What makes an OLS comparison informative and precise?

> Recover the adjusted slope from two sets of residuals.

> Use the slope’s observation weights to explain bias and precision.

First remove controls from the outcome

Save the score differences the controls do not predict.

> Fit score on the control columns; save $rY_i = Y_i - \widehat{Y}_{i,controls}$.

> rY remains in score points.

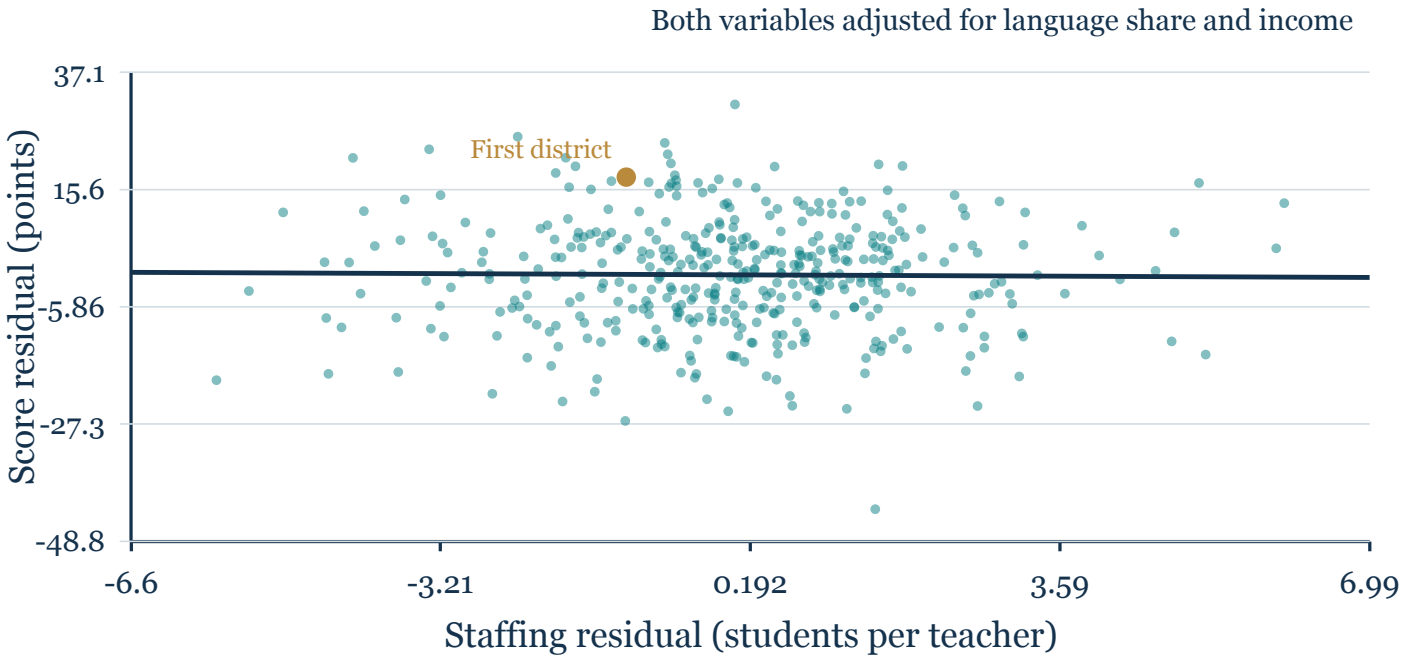
Then remove the same controls from staffing

An adjusted outcome must be compared with an adjusted regressor.

> $$rX_i = X_i - \widehat{E}_{\text{linear}}(X_i \mid Z_i, W_i)$$

Follow one district through both subtractions

The residual plot uses actual score points and staffing units.



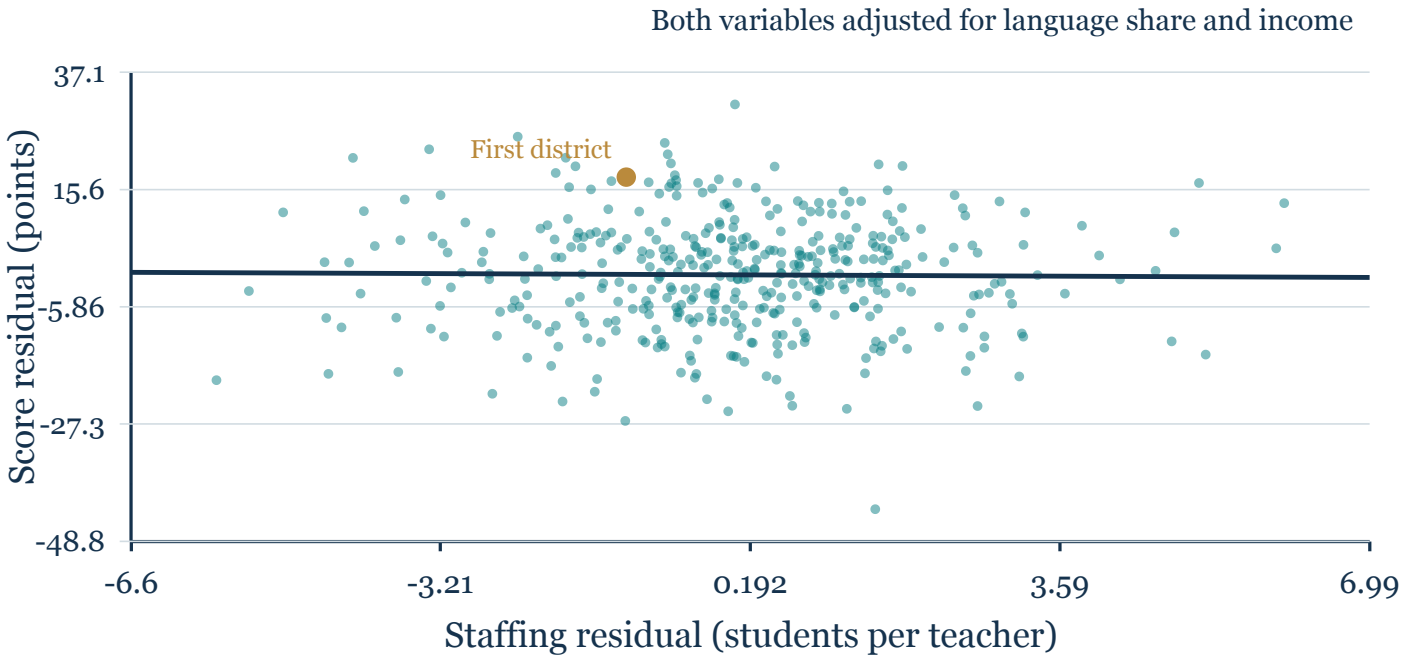
Actual residuals after removing English-learner share and income from both variables. The slope matches the full regression.

Actual first district: $rY \approx 690.80 - 672.92 = 17.88$ score points.

$rX \approx 17.89 - 19.06 = -1.17$ students per teacher.

Regress the two remainders on each other

The partialled slope equals the full-model staffing slope.



Actual residuals after removing English-learner share and income from both variables. The slope matches the full regression.

$$\hat{\beta}_1 = \frac{\sum_i rX_i rY_i}{\sum_i rX_i^2}$$

Read the full regression first

Compare coefficients on the same observations and controls.

```
. quietly regress score str el income

. matrix list e(b), format(%10.6f)

e(b)[1,4]
           str          el      income       _cons
y1   -0.068776   -0.488267    1.494517  640.315497
```

Actual Stata coefficient-only display; reproduce with coefficient_workshop.do.

> Full-model staffing coefficient: −0.068776

> Controls: language share, income and an intercept.

Verify the FWL identity in Stata

The slopes agree to numerical precision.

```
. quietly regress ry rx, noconstant

. matrix list e(b), format(%10.6f)

symmetric e(b)[1,1]
           rx
y1  -0.068776
```

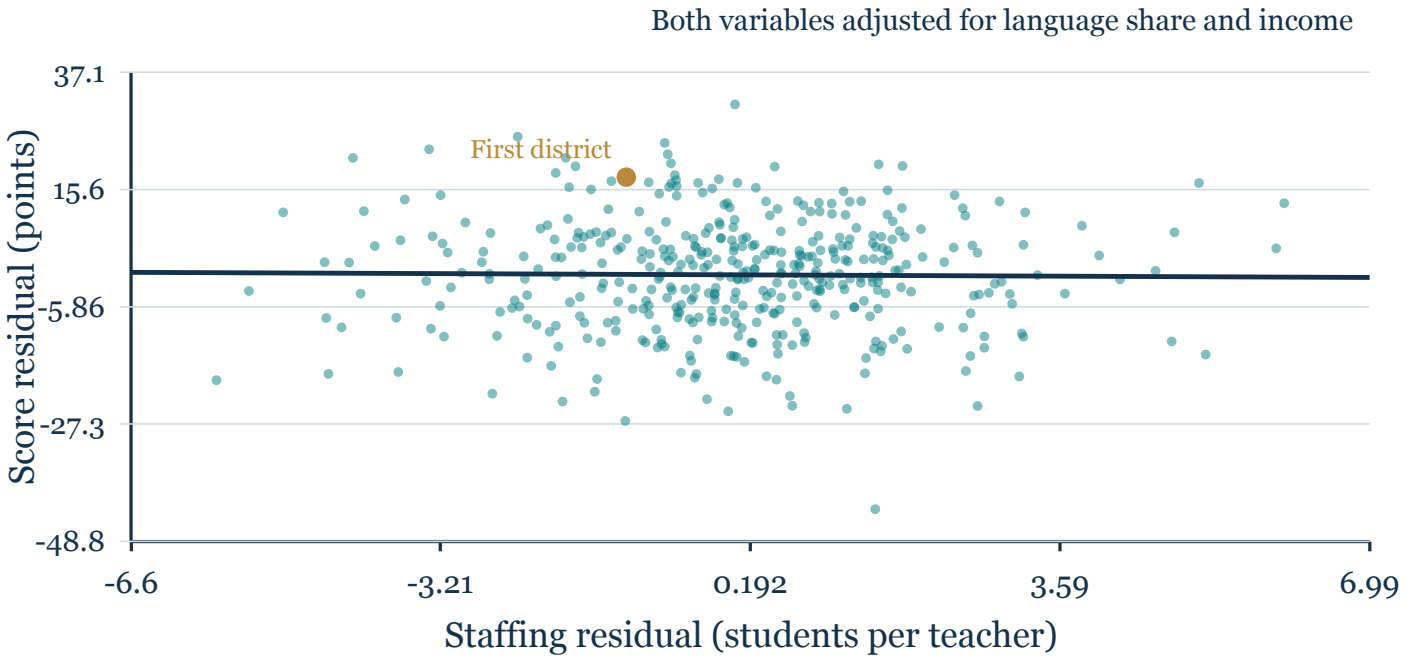
Actual Stata coefficient-only display; reproduce with coefficient_workshop.do.

> FWL: the same adjusted coefficient from compatible residuals.

> A coefficient identity does not make every printed uncertainty calculation identical.

Read the residual axes in real units

They are deviations, not new measurements.



Actual residuals after removing English-learner share and income from both variables. The slope matches the full regression.

Partialling out

Try the comparison before opening the explanation.

Practice

To estimate the schooling coefficient controlling for experience, describe the FWL procedure in words.

Partialling out: explanation

Follow the calculation and interpret its units.

> Regress wages on an intercept and experience and save the residuals. Regress schooling on the same intercept and experience and save those residuals. Regress wage residuals on schooling residuals without another intercept. Its slope equals the schooling coefficient in the full regression on the same observations. Both subtractions must remove exactly the same control columns.

An outcome distribution is not a sampling distribution

The random object changes.

> A distribution describes the possible values of a random quantity and how often they occur.

> These are different objects.

Describe spread before estimating uncertainty

Variance and standard deviation summarize different kinds of variation.

> $\mathrm{Var}(0, 1, 2) = (1 + 0 + 1)/3 = 2/3$

> Standard deviation: $\sqrt{2/3} \approx 0.816$

> Residual degrees of freedom: $n - p$, with $p = k + 1$ including the intercept.

A standard error estimates the estimator’s spread

It is not the standard deviation of the outcome.

> SE: estimated repeated-sample standard deviation of the coefficient.

> Normal reference: about 95% between -1.96 and 1.96.

Unbiasedness concerns repeated samples

It does not promise an accurate estimate in every sample.

> Let $\mathcal{X}$ denote all observed regressor values. Put $w_i = rX_i / \sum_j rX_j^2$.

> $\widehat{\beta}_1 - \beta_1 = \sum_i w_i u_i$.

> If $E(u_i \mid \mathcal{X}) = 0$ for each i, then $E(\widehat{\beta}_1 \mid \mathcal{X}) = \beta_1$.

Precision depends on usable regressor variation

Little variation after adjustment means a weak comparison.

> Given equal conditional variances and zero error correlation across districts, $\mathrm{Var}(\widehat{\beta}_1 \mid \mathcal{X}) = \sigma^2 / \sum_i r X_i^2$.

> $s^2 = \sum_i \widehat{u}_i^2 / (n - p)$, with p counting the intercept.

> Conventional $SE(\widehat{\beta}_1) = \sqrt{s^2 / \sum_i r X_i^2}$.

Multicollinearity reduces separate information

High correlation is different from exact dependence.

> Multicollinearity describes strong linear relationships among regressors.

> For example, including income measured in dollars and the same income divided by one thousand creates exact dependence.

Efficiency needs additional error assumptions

Small variance is a comparison among estimators.

> Linear estimators weight the observed outcomes using the regressors.

> Gauss–Markov gives OLS minimum conditional variance among linear unbiased estimators under its assumptions.

> BLUE means best linear unbiased estimator.

Use VIF as a description, not a decision rule

It measures linear overlap among included regressors.

```
. quietly regress score str el income

. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
      income |      1.14    0.873929
          el |      1.12    0.891206
         str |      1.07    0.931159
-------------+----------------------
    Mean VIF |      1.11
```

> $VIF_j = 1/(1 - R_j^2)$

> R_j^2: fit from predicting regressor j with the other control columns.

> No universal VIF cutoff certifies a causal model.

Reconstruct precision and the degrees of freedom

Try the comparison before opening the explanation.

Practice

Run score on str, el and income using schools.csv. Reconstruct s squared as SSR/(n-p), identify p, and explain the staffing VIF. Then compare the full staffing coefficient with the residual-on-residual coefficient. State which equality is algebraic and which variance claims require additional assumptions.

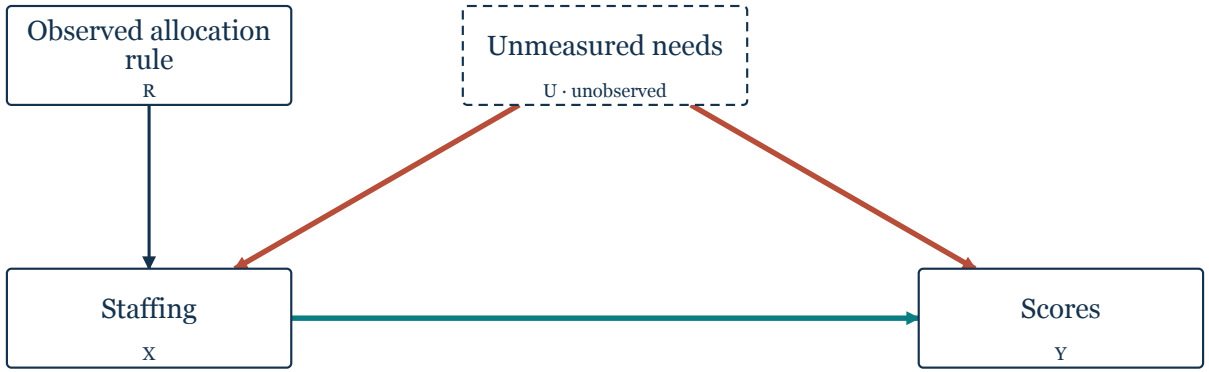
Reconstruct precision and the degrees of freedom: explanation

Follow the calculation and interpret its units.

> The model has an intercept and three slopes, so p is 4 and n minus p is 416. The fitted s squared is about 107.0691, the staffing VIF about 1.07, and both FWL and full staffing coefficients are about -0.068776. The conventional covariance formula requires equal conditional error variance and no cross-observation correlation; it is not established by this equality. The reference overleaf_technical.do prints s squared, VIF and the direct and residualised coefficients.

A residual is not an automatically valid control

Its construction does not establish a causal role.



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

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

Explain what the estimate uses

Describe the remaining comparison in ordinary language.

> For the school application, say that the staffing coefficient uses variation in student–teacher ratios left after their linear relationships with income and language composition are accounted for.

> Then state the remaining assumption: the adjusted staffing variation must not be related to relevant unobserved determinants in the way ruled out by the chosen error restriction.

Verify partialling out without misusing its SE

Reproduce the evidence, then explain its economic meaning.

Practice

Replace income by log income in both residualising equations and in the full model. Verify the new slope identity and report uncertainty from the complete model. Keep your commands, log, output and interpretation.

Verify partialling out without misusing its SE: explanation

A reproducible result needs both a calculation and an interpretation.

> Original staffing slope: −0.068776.

> Using log income: −0.385338, full-model robust SE 0.289855.

> FWL requires the same control columns in both residualising fits.

From OLS properties to statistical inference

> From OLS properties to statistical inference

Residual versus error

Try the comparison before opening the explanation.

Practice

OLS with an intercept gives residuals summing to zero. Does that verify $E(u \mid X) = 0$ in the population?

Residual versus error: explanation

Follow the calculation and interpret its units.

> No. The zero residual sum is a normal-equation identity imposed by least squares. The unobserved population error and its relation to X are a substantive assumption. OLS can satisfy its sample identities in a badly confounded model.

Matrix notation stacks the same regression equations

Define dimensions and operations before using the compact estimator.

> $y = X\beta + u$; $X : n \times p$, $\beta : p \times 1$, $p = k + 1$.

> Normal equations: $X\prime(y - X\widehat{\beta}) = 0$.

> Under full column rank, $\widehat{\beta} = (X\prime X)^{-1} X\prime y$.

The residual-maker gives a compact FWL proof

Remove the same control space from the outcome and target regressor.

> $P_Z = Z(Z\prime Z)^{-1}Z\prime$, $M_Z = I - P_Z$; $M_Z Z = 0$.

> $M_Z\prime = M_Z$, $M_Z^2 = M_Z$.

> $\widehat{\beta}_x = (x\prime M_Z x)^{-1} x\prime M_Z y$.

The estimator-error identity shows where unbiasedness comes from

Separate the algebra from the conditional-mean assumption.

> $\widehat{\beta} - \beta = (X\prime X)^{-1} X\prime u.$

> $E[u \mid X] = 0 \Rightarrow E[\widehat{\beta} \mid X] = \beta.$

> Linearity of the estimator in outcomes and linearity of the model in parameters are distinct statements.

The conventional variance formula uses usable regressor variation

Make the assumptions behind the precision calculation explicit.

> If $\mathrm{Var}(u \mid X) = \sigma^2 I$, $\mathrm{Var}(\widehat{\beta} \mid X) = \sigma^2 (X\prime X)^{-1}$.

> $\mathrm{Var}(\widehat{\beta}_j \mid X) = \dfrac{\sigma^2}{SST_j (1 - R_j^2)}$, $SST_j = \sum_i (X_{ij} - \overline{X}_j)^2$.

> $s^2 = SSR/(n-p)$, $SE_j = \sqrt{s^2 [(X\prime X)^{-1}]_{jj}}$.

Best linear unbiased has a precise comparison class

An efficiency theorem is conditional on a model and a class of estimators.

> BLUE: best linear unbiased estimator under the stated mean, rank and covariance assumptions.

> “Best” means minimum conditional variance within the linear unbiased class.

> Normality is not needed for this theorem.

Collinearity

Try the comparison before opening the explanation.

Practice

Two controls are highly correlated but neither is an exact linear combination of the others. Does this alone create omitted-variable bias?

Collinearity: explanation

Follow the calculation and interpret its units.

> No. Under exogeneity, it need not bias OLS. It can reduce precision because there is little independent regressor variation. Exact collinearity is different: a separate coefficient is not identified in that specification.

Make the computation reproducible

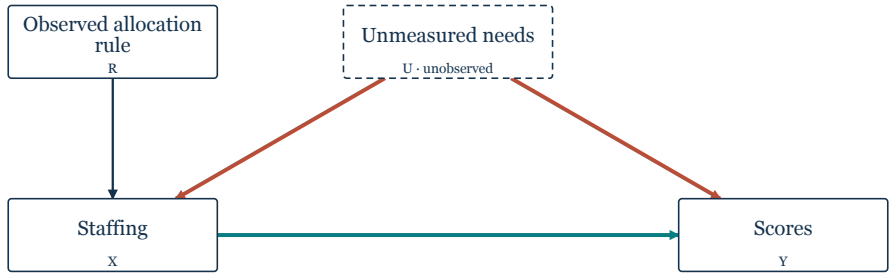
The same sample and control set are essential.

> Record the sample, control columns, residual definitions and coefficient check.

> Reproducibility of a calculation does not establish its causal assumptions.

DAG practice: partialling out and hidden needs

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

The observed allocation rule R helps explain staffing. Unmeasured needs U affect both staffing and scores. After residualising staffing and scores on R, has the graph-based identification problem necessarily disappeared? State the route and distinguish the calculation from the assumption.

DAG practice: partialling out and hidden needs: explanation

Explain what conditioning changes in this stated graph.

> The route $X \leftarrow U \rightarrow Y$ is still open. Removing the fitted contributions of R does not condition on the unobserved common cause U. The FWL coefficient identity can hold exactly while that identification problem remains. Residualisation describes which observed linear variation drives a coefficient. Exogeneity concerns whether the remaining comparison is unrelated to relevant omitted outcome determinants under the model. Those are different statements. A clear empirical report can verify the first by calculation while explaining why the second is, or is not, credible.