Multiple Regression Applications in Stata

05 Oct 2026 | Chapter 6

ECO 351 – Introduction to Econometrics

Shi Qiu – Fall 2026

Recall: what did our previous comparison leave open?

In Lecture 12, moving from STR 20 to STR 18 at the same English-learner share gave a fitted score increase of 2.20 points.

A question from the school board

"Can we obtain that improvement without increasing spending per student?

1. Was spending held fixed in the regression we used?
2. What information would we need before answering the board?

The earlier result does not answer the budget question

What the earlier comparison held fixed

English-learner share. Spending could differ between the districts being compared.

What the board now wants to hold fixed

Both English-learner share and spending per student. Hiring more teachers may require cuts to other uses of the budget.

Today we will use a sequence of regression results to distinguish these questions and assess what the evidence supports.

Two plans for the same class-size reduction

The superintendent proposes two fewer students per teacher. Keep the English-learner share unchanged in both comparisons.

Plan A: reallocate the existing budget

Hire more teachers but reduce other spending. Total spending per student stays the same.

$$\Delta STR = -2, \qquad \Delta Spend = 0$$

Plan B: provide additional resources

Hire more teachers with a spending increase of 1,000 dollars per student.

$$\Delta STR = -2, \qquad \Delta Spend = 1000$$

Plan the columns before opening Stata

Column	Regressors	Purpose
1	STR	Record the uncontrolled comparison.
2	STR, EL	Retain the language-composition control from Lecture 12.
3	STR, EL, Spend	Separate the class-size and spending contributions.

Keep the comparison transparent

Use Score as the outcome in every column. Keep the observations fixed. Report which controls are present, not just which coefficients have small p-values.

Stata: fix the sample before comparing models

Use the same districts in every column. Check N and the units before interpreting a coefficient.

```
do "setup.do" schools
keep if !missing(score,str,el,spend)
count
summarize score str el spend
```

Read the result

420 districts; EL is in percentage points; Spend is dollars per student.

Stata output: the common sample and variable units

```
. keep if !missing(score,str,el,spend)
(0 observations deleted)

. count
  420

. summarize score str el spend
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       score |        420    654.1565    19.05335      605.55     706.75
         str |        420    19.64043    1.891812          14       25.8
          el |        420    15.76816    18.28593           0   85.53972
       spend |        420    5312.408    633.9371     3926.07   7711.507
```

Stata: estimate the three comparisons

Each command adds the controls from one equation. Store each result so it can be compared later.

```
regress score str, vce(robust)
estimates store m1
regress score str el, vce(robust)
estimates store m2
regress score str el spend, vce(robust)
estimates store m3
```

Read the result

On the next three pages, follow the STR row: coefficient, standard error, then p-value.

Model 1: the uncontrolled class-size comparison

```
. regress score str, vce(robust)
Number of obs     =         420

------------------------------------------------------------------------------
             |               Robust
       score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |  -2.279808   .5194892    -4.39   0.000    -3.300945   -1.258671
       _cons |   698.933   10.36436     67.44   0.000     678.5602    719.3057
------------------------------------------------------------------------------
```

Model 2: hold language composition fixed

```
. regress score str el, vce(robust)
Number of obs     =         420

------------------------------------------------------------------------------
             |               Robust
       score | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         str |  -1.101296   .4328472    -2.54   0.011      -1.95213   -.2504616
          el |  -.6497768   .0310318   -20.94   0.000      -.710775   -.5887786
       _cons |   686.0322   8.728224    78.60   0.000      668.8754     703.189
------------------------------------------------------------------------------
```

Model 3: also hold spending fixed

```
. regress score str el spend, vce(robust)
Number of obs     =         420

------------------------------------------------------------------------------
             |               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
------------------------------------------------------------------------------
```

Read the units before interpreting the spending slope

The fitted third model, rounded for reading, is

$$\widehat{Score} = 649.578 - 0.2864 STR - 0.6560 EL + 0.003868 Spend.$$

A meaningful spending comparison

$$\Delta Spend = \$1{,}000 \quad \Rightarrow \quad \Delta \widehat{Score} = 0.003868(1000) = 3.868$$

The fitted score is about 3.87 points higher, holding STR and EL fixed.

The coefficient is per dollar, not per thousand dollars and not a percentage change.

Stata: put the stored results side by side

A coefficient appears first; its standard error is printed directly below. Read the controls and N with it.

```
estimates table m1 m2 m3, ///
    b(%9.3f) se(%9.3f) stats(N r2)
```

Read the result

The three columns use the same 420 districts. A blank cell means that regressor was not included.

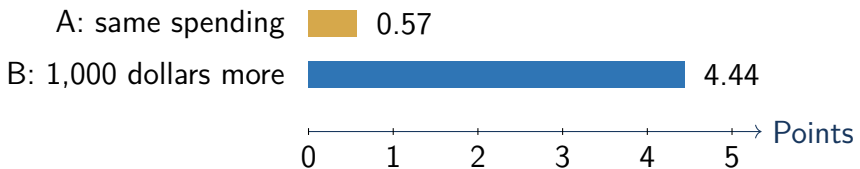
Stata output: compare all three stored models

```
. estimates table m1 m2 m3, ///
>     b(%9.3f) se(%9.3f) stats(N r2)

----------------------------------------------------
    Variable |     m1          m2          m3
-------------+--------------------------------------
         str |   -2.280      -1.101      -0.286
             |    0.519       0.433       0.482
          el |               -0.650      -0.656
             |                0.031       0.032
       spend |                           0.004
             |                           0.002
       _cons |  698.933     686.032     649.578
             |   10.364       8.728      15.458
-------------+--------------------------------------
           N |      420         420         420
          r2 |    0.051       0.426       0.437
----------------------------------------------------
                                      Legend: b/se
```

Use the full model to compare the two budget plans

$$\widehat{\Delta Score} = -0.2864\Delta STR + 0.003868\Delta Spend \quad (\Delta EL = 0)$$



Where the difference comes from

Both plans reduce STR by 2. Plan B adds the fitted spending contribution: $0.57 + 3.87 = 4.44$ points.

Your turn: assess a smaller budget increase

Use model 3: the STR slope is -0.2864. The spending slope per dollar is 0.0038679, with SE 0.0015807.

1. A third plan reduces STR by 2 and increases spending by 500 dollars per student, with EL unchanged. Calculate the fitted score change.
2. Let δ_3 be the population spending slope in model 3. Test $H_0 : \delta_3 = 0$ against a two-sided alternative at 5%.
3. Give a large-sample 95% interval for the score change associated with 1,000 dollars more spending, at fixed STR and EL.

Use 1.96. Work for two minutes before we discuss the answer.

Solution: separate the plan from the spending test

$$\Delta\widehat{Score} = (-0.2864)(-2) + 0.0038679(500) = 2.51 \text{ points}$$

$$t = \frac{0.0038679}{0.0015807} = 2.45 \quad \Rightarrow \quad \text{reject } H_0 : \delta_3 = 0 \text{ at } 5\%.$$

$$\text{CI per 1,000 dollars :} \quad 3.8679 \pm 1.96(1.5807) = [0.77,\ 6.97]$$

What the interval describes

A spending increase at fixed STR and EL. It is not the interval for the plan that changes both class size and spending.

Write a conclusion the school board can use

State the comparison

At fixed EL and spending, two fewer students per teacher correspond to a fitted increase of about 0.57 score point.

State the uncertainty

The STR p-value is 0.553. We cannot distinguish its fixed-budget slope from zero at 5%; this does not prove the slope is zero.

State the limitation

These district comparisons do not by themselves show what an actual intervention would achieve. Teacher quality, for example, might still differ systematically across districts.