F Tests and Causal Model Building

18 Oct 2026 | Chapter 7

Restrictions, loss of fit, and choosing controls

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

Restricted and unrestricted fits

Use the same 420 districts. Score is in test points, STR is students per teacher, EL is in percentage points, and Spend is dollars per pupil.

$$H_0 : \beta_{STR} = \beta_{Spend} = 0.$$

Unrestricted (U): estimate all four coefficients.

$$\widehat{Score}_{U,i} = \widehat{\beta}_0 + \widehat{\beta}_{STR} STR_i + \widehat{\beta}_{EL} EL_i + \widehat{\beta}_{Spend} Spend_i.$$

Restricted (R): impose the null and refit the remaining coefficients.

$$\widehat{Score}_{R,i} = \widetilde{\beta}_0 + \widetilde{\beta}_{EL} EL_i.$$

For either fit, the **sum of squared residuals** is

$$SSR_j = \sum_{i=1}^{420} (Score_i - \widehat{Score}_{j,i})^2, \qquad j = R, U.$$

Restrictions and the least-squares fit

The unrestricted model can reproduce the restricted fit by setting its STR and Spend slopes to zero.

OLS chooses the smallest SSR among this fit and all other unrestricted fits. Therefore,

$$SSR_U \leq SSR_R, \qquad SSR_R - SSR_U \geq 0.$$

Removing restrictions gives OLS more possible fits. Even irrelevant regressors can fit some sampling noise.

The F statistic asks whether the improvement is large relative to residual variation. A positive improvement alone is insufficient.

Discuss: would this comparison work if the regressions used different districts?

Stata: the residual SS is the quantity we need

Unrestricted: STR, EL, and Spend, with an intercept.

```
. regress score str el spend
      Source |       SS           df       MS      Number of obs   =       420
-------------+----------------------------------   F(3, 416)       =    107.45
    Residual |  85699.7099       416  206.008918   R-squared       =    0.4366
-------------+----------------------------------   Adj R-squared   =    0.4325
```

Restricted: EL and an intercept.

```
. regress score el
      Source |       SS           df       MS      Number of obs   =       420
-------------+----------------------------------   F(1, 418)       =    296.40
    Residual |  89000.0248       418   212.91872   R-squared       =    0.4149
-------------+----------------------------------   Adj R-squared   =    0.4135
```

Imposing the null adds $89{,}000.0248 - 85{,}699.7099 = 3{,}300.3149$ to SSR.

The SSR construction of the classical F statistic

Let q be the number of restrictions and k_U the number of unrestricted slopes, excluding the intercept. Here $q = 2$, $k_U = 3$, and $n = 420$ districts.

$$F_{\text{classical}} = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n - k_U - 1)}.$$

- Numerator: extra squared error per restriction, $3300.3149/2 = 1650.1575$.
- Denominator: unrestricted residual mean square, $85699.7099/416 = 206.0089$.

$$F_{\text{classical}} = \frac{1650.1575}{206.0089} = 8.01.$$

The fit loss per restriction is about eight times the residual mean square. This formula requires homoskedasticity for valid classical inference.

Fit loss, robust inference, and the choice of controls

For the same school-resource null, Stata gives:

```
. regress score str el spend
. test str spend
       F(  2,   416) =    8.01
            Prob > F =    0.0004
. regress score str el spend, vce(robust)
. test str spend
       F(  2,   416) =    5.43
            Prob > F =    0.0047
```

The coefficients and SSRs are unchanged. The uncertainty calculation changes. The SSR shortcut gives the classical result, not the robust result.

Neither result decides which controls give a credible causal comparison.

Return to the hospital question: for the total effect of adding nurses on waiting time, should we control for prior demand, newly created capacity, or both?

DAG vocabulary: variables, arrows, and no feedback loop

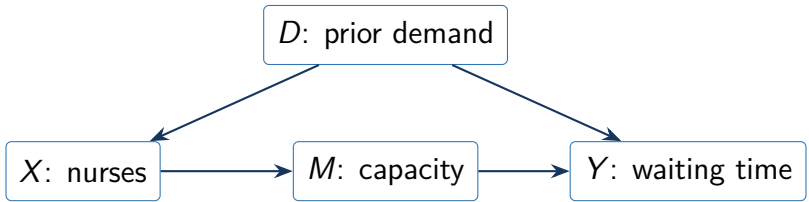
Directed acyclic graph (DAG)

A set of **nodes** (variables) joined by **directed edges** (arrows), with no directed cycle.

- $X \to Y$: we assume that changing X can affect Y directly.
- **Acyclic**: following the arrows cannot lead back to where we started.
- No arrow from X to Y means no *direct* effect in this graph. An indirect effect may still exist.

Draw causes before effects. Show plausible unobserved common causes too.

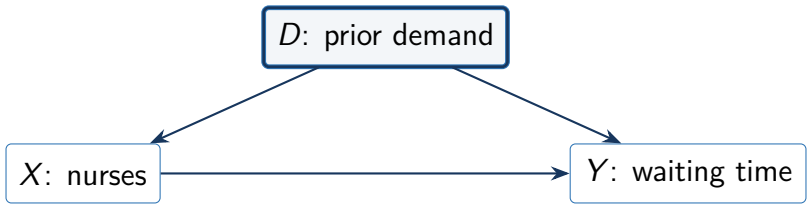
Parents, descendants, and two kinds of path



- **Parent/child**: one arrow connects them. Nurses are a parent of capacity.
- **Ancestor/descendant**: follow one or more arrows. Waiting time is a descendant of nurses through capacity.
- A **path** is a route along connected lines; we may trace it against an arrow.

Causal path: $X \to M \to Y$. Back-door path: $X \leftarrow D \to Y$.

A fork: control for a shared cause

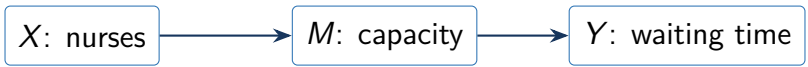


The **fork** $X \leftarrow D \rightarrow Y$ can create a noncausal association.

- Busy hospitals may hire more nurses *and* have longer waits.
- **Control for** D: compare hospitals with similar prior demand. This blocks the shared-cause path, not the staffing effect.

Here D is a **confounder**: a shared cause of staffing and waiting time. A double outline means we control for it.

A chain: a mediator carries part of the effect

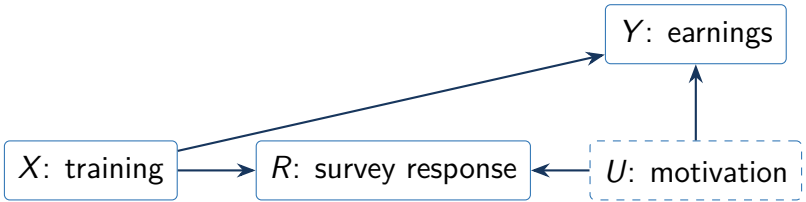


$X \to M \to Y$ is a **chain**; M is a **mediator** of the staffing effect.

- The **total effect** includes changes that work through capacity.
- Controlling for capacity holds that channel fixed, so it no longer measures the full staffing effect.

To study a direct effect with capacity held fixed, we would need a different question and additional assumptions.

A collider: predict what sample selection will do

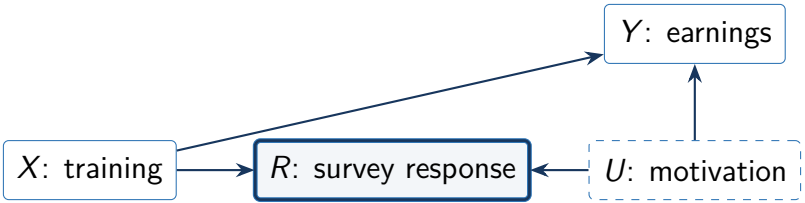


Training is randomly assigned. Motivation affects earnings and the chance of answering the evaluation survey; training also affects survey response.

Pair discussion – 2 minutes

On $X \to R \leftarrow U \to Y$, is R a common cause or a common effect? What might change if we keep only respondents ($R = 1$)?

Debrief: a common effect is not a confounder



R is a **collider**: two arrows on the path point into it.

- Before conditioning, R blocks $X \to R \leftarrow U \to Y$.
- Selecting $R = 1$ opens this path and can associate training with motivation among respondents.

Random assignment in the full sample does not guarantee a valid respondents-only comparison.

Two rules for choosing controls

Rule 1: all back-door paths must be blocked. We want to close the unwanted routes without blocking the causal effect we are studying.

Rule 2: controlling for the middle variable flips the path.

Pattern	Before control	After control
Chain: $X \to M \to Y$	Open	Closed
Fork: $X \leftarrow D \to Y$	Open	Closed
Collider: $X \to S \leftarrow Y$	Closed	Open

Apply the flip rule at the variable you control. If another variable still blocks the route, the whole path stays closed.

From graph to controls: the back-door criterion

Back-door path: a route from X to Y that starts with an arrow pointing *into* X.

The **back-door criterion** gives a rule for choosing controls for a total effect:

1. Do not include variables caused by X, directly or indirectly.
2. Choose controls that block **every back-door path**, without opening another one.

For our hospital graph: control for prior demand D, not capacity M.

A control must improve the causal comparison. A small p-value or a higher R^2 is not a reason by itself. This rule works if the assumed graph is adequate.

Group task: draw the DAG, then choose the model

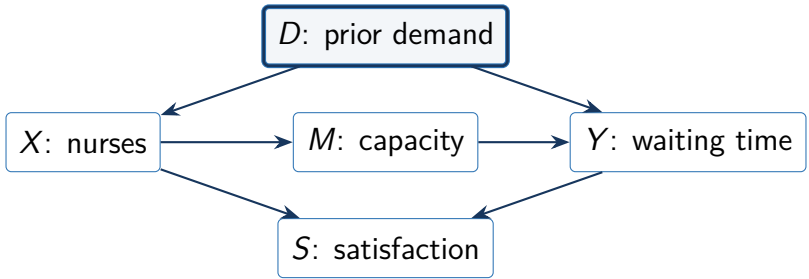
Hospital model-building task – 6 minutes

For this exercise, assume the following story is complete:

- Prior demand D affects nurses X and waiting time Y.
- Nurses affect capacity M, which affects waiting time.
- Nurses and waiting time both affect later patient satisfaction S.

1. Draw the arrows. Which set of D, M, S would you control for to estimate the total staffing effect?
2. Trace a path that explains each inclusion or exclusion.

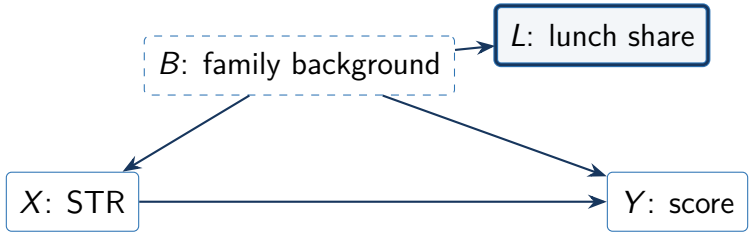
Debrief: choose demand, not every measured variable



- Include D: block $X \leftarrow D \rightarrow Y$.
- Leave M out for the total effect: preserve $X \rightarrow M \rightarrow Y$.
- Leave S out: do not open $X \rightarrow S \leftarrow Y$.

Choose prior demand D as the control, if this causal story is adequate.

A background proxy is not the confounder itself



Here L is the observed lunch share; B is unobserved family background. The double outline conditions on L.

- The path $X \leftarrow B \rightarrow Y$ does *not* pass through L.
- Therefore conditioning on L does not block that path in this DAG.

Lunch share may be informative, but better fit is not proof of no remaining confounding.

Turn the hospital graph into a regression

For hospital i, let $Waiting_i$ be average waiting time in minutes, $Nurses_i$ staffing, and $Demand_i$ prior demand.

$$Waiting_i = \beta_0 + \beta_1 Nurses_i + \beta_2 Demand_i + u_i.$$

- β_0 is the intercept; β_2 adjusts for demand; u_i contains other influences on waiting time.
- β_1 compares hospitals with one more nurse, *holding prior demand fixed*. We leave capacity and satisfaction out.

To call β_1 a causal effect, we must defend the graph and the regression assumptions. We also need hospitals with different staffing but comparable demand.

Group recommendation: separate the evidence from the model

School-policy discussion – 4 minutes

A report states: "The joint F test is significant, so smaller classes and higher spending will raise scores. We should control for every available variable to be safe."

1. State what the F result actually supports.
2. Use the family-background DAG to explain what remains uncertain.
3. Use the hospital DAG to explain why "control for everything" can fail.

Write a two-sentence replacement for the report's recommendation.

Debrief: a test result and a causal argument are different

A defensible replacement

The data reject the claim that both resource coefficients are zero in the specified regression. A causal recommendation still requires credible controls and a suitable model; including every measured variable is not a solution.

- Close shared-cause paths, keep total-effect pathways, and avoid opening collider paths.
- The graph guides *which variables* belong. It does not tell us whether their relationship is a straight line.

Next class: keep estimation and testing, but allow the fitted relationship to curve.

Optional B1: Full school regression output

STR is students per teacher; EL is percentage points; spending is dollars per pupil. The sample is 420 districts.

```
. regress score str el spend, vce(robust)
Linear regression                               Number of obs     =        420
                                                R-squared         =     0.4366
------------------------------------------------------------------------------
             |               Robust
       score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234002     .661203
          el |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       spend |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

The graph has *not* justified this three-regressor model as causal.

Optional B2: A combined policy contrast

$$\Delta = -2\beta_{STR} + 1000\beta_{Spend}, \qquad \widehat{\Delta} = 4.441.$$

```
. lincom -2*str + 1000*spend
 ( 1)  - 2*str + 1000*spend = 0
------------------------------------------------------------------------------
       score | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |     4.4407   1.351619     3.29   0.001     1.783845    7.097555
------------------------------------------------------------------------------
```

`lincom` uses the coefficient variances *and covariance*. The 95% interval is $[1.784, 7.098]$ score points, not a guaranteed causal gain.

Optional B3: Scalar covariance calculation

The combined change is $\widehat{\Delta} = -2\widehat{\beta}_{STR} + 1000\widehat{\beta}_{Spend}$.

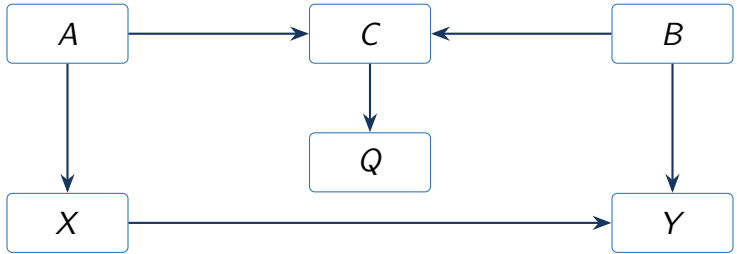
$$\widehat{Var}(\widehat{\Delta}) = 4\widehat{Var}(\widehat{\beta}_{STR}) + 1000^2\widehat{Var}(\widehat{\beta}_{Spend}) - 4000\widehat{Cov}(\widehat{\beta}_{STR}, \widehat{\beta}_{Spend}).$$

The three contributions from the school regression are

$$0.9296 + 2.4987 - 1.6014 = 1.8269, \qquad SE = \sqrt{1.8269} = 1.3516.$$

Do not add standard errors. Stata also accounts for how the two coefficient estimates move together.

Optional B4: A collider descendant before treatment



Assume A, B, C, Q are measured before X. On $X \leftarrow A \rightarrow C \leftarrow B \rightarrow Y$, C is a collider and Q its descendant.

- With neither conditioned on, this back-door path is blocked.
- Conditioning on C *or* Q opens it, unless another node blocks it.

Pre-treatment timing alone does not make a variable a good control.