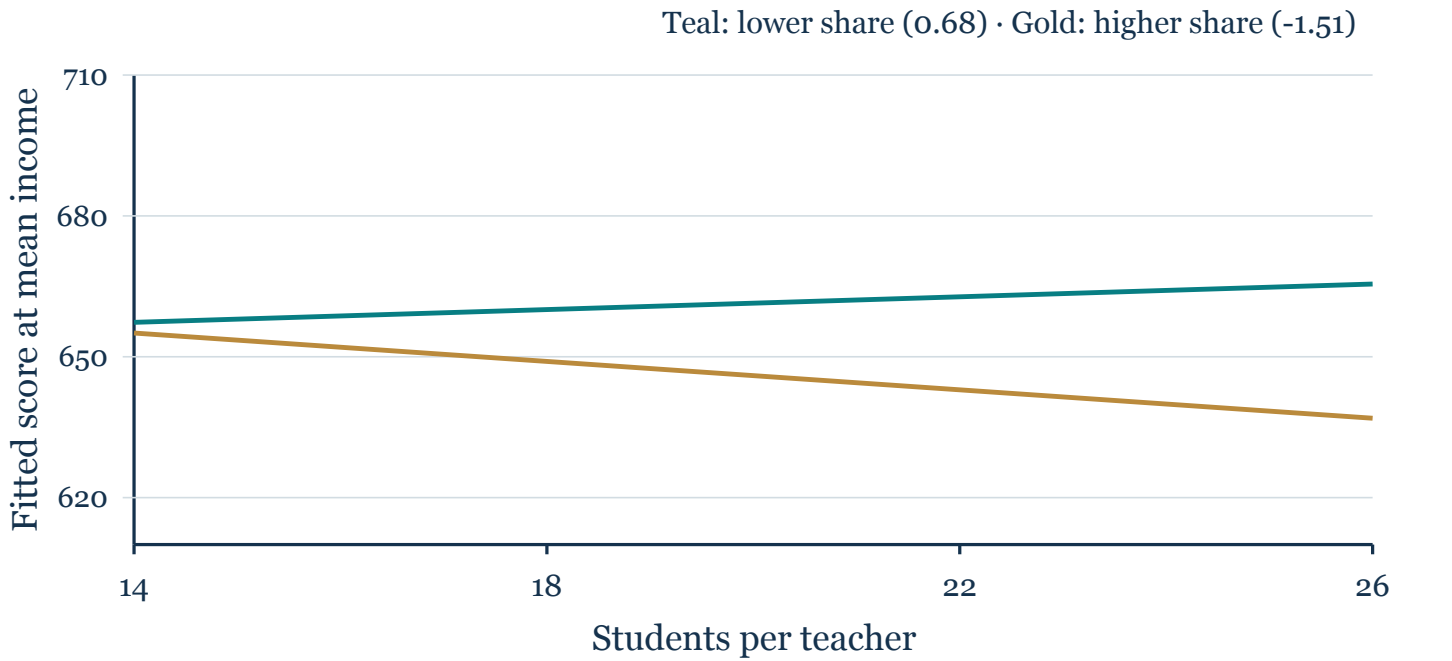
> A 20,000-dollar income difference gives a 0.018147-dollar level-model contrast.

> From 30,000 to 50,000: exponentiated-log contrast 1.932756%.

> A fitted log-value ratio needs extra assumptions to be an arithmetic-mean ratio.

Does the staffing association differ across district groups?

An interaction lets slopes differ in a stated way.



Actual group-specific fitted lines at the overall sample mean of income. High language share means at least 10 percent English learners.

Define the indicator and reference group

Coding determines which coefficients describe the baseline.

> $D_i = 1\{\mathrm{EL}_i \geq 10\}$

Multiple categories require a base category when there is an intercept

An omitted category supplies the reference comparison.

> $D_1 + \cdots + D_G = 1$: all G indicators and an intercept are linearly dependent.

> Use an intercept and $G - 1$ indicators; interpret against the omitted group.

> Stata factor notation i.group records the category structure.

Write the interacted equation

Include the main terms as well as their product.

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i D_i + \gamma W_i + u_i$$

X: staffing; D: group; W: income; XD: their interaction.

Substitute zero and one before interpreting

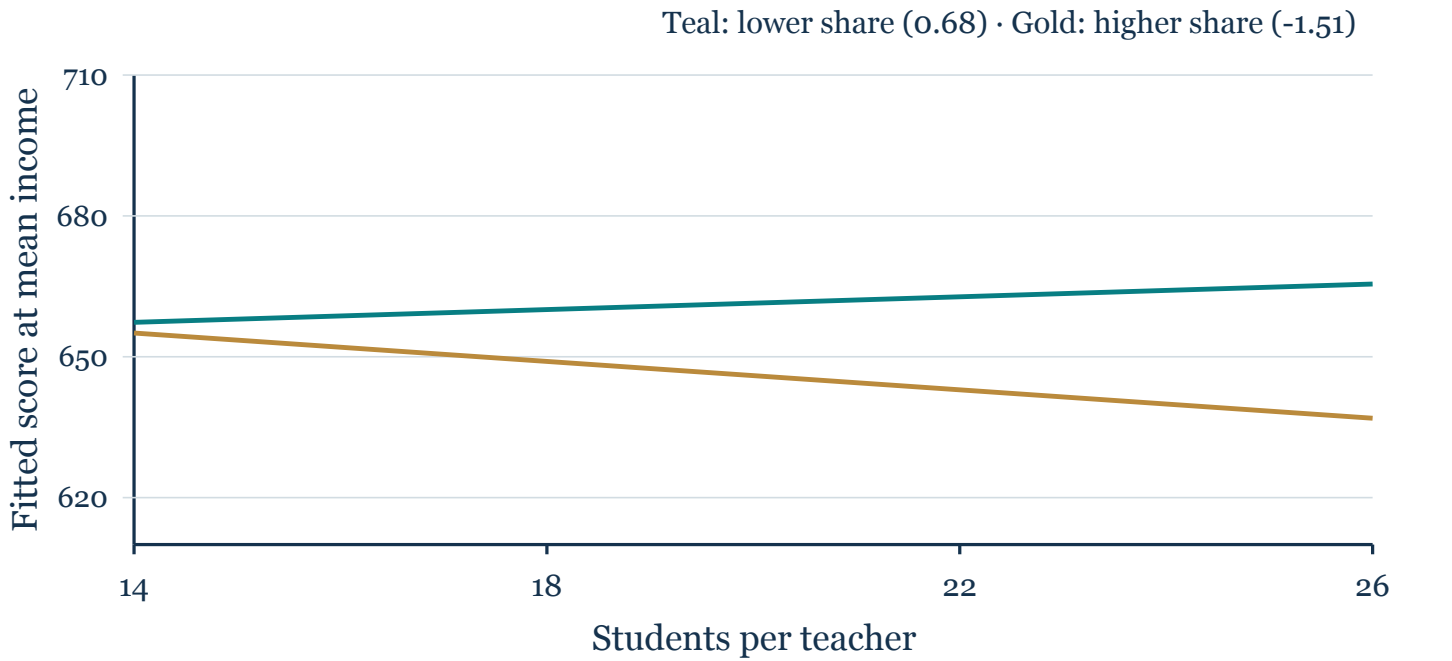
Two group equations reveal the relevant combinations.

> $D = 0: \quad \partial E(Y \mid X, D, W)/\partial X = \beta_1$

> $D = 1: \quad \partial E(Y \mid X, D, W)/\partial X = \beta_1 + \beta_3$

Read both fitted lines in the actual school data

The reference slope and slope difference can have different signs.



Actual group-specific fitted lines at the overall sample mean of income. High language share means at least 10 percent English learners.

Factor-variable syntax keeps the model explicit

Use Stata's interaction notation to include the needed terms.

```
. regress score c.str##i.high income, vce(robust)

------------------------------------------------------------------------------
             |               Robust
       score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |   .6800597   .3532615     1.93   0.055    -.0143453    1.374465
      1.high |   28.41596   12.83889     2.21   0.027     3.178593    53.65332
             |
  high#c.str |
          1  |  -2.192258   .6399264    -3.43   0.001    -3.450159   -.9343566
             |
      income |   1.651672   .1034089    15.97   0.000     1.448401    1.854942
       _cons |   622.5112   7.209389    86.35   0.000     608.3397    636.6826
------------------------------------------------------------------------------
```

Calculate a policy contrast within a group

A two-student change uses that group's total slope.

```
. lincom -2*str-2*1.high#c.str

------------------------------------------------------------------------------
       score | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   3.024396   1.071979     2.82   0.005       .91721    5.131582
------------------------------------------------------------------------------
```

$$\Delta\widehat{Y}_{D=1} = -2(\widehat{\beta}_1 + \widehat{\beta}_3)$$

The SE of a sum needs covariance

Coefficient estimates can move together across samples.

> $\mathrm{Var}(\widehat{\beta}_1 + \widehat{\beta}_3) = V_{11} + V_{33} + 2V_{13}$

> Example: $SE = \sqrt{0.04 + 0.09 + 2(-0.03)} = \sqrt{0.07} \approx 0.265$

Read the higher-group slope as one tested quantity

Use the output for the combination the question asks about.

```
. lincom str+1.high#c.str

------------------------------------------------------------------------------
       score | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |  -1.512198   .5359897    -2.82   0.005    -2.565791    -.458605
------------------------------------------------------------------------------
```

A group gap can depend on X

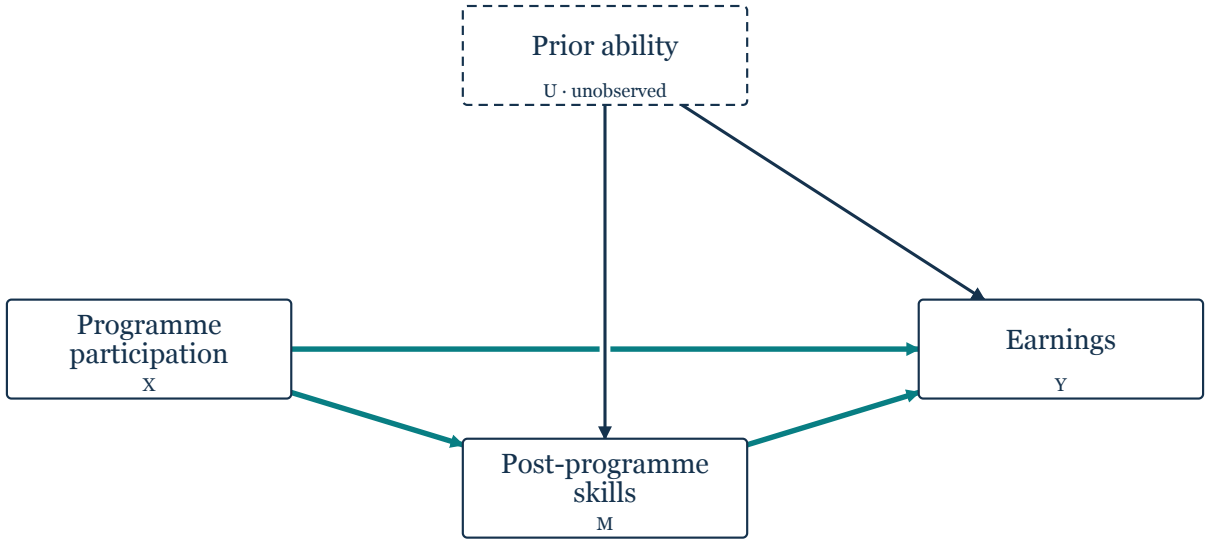
The group main effect is not the gap at every staffing ratio.

> $\text{Gap}(X) = \beta_2 + \beta_3 X$

> Center at 20: $X^c = X - 20$

Separate group heterogeneity from causal identification

A different slope is still a fitted relationship.



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

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

Turn an interaction into two policy contrasts

Reproduce the evidence, then explain its economic meaning.

Practice

Centre staffing at 20 and refit the same interaction model. Interpret the resulting group-indicator coefficient and verify that fitted group slopes have not changed. Keep your commands, log, output and interpretation.

Turn an interaction into two policy contrasts: explanation

A reproducible result needs both a calculation and an interpretation.

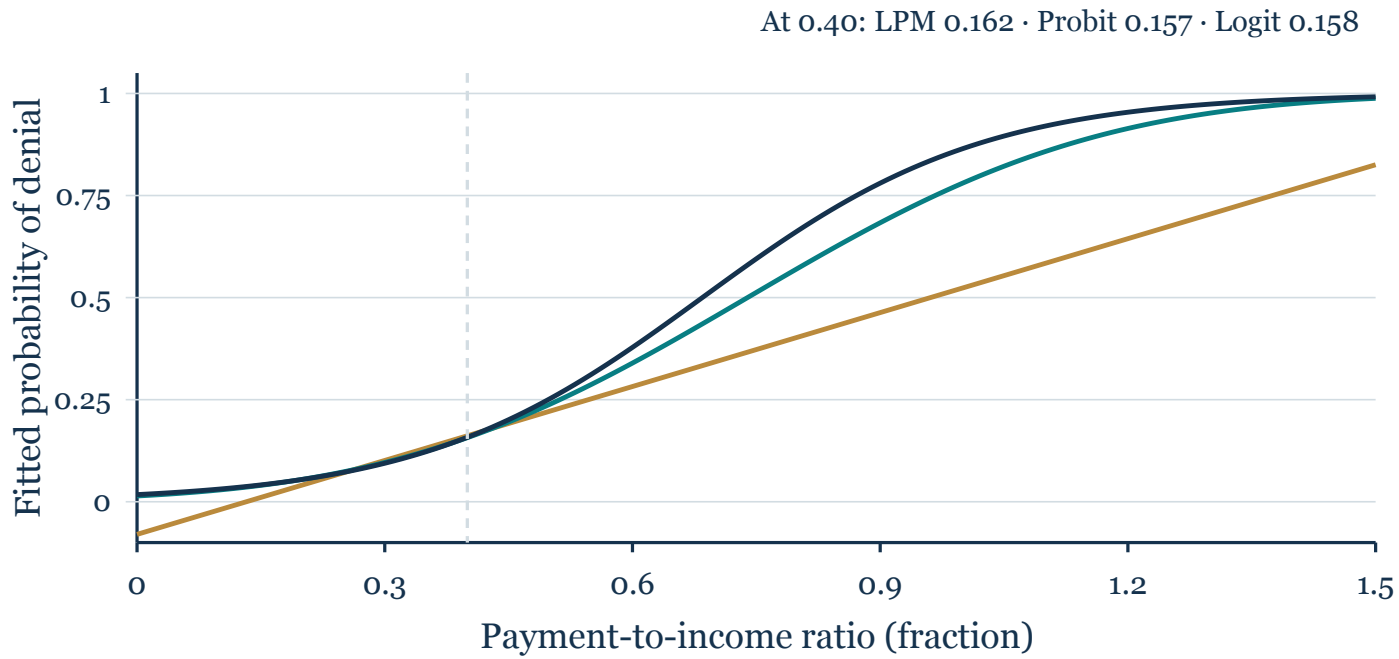
> Original subgroup staffing slopes: 0.680060 and −1.512198.

> Group gap at ratio 20: −15.429199, robust SE 1.128935.

> Centring changes the reference point, not the predictions.

How does payment burden predict mortgage denial?

A zero-or-one outcome is an event indicator.



Actual fitted LPM, probit and logit predictions from the same 2,380 applications. Payment-to-income is a fraction; these are descriptive fits.

- Outcome: denied = 1, not denied = 0.
- Payment-to-income ratio: a fraction, with 0.40 representing 40%.
- Define prediction before making an intervention claim.

The conditional mean of an indicator is a probability

This gives the linear model a useful starting interpretation.

> $Y \in \{0,1\}: \quad E(Y \mid X) = \Pr(Y = 1 \mid X) = p(X)$

> LPM: $p(X) \approx a + bX$

A probability change is not a relative percent change

Keep percentage points and percent distinct.

> $0.23 - 0.18 = 0.05 = 5$ percentage points

> Relative percent: $100(0.05/0.18) \approx 27.8\%$

Read the actual LPM coefficient

Use a feasible ratio change, not an entire unit by default.

```
. regress deny pir, vce(robust)

------------------------------------------------------------------------------
             |               Robust
        deny | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         pir |   .6035349   .0984826     6.13   0.000     .4104144    .7966555
       _cons |  -.0799096   .0319666    -2.50   0.012    -.1425949   -.0172243
------------------------------------------------------------------------------
```

The LPM error variance is generally unequal

Robust SEs address uncertainty, not invalid fitted probabilities.

> $Y^2 = Y \Rightarrow \mathrm{Var}(Y \mid X) = p(X) - p(X)^2$

> $p = 0.1 \Rightarrow \mathrm{Var} = 0.09; \quad p = 0.5 \Rightarrow \mathrm{Var} = 0.25$

Can we predict a district left out of fitting?

A prediction question has a different target from a staffing intervention.

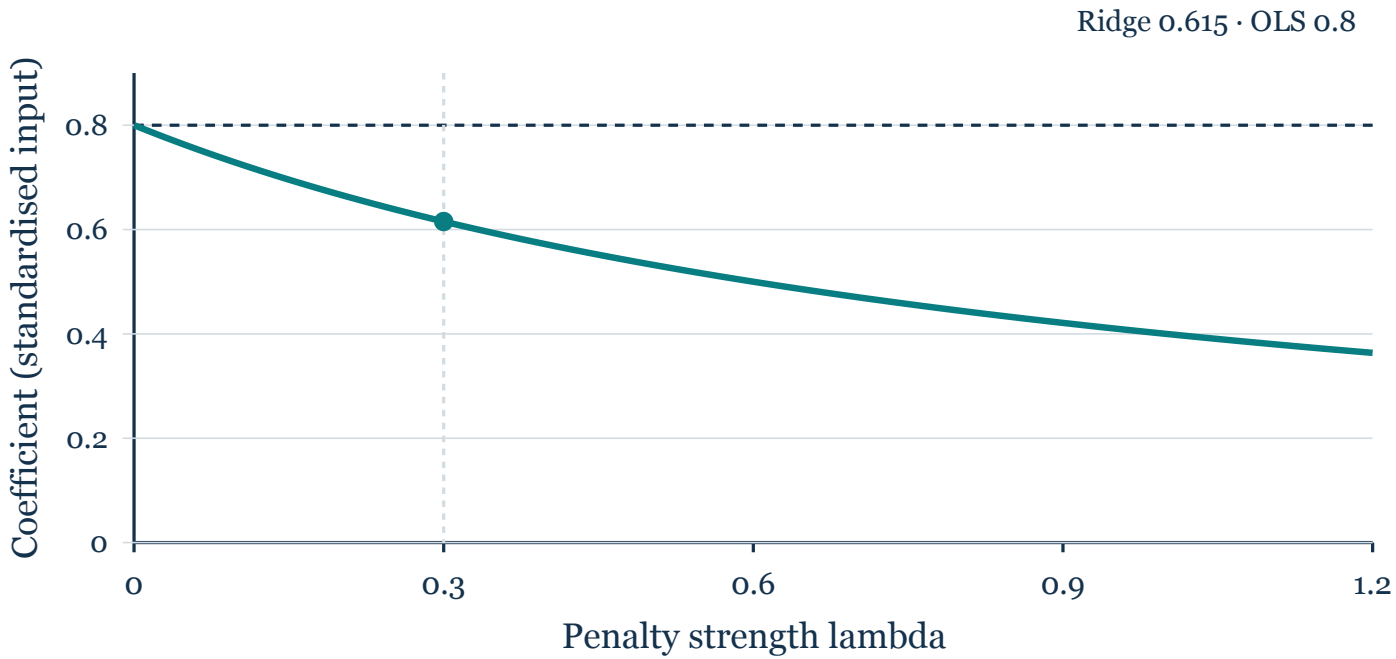
> Training: 280 districts for fitting and tuning.

> Test: 140 districts kept out of fitting and tuning.

> Base: three predictors; expanded: nine predictor columns specified before test evaluation.

Ridge shrinks coefficients with a squared penalty

Accept some bias if it reduces enough estimation variation.

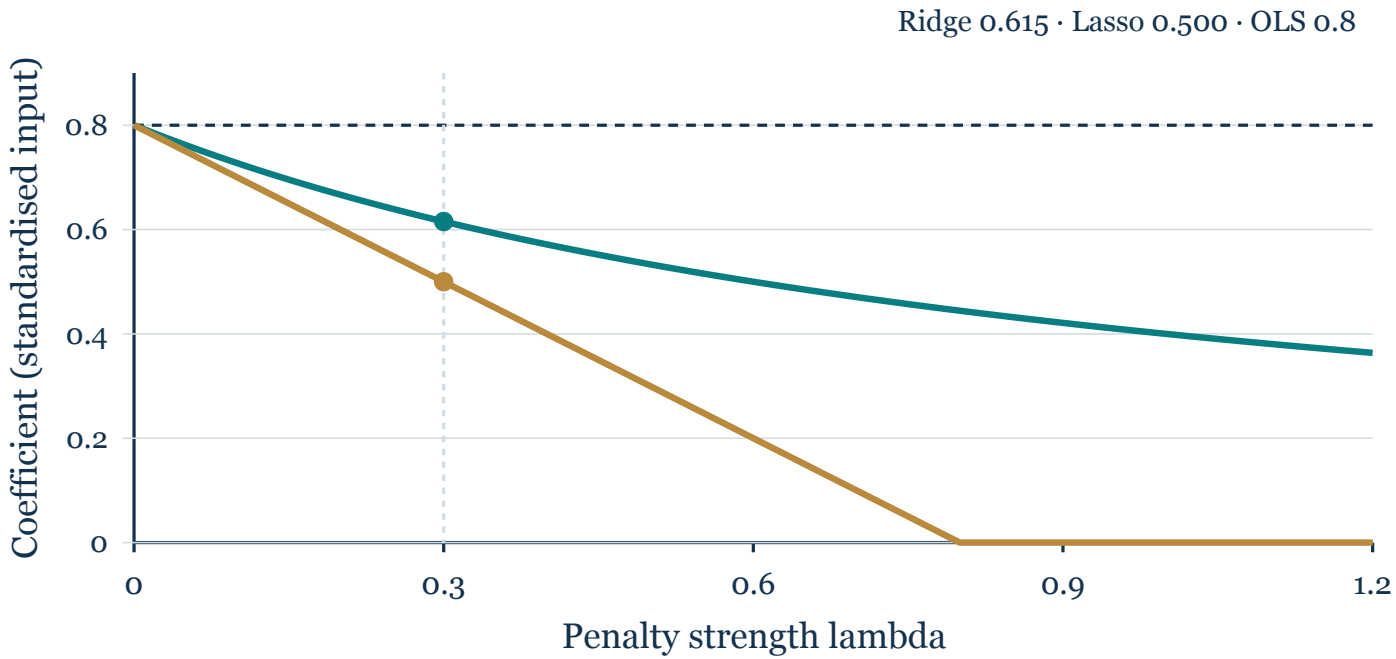


Constructed one-standardised-predictor ridge illustration with OLS slope 0.8. The real nine-column model need not follow this one-column formula.

- Loss: $L(a,b) = \frac{1}{2n}\sum_i [y_i - a - \sum_j b_j z_{ij}]^2$.
- Ridge objective: $L(a,b) + \frac{\lambda}{2}\sum_j b_j^2$.
- Standardise predictor columns using training information; leave the intercept unpenalised.

Lasso can set a predictor coefficient to zero

Absolute and squared penalties behave differently near zero.

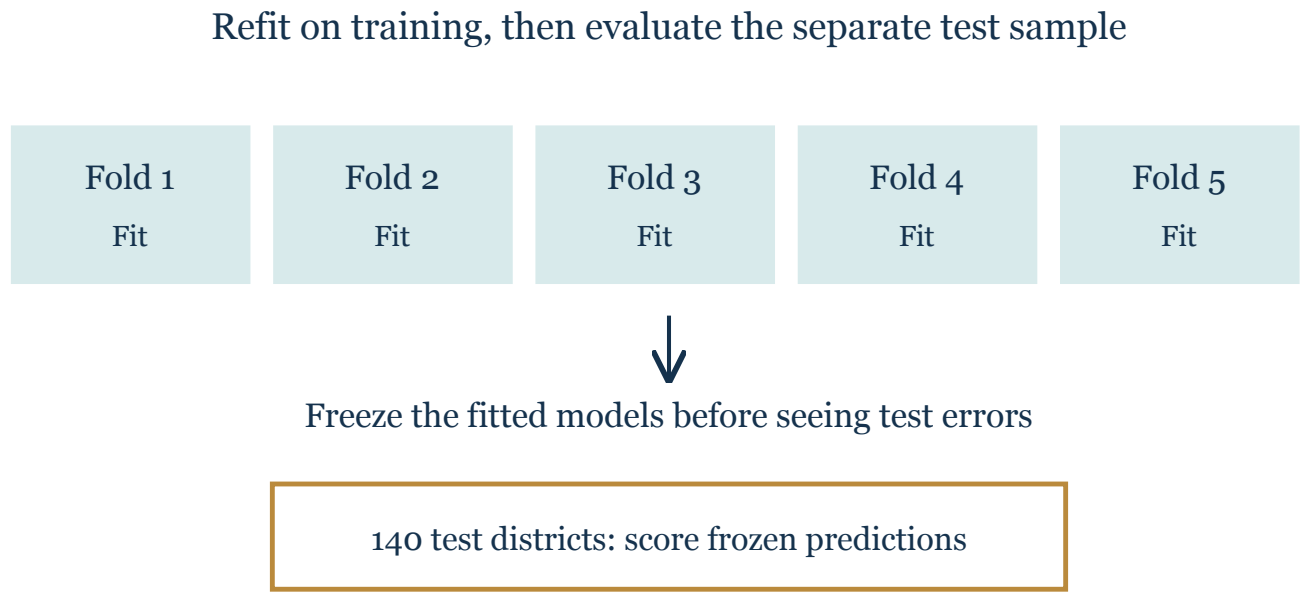


Constructed one-standardised-predictor illustration with OLS slope 0.8. Equal numerical penalties do not imply equal model complexity across methods.

- Lasso objective: $L(a, b) + \lambda \sum_j |b_j|$.
- One standardised predictor: $\hat{b}_R = 0.8/(1 + \lambda)$, $\hat{b}_L = \max(0.8 - \lambda, 0)$.
- These simple paths illustrate the penalties; the actual model has nine correlated columns.

Choose the penalty inside the training sample

Cross-validation reserves the test answers for a later check.

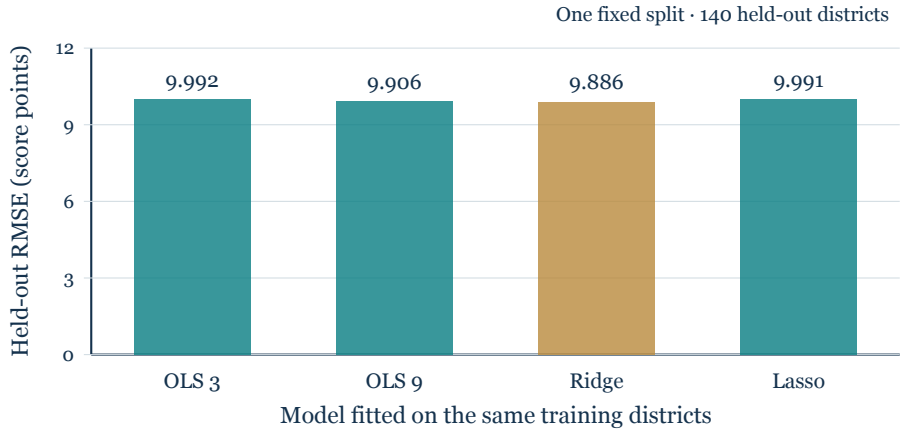


Training-only cross-validation. The separate test sample is used after the fitting and tuning choices are frozen.

- Five training folds: fit on four, validate on one, then rotate.
- Choose the penalty and refit using training observations.
- Score frozen predictions on the separate 140-district test sample.

Compare actual predictions on the same held-out districts

A small error difference is evidence for this run, not a universal ranking.



Actual Stata errors on one fixed split: 280 training and 140 held-out districts. Lower RMSE is better for this prediction task.

- Same training/test split and the same outcome scale for every model.
- Penalties are selected inside training; the test set is evaluated afterwards.
- No test-error ranking identifies a causal staffing effect.

Practice

Does this one test split establish a universally best estimator?

Explanation: No. These small RMSE differences describe this particular evaluation.

Predictive selection does not choose a causal adjustment set

Return to the DAG when the question changes from prediction to intervention.

> Prediction selection depends on the candidate columns, penalty and sample.

> Causal controls follow the target and a defensible identification argument.

> Keep prediction accuracy separate from evidence about an intervention.

Practice

Can prediction lasso justify removing a necessary confounder?

Explanation: No. Causal controls follow the intervention target and identification assumptions.

From flexible models to variance and data problems

From flexible models to variance and data problems

Exact log change

Try the comparison before opening the explanation.

Practice

In a log-outcome model, a policy dummy has coefficient 0.08. Give the usual approximate percentage contrast and the exact multiplicative contrast in the fitted exp(linear predictor).

Exact log change: explanation

Follow the calculation and interpret its units.

> Approximate: +8%. Exact: $100[\exp(0.08) - 1] = 8.3287\%$. This is a contrast in the exponentiated fitted log outcome. A conditional-mean claim in outcome levels additionally needs a justified retransformation treatment of the error distribution.

Elasticity

Try the comparison before opening the explanation.

Practice

In $\log(quantity) = a - 0.7\log(price) + u$, what is the local elasticity? What is the exact fitted quantity change for a 10% price rise?

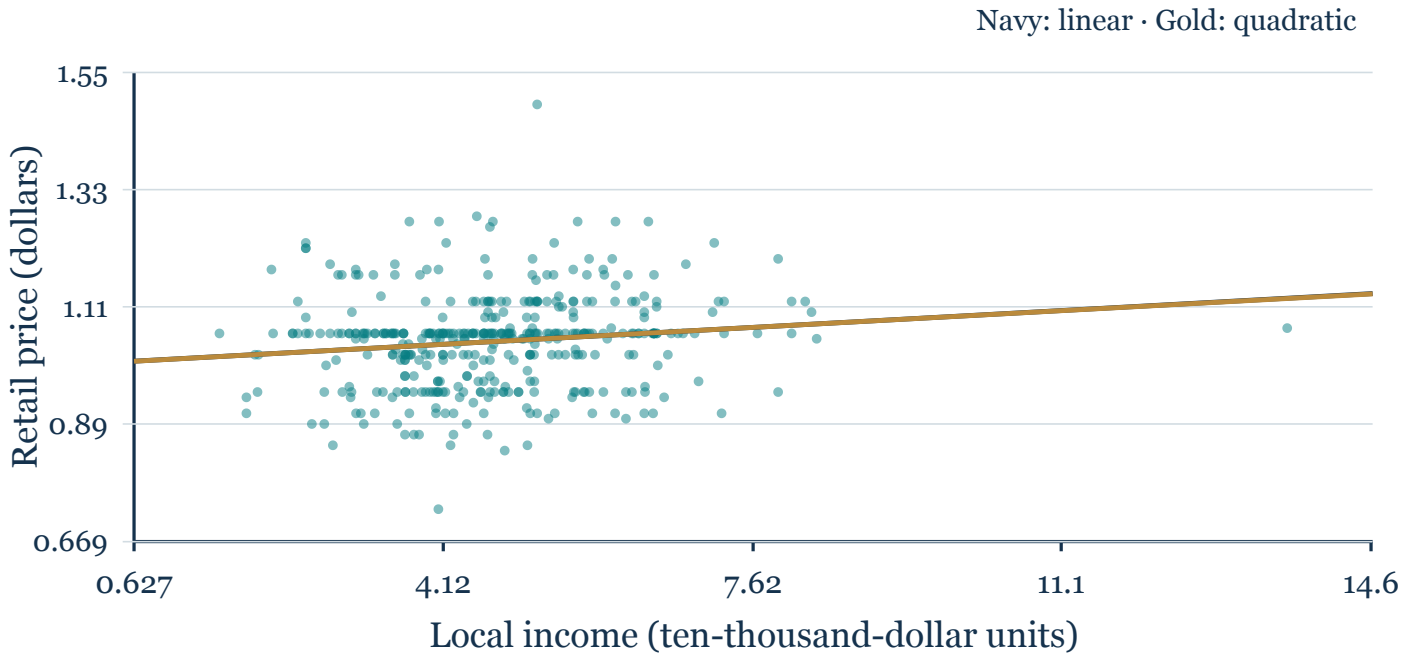
Elasticity: explanation

Follow the calculation and interpret its units.

> The elasticity is −0.7. A 10% price rise multiplies fitted exponentiated log quantity by $1.1^{-0.7}$; the percentage contrast is $100[1.1^{-0.7} - 1]$, about −6.45%, not exactly −7%. Causal interpretation requires price exogeneity or a valid identification strategy.

Compare predictions for the same realistic change

Use the equation, not just the coefficient label.



401 observed positive-price and positive-income records. Lines use the actual level and quadratic regressions; no geographic sampling claim is made.

Predicting a mean and predicting a new outcome have different uncertainty

A new observation brings its own error as well as coefficient uncertainty.

> Mean prediction: $\widehat{m}_0 = x_0' b$; estimated variance $x_0' \widehat{V} x_0$.

> Under the stated independent homoskedastic model, prediction-error variance $\approx x_0' \widehat{V} x_0 + s^2$.

> State whether the target is an average response or a new individual outcome.

Quadratic slope

Try the comparison before opening the explanation.

Practice

For $\widehat{score} = 500 + 8X - 0.2X^2$, give the slope at $X = 10$ and the turning point. What must be checked before using the turning point?

Quadratic slope: explanation

Follow the calculation and interpret its units.

> The slope is $8 - 0.4X$, which equals 4 at $X = 10$. The turning point is $X = 20$, a maximum because the squared coefficient is negative. Check that $X = 20$ lies within the observed and substantively meaningful range; extrapolated turning points are unreliable.

Choose a form for a decision-sized comparison

Keep fit, interpretation and design connected.

> An analyst studying affordability may want dollar price differences.

> Use graphs and stated prediction contrasts to communicate the model.

Adjusted R squared changes the fit penalty, not the identification argument

Compare the same outcome and observations before using a fit criterion.

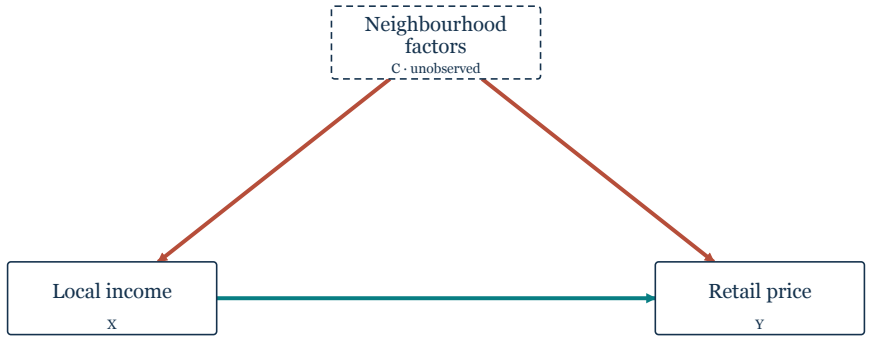
$$\overline{R}^2 = 1 - \frac{SSR/(n-p)}{SST/(n-1)}.$$

An added regressor can raise R^2 and lower adjusted R^2.

Use the same outcome and sample; defend the control set separately.

DAG practice: logs and a common cause

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

Neighbourhood characteristics influence local income and retail prices. Would replacing prices and income by their logarithms close the common-cause route? Explain the distinct roles of the graph and the functional form.

DAG practice: logs and a common cause: explanation

Explain what conditioning changes in this stated graph.

> The route $X \leftarrow C \rightarrow Y$ remains. A numerical transformation does not remove the economic mechanism represented by C. A better fit or a convenient elasticity therefore does not establish causal income variation. The graph states the proposed causal relationships and helps assess controls or alternative designs. The equation chooses how a given comparison is represented numerically and which finite or local contrast is reported. An applied analysis needs both a defensible comparison and a useful specification.

From curvature to indicators and interactions

> From curvature to indicators and interactions

Interaction contrast

Try the comparison before opening the explanation.

Practice

In $Y = a + 0.4X - 0.9(X \times near) + c\,near + u$, near is 0 or 1. What are the x slopes far from and near the border?

Interaction contrast: explanation

Follow the calculation and interpret its units.

> Far: 0.4. Near: $0.4 - 0.9 = -0.5$. The interaction is the difference in slopes, −0.9. The near-group intercept shift c is a different quantity.

Uncertainty of a sum

Try the comparison before opening the explanation.

Practice

In the lesson’s interaction model, the higher-group staffing slope is beta one plus beta three. Can its standard error be obtained by adding the two coefficient standard errors?

Uncertainty of a sum: explanation

Follow the calculation and interpret its units.

> No. Add their variances and twice their covariance, then take the square root. With variances 0.04 and 0.09 and covariance −0.03, the result is the square root of 0.07, about 0.265. Stata’s lincom uses the fitted covariance matrix for the actual model.

An interaction between two continuous inputs changes both partial slopes

The main coefficient is the slope at a specified value of the other input.

> $\widehat{Y} = b_0 + b_1 X + b_2 W + b_3 XW.$

> $\partial \widehat{Y} / \partial X = b_1 + b_3 W, \partial \widehat{Y} / \partial W = b_2 + b_3 X.$

> Evaluate slopes or finite contrasts at explicitly stated input values.

Centering changes coefficient labels, not fitted values

Choose a reference value that helps interpretation.

> $X_i^c = X_i - 20$

Build an empirical comparison before reading stars

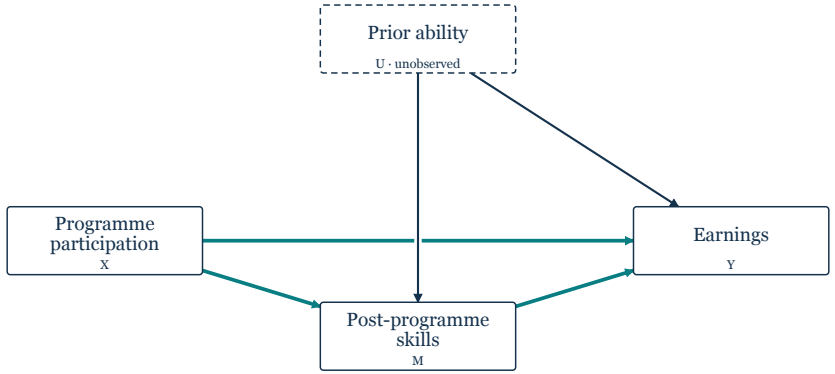
List the coefficient combination and its units.

> For each question, write the required expression: a reference slope, a second-group slope, a slope difference or a group gap at a stated X.

> In a policy memo, report a feasible change within each group and the evidence for any difference between them.

DAG practice: a binary treatment still needs a control argument

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

Programme participation affects post-programme skills and earnings. Prior ability affects skills and earnings. If the programme is initially independent of ability, what two changes can occur when skills are held fixed? Would adding a group interaction automatically resolve them?

DAG practice: a binary treatment still needs a control argument: explanation

Explain what conditioning changes in this stated graph.

> Conditioning on skills blocks the mediated programme route $X \to M \to Y$. It can also open $X \to M \leftarrow U \to Y$, where skills are a collider and ability influences earnings. A regression can consequently mix a changed target with new selection. A group interaction changes the specified shape or subgroup slope. It does not automatically supply the missing causal control argument. Define the intervention and whether its total or a suitable direct effect is wanted, then assess assignment and mediator–outcome assumptions before interpreting the interaction causally.

Probability scale

Try the comparison before opening the explanation.

Practice

An LPM predicts 0.18 at one covariate vector and 0.23 at another. State the change in percentage points and relative percent.

Probability scale: explanation

Follow the calculation and interpret its units.

> The change is 0.05, or 5 percentage points. Relative to 0.18 it is $100 \times 0.05/0.18 = 27.78\%$. These are different scales. The contrast is predictive unless identification assumptions support a causal claim.