ECON 640: Multiple Regression Model

Yang Liang

Department of Economics
San Diego State University

Multiple Regression

- So far, we have focused on the situation where we have one dependent and one explanatory variable.

Multiple Regression

- So far, we have focused on the situation where we have one dependent and one explanatory variable.
- The real power of regression analysis, however, is that we are able to hold other factors constant when estimating correlations.

Multiple Regression

- So far, we have focused on the situation where we have one dependent and one explanatory variable.
- The real power of regression analysis, however, is that we are able to hold other factors constant when estimating correlations.
- These ceteris paribus comparisons help get us closer to establishing causal relationships.

Multiple Regression

- So far, we have focused on the situation where we have one dependent and one explanatory variable.
- The real power of regression analysis, however, is that we are able to hold other factors constant when estimating correlations.
- These ceteris paribus comparisons help get us closer to establishing causal relationships.
- For this, we need to expand our model to allow for more than one explanatory variable - which changes quite a bit in how we think about OLS.

Multiple Regression - Basic framework

- In the multiple regression model, the population regression model is given by

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_i$$

Multiple Regression - Basic framework

- In the multiple regression model, the population regression model is given by

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_i$$

- This framework allows us to estimate the effect on Y_i of changing one variable (X_{1i} for example), while holding the other regressors ($X_{2i}, X_{3i}, ..$) constant.

Multiple Regression - Basic framework

- In the multiple regression model, the population regression model is given by

$$Y_i = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + u_i$$

- This framework allows us to estimate the effect on Y_i of changing one variable (X_{1i} for example), while holding the other regressors ($X_{2i}, X_{3i}, ..$) constant.
- The population regression line (the relationship that holds between Y and the X's on average) is given by

$$E\left(Y_i \mid X_{1i}, ..., X_{ki}\right) = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki}$$

  and the error term u_i by

$$u_i \equiv Y_i - E\left(Y_i \mid X_{1i}, ..., X_{ki}\right)$$

Multiple Regression

- As in the univariate case, β_0 is the intercept and β_k is the slope coefficient of X_k.
- Multivariate OLS estimates the parameters (the β's) in the same way as univariate OLS: by minimizing the sum of the squared errors (prediction mistakes)

$$\widehat{u}_i = Y_i - \widehat{Y}_i$$

where $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + ... + \widehat{\beta}_k X_{ki}$

- OLS minimizes the sum of the squared errors $\left(\sum \widehat{u}_i^2\right)$, yielding explicit formulas for the estimators $\widehat{\beta}_0, \widehat{\beta}_1, ..., \widehat{\beta}_k$
- As before, pending some assumptions, the OLS estimates will be unbiased, consistent, and asymptotically normal.

The OLS Assumptions - Multivariate

OLS Assumption 1 Linearity (in parameters)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \text{ is the DGP}$$

The OLS Assumptions - Multivariate

OLS Assumption 1 **Linearity (in parameters)**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \text{ is the DGP}$$

OLS Assumption 2 Simple random sample

$$(X_{1i}, X_{2i}, X_{3i}, \ldots, X_{ki}, Y_i) \text{ are } \textit{iid} \text{ draws from their joint distribution}$$

The OLS Assumptions - Multivariate

OLS Assumption 1 Linearity (in parameters)

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$ is the DGP

OLS Assumption 2 Simple random sample

$(X_{1i}, X_{2i}, X_{3i}, \ldots, X_{ki}, Y_i)$ are *iid* draws from their joint distribution

OLS Assumption 3 No perfect multicolinearity

No X can be expressed as a perfect linear function of other Xs

The OLS Assumptions - Multivariate

OLS Assumption 1 **Linearity (in parameters)**

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$ is the DGP

OLS Assumption 2 Simple random sample

$(X_{1i}, X_{2i}, X_{3i}, \ldots, X_{ki}, Y_i)$ are *iid* draws from their joint distribution

OLS Assumption 3 **No perfect multicolinearity**

No X can be expressed as a perfect linear function of other Xs

OLS Assumption 4 **Zero Conditional Mean**

$\mathbb{E}(u|x_1, x_2, x_3, \ldots, x_k) = 0$

The OLS Assumptions - Multivariate

OLS Assumption 1 Linearity (in parameters)

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$ is the DGP

OLS Assumption 2 Simple random sample

$(X_{1i}, X_{2i}, X_{3i}, \ldots, X_{ki}, Y_i)$ are *iid* draws from their joint distribution

OLS Assumption 3 No perfect multicolinearity

No X can be expressed as a perfect linear function of other Xs

OLS Assumption 4 Zero Conditional Mean

$\mathbb{E}(u|x_1, x_2, x_3, \ldots, x_k) = 0$

OLS Assumption 5 Homoskedasticity

$Var(u|x_1, x_2, x_3, \ldots, x_k) = \sigma_u^2$

Imperfect Multicollinearity

- OLS A3 rules out perfect collinearity.

Imperfect Multicollinearity

- OLS A3 rules out perfect collinearity.
- But what if two or more of the regressors are highly (but not perfectly) collinear?

Imperfect Multicollinearity

- OLS A3 rules out perfect collinearity.
- But what if two or more of the regressors are highly (but not perfectly) collinear?
- In other words, what if there's a linear function of the regressors that is highly correlated with another regressor?

Imperfect Multicollinearity

- OLS A3 rules out perfect collinearity.
- But what if two or more of the regressors are highly (but not perfectly) collinear?
- In other words, what if there's a linear function of the regressors that is highly correlated with another regressor?
    - Example: Hedonic Regression - Regressing the price of a good on its characteristics (e.g. cars, diamonds, houses)

Imperfect Multicollinearity

- OLS A3 rules out perfect collinearity.
- But what if two or more of the regressors are highly (but not perfectly) collinear?
- In other words, what if there's a linear function of the regressors that is highly correlated with another regressor?
  - Example: Hedonic Regression - Regressing the price of a good on its characteristics (e.g. cars, diamonds, houses)
    - How different are horsepower and acceleration?
- This situation, called imperfect multicollinearity, does not pose any problem for the theory of the OLS estimators.

Imperfect Multicollinearity

- OLS A3 rules out perfect collinearity.
- But what if two or more of the regressors are highly (but not perfectly) collinear?
- In other words, what if there's a linear function of the regressors that is highly correlated with another regressor?
  - Example: Hedonic Regression - Regressing the price of a good on its characteristics (e.g. cars, diamonds, houses)
    - How different are horsepower and acceleration?
- This situation, called imperfect multicollinearity, does not pose any problem for the theory of the OLS estimators.
- In fact, a purpose of OLS is to sort out the independent effects of the various regressors when they are potentially correlated.

Imperfect Multicollinearity

- However, with a high degree of imperfect multicollinearity among the regressors, $\sigma_{\widehat{\beta}_j}$ will tend to be high (at least among the correlated regressors).

Imperfect Multicollinearity

- However, with a high degree of imperfect multicollinearity among the regressors, $\sigma_{\widehat{\beta}_j}$ will tend to be high (at least among the correlated regressors).
  - This can make it difficult to obtain precise estimates of the separate effects (at least in small samples).

Imperfect Multicollinearity

- However, with a high degree of imperfect multicollinearity among the regressors, $\sigma_{\widehat{\beta}_j}$ will tend to be high (at least among the correlated regressors).
  - This can make it difficult to obtain precise estimates of the separate effects (at least in small samples).
- Why? Recall that a coefficient is an estimate of the partial effect of one regressor, holding the other ones constant.

Imperfect Multicollinearity

- However, with a high degree of imperfect multicollinearity among the regressors, $\sigma_{\widehat{\beta}_j}$ will tend to be high (at least among the correlated regressors).
  - This can make it difficult to obtain precise estimates of the separate effects (at least in small samples).
- Why? Recall that a coefficient is an estimate of the partial effect of one regressor, holding the other ones constant.
  - If the regressors tend to move together, this effect will be hard to estimate precisely.

Another example

Suppose we add *EXPN_STU* (the total annual expenditures per pupil in the district in dollars) to our regression

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

95% CI for $\beta_1 : -.29 \pm 1.96 \cdot .48 = (-1.23, .65)$

Another example

Suppose we add *EXPN_STU* (the total annual expenditures per pupil in the district in dollars) to our regression

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

95% CI for $\beta_1 : -.29 \pm 1.96 \cdot .48 = (-1.23, .65)$

90% CI for $\beta_1 : -.29 \pm 1.645 \cdot .48 = (-1.08, .50)$

Another example

Suppose we add *EXPN_STU* (the total annual expenditures per pupil in the district in dollars) to our regression

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

95% CI for $\beta_1 : -.29 \pm 1.96 \cdot .48 = (-1.23, .65)$

90% CI for $\beta_1 : -.29 \pm 1.645 \cdot .48 = (-1.08, .50)$

p-value for $\beta_1 = 2\Phi\left(-\left|\frac{-.29-0}{.48}\right|\right) = 2\Phi(-.6) = .546$

Another example

Suppose we add *EXPN_STU* (the total annual expenditures per pupil in the district in dollars) to our regression

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

95% CI for $\beta_1 : -.29 \pm 1.96 \cdot .48 = (-1.23, .65)$

90% CI for $\beta_1 : -.29 \pm 1.645 \cdot .48 = (-1.08, .50)$

p-value for $\beta_1 = 2\Phi\left(-\left|\frac{-.29-0}{.48}\right|\right) = 2\Phi(-.6) = .546$

p-value for $\beta_2 = 2\Phi\left(-\left|\frac{.39-0}{.0016}\right|\right) = 2\Phi(-2.45) = .0143$

Another example

- This new result

$$\widehat{TS} = \underset{(15.5)}{649.6} - \underset{(.48)}{.29}\,STR + \underset{(.002)}{.004}\,EXPN\_STU - \underset{(.032)}{.656}\,EL\_PCT, \;\; R^2 = .44$$

  is a very different from what we found earlier

$$\widehat{TS} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28} \cdot STR, \;\; R^2 = .05$$

- So what's going on?

Another example

- This new result

$$\widehat{TS} = \underset{(15.5)}{649.6} - \underset{(.48)}{.29}\, STR + \underset{(.002)}{.004}\, EXPN\_STU - \underset{(.032)}{.656}\, EL\_PCT, \;\; R^2 = .44$$

  is a very different from what we found earlier

$$\widehat{TS} = \underset{(10.4)}{698.9} - \underset{(0.52)}{2.28} \cdot STR, \;\; R^2 = .05$$

- So what's going on?
- Teacher salaries are a big component of EXPN STU, leading to a large negative correlation between STR and EXPN STU.
- To have more teachers per pupil you need to increase expenditure. If you don't you need to cut other expenses.
- The model is having a hard time separating their effects

Goodness of Fit

- You probably noticed that the R^2 went way up when we added $EXPN\_STU$ to the Test Scores regression.

Goodness of Fit

- You probably noticed that the R^2 went way up when we added $EXPN\_STU$ to the Test Scores regression.
- If we use our reasoning from the univariate case, $R^2 = .44$ means that we are explaining about 44% of the variation in test scores with our regression, up from about 5% with only STR.

Goodness of Fit

- You probably noticed that the R^2 went way up when we added $EXPN\_STU$ to the Test Scores regression.
- If we use our reasoning from the univariate case, $R^2 = .44$ means that we are explaining about 44% of the variation in test scores with our regression, up from about 5% with only STR.
- So it seems that we are doing a much better job of explaining test scores now. But is this logic okay?

Goodness of Fit

- You probably noticed that the R^2 went way up when we added *EXPN_STU* to the Test Scores regression.
- If we use our reasoning from the univariate case, $R^2 = .44$ means that we are explaining about 44% of the variation in test scores with our regression, up from about 5% with only *STR*.
- So it seems that we are doing a much better job of explaining test scores now. But is this logic okay?
- The answer is yes, but we need to revisit the notion of *goodness of fit* to convince ourselves of why.

Goodness of Fit

- You probably noticed that the R^2 went way up when we added *EXPN_STU* to the Test Scores regression.
- If we use our reasoning from the univariate case, $R^2 = .44$ means that we are explaining about 44% of the variation in test scores with our regression, up from about 5% with only *STR*.
- So it seems that we are doing a much better job of explaining test scores now. But is this logic okay?
- The answer is yes, but we need to revisit the notion of *goodness of fit* to convince ourselves of why.
- To assess goodness of fit, we ask how well our OLS estimates fit the data?
- There are 3 ways to answer this question

Goodness of Fit

- You probably noticed that the R^2 went way up when we added $EXPN\_STU$ to the Test Scores regression.
- If we use our reasoning from the univariate case, $R^2 = .44$ means that we are explaining about 44% of the variation in test scores with our regression, up from about 5% with only STR.
- So it seems that we are doing a much better job of explaining test scores now. But is this logic okay?
- The answer is yes, but we need to revisit the notion of *goodness of fit* to convince ourselves of why.
- To assess goodness of fit, we ask how well our OLS estimates fit the data?
- There are 3 ways to answer this question
    - The standard error of the regression (SER)

Goodness of Fit

- You probably noticed that the R^2 went way up when we added *EXPN_STU* to the Test Scores regression.
- If we use our reasoning from the univariate case, $R^2 = .44$ means that we are explaining about 44% of the variation in test scores with our regression, up from about 5% with only *STR*.
- So it seems that we are doing a much better job of explaining test scores now. But is this logic okay?
- The answer is yes, but we need to revisit the notion of *goodness of fit* to convince ourselves of why.
- To assess goodness of fit, we ask how well our OLS estimates fit the data?
- There are 3 ways to answer this question
  - The standard error of the regression (*SER*)
  - R^2

Goodness of Fit

- You probably noticed that the R^2 went way up when we added *EXPN_STU* to the Test Scores regression.
- If we use our reasoning from the univariate case, $R^2 = .44$ means that we are explaining about 44% of the variation in test scores with our regression, up from about 5% with only *STR*.
- So it seems that we are doing a much better job of explaining test scores now. But is this logic okay?
- The answer is yes, but we need to revisit the notion of *goodness of fit* to convince ourselves of why.
- To assess goodness of fit, we ask how well our OLS estimates fit the data?
- There are 3 ways to answer this question
    - The standard error of the regression (*SER*)
    - R^2
    - *adjusted*-R^2

The Standard Error of the Regression (SER)

- The standard error of the regression (SER) estimates the standard deviation of the error term u_i.
- Thus, the SER is a measure of the spread of the distribution of Y around the regression line. In the multiple regression model with k regressors,

  $$SER = s_{\widehat{u}}$$

  where

  $$s_{\widehat{u}}^2 = \frac{1}{n-k-1}\sum \widehat{u}_i^2 = \frac{SSR}{n-k-1}$$

  where

  $$SSR = \sum \widehat{u}_i^2 = \sum \left(Y_i - \widehat{Y}_i\right)^2$$

  The divisor $(n-k-1)$ adjusts for the downward bias introduced by estimating the $k+1$ coefficients.
- So SER is one option, but like the variance of Y, it depends on the units of Y, which makes it hard to compare across applications.

R-squared

- As before, the regression R^2 is the fraction of the sample variance of Y_i explained by the regressors.

$$R^2 = \frac{\sum \left(\widehat{Y}_i - \overline{Y}\right)^2}{\sum \left(Y_i - \overline{Y}\right)^2} = \frac{ESS}{TSS} = 1 - \frac{SSR}{TSS}$$

- However, in multiple regression, the R^2 increases whenever a regressor is added.
- So an unscrupulous econometrician might be tempted to keep adding regressors to inflate R^2.

Adjusted R-squared

- One way to adjust for this is to deflate the R^2 by some factor, which is what the *adjusted-R^2* or $\overline{R}^2$ does.

$$\overline{R}^2 = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS} = 1 - \frac{s_{\widehat{u}}^2}{s_Y^2}$$

Notice that:

1. $\overline{R}^2$ is always less than R^2 :
2. Adding a regressor has two effects on $\overline{R}^2$: 1.) SSR falls, which increases $\overline{R}^2$, 2.) $\frac{n-1}{n-k-1}$ increases. So the total effect on $\overline{R}^2$ depends on which effect is bigger.
3. $\overline{R}^2$ can be negative.

Adjusted R-squared

- To see how $\overline{R}^2$ works, let's see what happens when we add irrelevant regressors to the Test Scores regression.

[1] e.g. gen junk1 = invnorm(uniform())

Adjusted R-squared

- To see how $\overline{R}^2$ works, let's see what happens when we add irrelevant regressors to the Test Scores regression.
- Using the *gen* command[1] in Stata, let us created three independent $N(0,1)$ variables (junk1-junk3) and then add them to the regression of Test Scores on STR and EL_PCT.

[1] e.g. gen junk1 = invnorm(uniform())

Adjusted R-squared

- To see how $\overline{R}^2$ works, let's see what happens when we add irrelevant regressors to the Test Scores regression.
- Using the *gen* command[1] in Stata, let us created three independent $N(0,1)$ variables (junk1-junk3) and then add them to the regression of Test Scores on STR and EL_PCT.
- Since these new variables were randomly generated, they shouldn't help explain any of the variation in Test Scores.

[1] e.g. gen junk1 = invnorm(uniform())

Adjusted R-squared

- To see how $\overline{R}^2$ works, let's see what happens when we add irrelevant regressors to the Test Scores regression.
- Using the *gen* command[1] in Stata, let us created three independent $N(0,1)$ variables (junk1-junk3) and then add them to the regression of Test Scores on STR and EL_PCT.
- Since these new variables were randomly generated, they shouldn't help explain any of the variation in Test Scores.
- Let's see what happens in practice.

[1] e.g. gen junk1 = invnorm(uniform())

```
      Source |       SS       df       MS              Number of obs =     420
-------------+------------------------------           F(  2,   417) =  155.01
       Model |  64864.3011     2  32432.1506           Prob > F      =  0.0000
    Residual |  87245.2925   417  209.221325           R-squared     =  0.4264
-------------+------------------------------           Adj R-squared =  0.4237
       Total |  152109.594   419  363.030056           Root MSE      =  14.464

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -1.101296   .3802783    -2.90   0.004    -1.848797   -.3537946
      el_pct |   .6497768   .0393425    16.52   0.000     .7271112    .5724423
       _cons |   686.0322   7.411312    92.57   0.000     671.4641    700.6004
------------------------------------------------------------------------------

      Source |       SS       df       MS              Number of obs =     420
-------------+------------------------------           F(  3,   416) =  103.21
       Model |  64906.6894     3  21635.5631           Prob > F      =  0.0000
    Residual |  87202.9042   416  209.622366           R-squared     =  0.4267
-------------+------------------------------           Adj R-squared =  0.4226
       Total |  152109.594   419  363.030056           Root MSE      =  14.478

     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -1.094981   .3809016    -2.87   0.004    -1.843712   -.3462488
      el_pct |   .6502859   .0393965    16.51   0.000     .7277269    .5728449
       junk1 |   -.335101   .7451975    -0.45   0.653    -1.799923    1.129721
        cons |   685.9061   7.423711    92.39   0.000     671.3135    700.4988
------------------------------------------------------------------------------

      Source |       SS       df       MS              Number of obs =     420
-------------+------------------------------           F(  4,   415) =   77.44
       Model |  65010.6439     4   16252.661           Prob > F      =  0.0000
    Residual |  87098.9497   415  209.876987           R-squared     =  0.4274
-------------+------------------------------           Adj R-squared =  0.4219
       Total |  152109.594   419  363.030056           Root MSE      =  14.487

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |   1.097402   .3811484     2.88   0.004     1.846624    .3481797
      el_pct |   -.651411   .0394528   -16.51   0.000    -.7289633   -.5738587
       junk1 |  -.3239416   .7458185    -0.43   0.664    -1.789995    1.142111
       junk2 |   .4838163   .6874502     0.70   0.482    -.8675022    1.835135
       _cons |   685.9658   7.428702    92.34   0.000     671.3632    700.5684
------------------------------------------------------------------------------

      Source |       SS       df       MS              Number of obs =     420
-------------+------------------------------           F(  5,   414) =   62.34
       Model |  65335.3015     5  13067.0603           Prob > F      =  0.0000
    Residual |  86774.2921   414  209.599739           R-squared     =  0.4295
             |                                         Adj R-squared =  0.4226
       Total |  152109.594   419  363.030056           Root MSE      =  14.478

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -1.105399   .3809508    -2.90   0.004    -1.854238   -.3565603
      el_pct |  -.6511835   .0394272   -16.52   0.000    -.7286859   -.5736811
       junk1 |  -.3242784   .7453258    -0.44   0.664    -1.789373    1.140816
       junk2 |   .5627173   .6899149     0.82   0.415    -.7934557     1.91889
       junk3 |    .856148   .6879098     1.24   0.214    -.4960834     2.20838
       _cons |   686.1817   7.425821    92.40   0.000     671.5847    700.7787
------------------------------------------------------------------------------
```

Some Caveats

Some caveats about using R^2 and $\overline{R}^2$ in practice

Some Caveats

Some caveats about using R^2 and $\overline{R}^2$ in practice

1. An increase in R^2 or $\overline{R}^2$ does not mean that an added variable is statistically significant.

Some Caveats

Some caveats about using R^2 and $\overline{R}^2$ in practice

1. An increase in R^2 or $\overline{R}^2$ does not mean that an added variable is statistically significant.
    - R^2 always increases when we add regressors. $\overline{R}^2$ might not, but even if it does, it doesn't mean that the added variable is statistically significant. You need a hypothesis test to establish this.

Some Caveats

Some caveats about using R^2 and $\overline{R}^2$ in practice

1. An increase in R^2 or $\overline{R}^2$ does not mean that an added variable is statistically significant.
    - R^2 always increases when we add regressors. $\overline{R}^2$ might not, but even if it does, it doesn't mean that the added variable is statistically significant. You need a hypothesis test to establish this.
2. A high R^2 or $\overline{R}^2$ does not mean that the regressors are a true cause of the dependent variable.
    - Remember spurious correlation?

Some Caveats

Some caveats about using R^2 and $\overline{R}^2$ in practice

1. An increase in R^2 or $\overline{R}^2$ does not mean that an added variable is statistically significant.
    - R^2 always increases when we add regressors. $\overline{R}^2$ might not, but even if it does, it doesn't mean that the added variable is statistically significant. You need a hypothesis test to establish this.
2. A high R^2 or $\overline{R}^2$ does not mean that the regressors are a true cause of the dependent variable.
    - Remember spurious correlation?
3. A high R^2 or $\overline{R}^2$ does not mean that there is no omitted variable bias.

Some Caveats

Some caveats about using R^2 and $\overline{R}^2$ in practice

1. An increase in R^2 or $\overline{R}^2$ does not mean that an added variable is statistically significant.
    - R^2 always increases when we add regressors. $\overline{R}^2$ might not, but even if it does, it doesn't mean that the added variable is statistically significant. You need a hypothesis test to establish this.
2. A high R^2 or $\overline{R}^2$ does not mean that the regressors are a true cause of the dependent variable.
    - Remember spurious correlation?
3. A high R^2 or $\overline{R}^2$ does not mean that there is no omitted variable bias.
4. A high R^2 or $\overline{R}^2$ does not necessarily mean that you have the most appropriate set of regressors, nor does a low R^2 or $\overline{R}^2$ necessarily mean that you have a bad set of regressors.

Hypothesis Testing: Testing Joint Hypotheses

- Recall the example above

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

- Suppose we wanted to test the null hypothesis that *both* the coefficient on *STR* and the coefficient on *EXPN_STU* are zero.

Hypothesis Testing: Testing Joint Hypotheses

- Recall the example above

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

- Suppose we wanted to test the null hypothesis that *both* the coefficient on *STR* and the coefficient on *EXPN_STU* are zero.
- This is a **joint** hypothesis since we are imposing two restrictions on the regression model ($\beta_1 = 0$ *and* $\beta_2 = 0$)

Testing Joint Hypotheses

- Can we just use the two t-statistics to construct two t-tests?
- $H_0: \ \beta_1 = 0$ and $\beta_2 = 0$

Testing Joint Hypotheses

- Can we just use the two t-statistics to construct two t-tests?
- $H_0: \ \beta_1 = 0$ and $\beta_2 = 0$
  - Compare $\left| t_{\ \beta_1} \right|$ and $\left| t_{\ \beta_2} \right|$ to the 5% critical value (1.96) and reject if either one is bigger that it?

Testing Joint Hypotheses

- Can we just use the two t-statistics to construct two t-tests?
- $H_0: \ \beta_1 = 0$ and $\beta_2 = 0$
    - Compare $\left| t_{\ \beta_1} \right|$ and $\left| t_{\ \beta_2} \right|$ to the 5% critical value (1.96) and reject if either one is bigger that it?
- Why not? This test won't have the correct size $\Rightarrow$ you will be wrong more than 5% of the time.

Testing Joint Hypotheses

- Can we just use the two t-statistics to construct two t-tests?
- $H_0: \ \beta_1 = 0$ and $\beta_2 = 0$
    - Compare $\left| t_{\ \beta_1} \right|$ and $\left| t_{\ \beta_2} \right|$ to the 5% critical value (1.96) and reject if either one is bigger that it?
- Why not? This test won't have the correct size $\Rightarrow$ you will be wrong more than 5% of the time.
- Let's see why.

Testing Joint Hypotheses

$$H_0: \ \beta_1 = 0 \text{ and } \beta_2 = 0$$

- First of all, what is the alternative hypothesis here?

Testing Joint Hypotheses

$$H_0: \ \beta_1 = 0 \text{ and } \beta_2 = 0$$

- First of all, what is the alternative hypothesis here?
    - $H_A : \beta_1 \neq 0$ and $\beta_2 \neq 0$?

Testing Joint Hypotheses

$$H_0: \ \beta_1 = 0 \text{ and } \beta_2 = 0$$

- First of all, what is the alternative hypothesis here?
  - $H_A : \beta_1 \neq 0$ and $\beta_2 \neq 0$?
  - $H_A : \beta_1 \neq 0$ and/or $\beta_2 \neq 0$?

Testing Joint Hypotheses

$$H_0: \ \beta_1 = 0 \text{ and } \beta_2 = 0$$

- First of all, what is the alternative hypothesis here?
    - $H_A : \beta_1 \neq 0$ and $\beta_2 \neq 0$?
    - $H_A : \beta_1 \neq 0$ and/or $\beta_2 \neq 0$?
- Suppose we use the two t-statistics and reject the null hypothesis if either one exceeds 1.96 in absolute value.

Testing Joint Hypotheses

$$H_0: \ \beta_1 = 0 \text{ and } \beta_2 = 0$$

- First of all, what is the alternative hypothesis here?
    - $H_A : \beta_1 \neq 0$ and $\beta_2 \neq 0$?
    - $H_A : \beta_1 \neq 0$ and/or $\beta_2 \neq 0$?
- Suppose we use the two t-statistics and reject the null hypothesis if either one exceeds 1.96 in absolute value.
- Let's consider the simple case where the t-statistics are uncorrelated and therefore independent (because they are asymptotically normal).

Testing Joint Hypotheses

$$H_0: \ \beta_1 = 0 \text{ and } \beta_2 = 0$$

- First of all, what is the alternative hypothesis here?
  - $H_A: \beta_1 \neq 0$ and $\beta_2 \neq 0$?
  - $H_A: \beta_1 \neq 0$ and/or $\beta_2 \neq 0$?
- Suppose we use the two t-statistics and reject the null hypothesis if either one exceeds 1.96 in absolute value.
- Let's consider the simple case where the t-statistics are uncorrelated and therefore independent (because they are asymptotically normal).
- What, then, is the probability that you will reject the null hypothesis when the **null is in fact true**?

Testing Joint Hypotheses

- Remember that for a test with a 5% significance level this should equal .05

Testing Joint Hypotheses

- Remember that for a test with a 5% significance level this should equal .05
- This is called the *size* of the test (we have been denoting it α).

Testing Joint Hypotheses

- Remember that for a test with a 5% significance level this should equal .05
- This is called the *size* of the test (we have been denoting it α).
- The probability that we correctly reject the null when the **alternative is true** is called the *power* of the test (usually denoted β).

Testing Joint Hypotheses

- Remember that for a test with a 5% significance level this should equal .05
- This is called the *size* of the test (we have been denoting it α).
- The probability that we correctly reject the null when the **alternative is true** is called the *power* of the test (usually denoted β).
- Intuitively, you would like α to be small and β to be big, but there is a trade-off between them.

Testing Joint Hypotheses

- Remember that for a test with a 5% significance level this should equal .05
- This is called the *size* of the test (we have been denoting it α).
- The probability that we correctly reject the null when the **alternative is true** is called the *power* of the test (usually denoted β).
- Intuitively, you would like α to be small and β to be big, but there is a trade-off between them.
- So what is the size (α) of the *overall* test when we use the two t-ratios?
- We know that the null is *not* rejected only if *both*

$$\left| t_{\widehat{\beta}_1} \right| \leq 1.96 \text{ and } \left| t_{\widehat{\beta}_2} \right| \leq 1.96$$

Testing Joint Hypotheses

- Since the t-ratios are independent here (by assumption)

$$
\begin{aligned}
P\left(\left|t_{\widehat{\beta}_1}\right| < 1.96, \left|t_{\widehat{\beta}_2}\right| < 1.96\right) &= P\left(\left|t_{\widehat{\beta}_1}\right| < 1.96\right) P\left(\left|t_{\widehat{\beta}_2}\right| < 1.96\right) \\
&= .95^2 = .9025
\end{aligned}
$$

- So the probability of rejecting the null when the null is true is: 9.75%

Testing Joint Hypotheses

- Since the t-ratios are independent here (by assumption)

$$\begin{aligned} P\left(\left|t_{\widehat{\beta}_1}\right| < 1.96, \left|t_{\widehat{\beta}_2}\right| < 1.96\right) &= P\left(\left|t_{\widehat{\beta}_1}\right| < 1.96\right) P\left(\left|t_{\widehat{\beta}_2}\right| < 1.96\right) \\ &= .95^2 = .9025 \end{aligned}$$

- So the probability of rejecting the null when the null is true is: 9.75%
- The joint test will then have a 9.75% significance level, not 5%

Testing Joint Hypotheses

- Since the t-ratios are independent here (by assumption)

$$\begin{aligned} P\left(\left|t_{\widehat{\beta}_1}\right| < 1.96, \left|t_{\widehat{\beta}_2}\right| < 1.96\right) &= P\left(\left|t_{\widehat{\beta}_1}\right| < 1.96\right) P\left(\left|t_{\widehat{\beta}_2}\right| < 1.96\right) \\ &= .95^2 = .9025 \end{aligned}$$

- So the probability of rejecting the null when the null is true is: 9.75%
- The joint test will then have a 9.75% significance level, not 5%
- You will be wrong almost twice as often as you thought!

Testing Joint Hypotheses

- If the t-statistics are not independent, the situation is worse since we need to account for the correlation between them.

Testing Joint Hypotheses

- If the t-statistics are not independent, the situation is worse since we need to account for the correlation between them.
- What now?

Testing Joint Hypotheses

- If the t-statistics are not independent, the situation is worse since we need to account for the correlation between them.
- What now?
- One option would be to modify this "one at a time" method so that its size is equal to its significance level (i.e. pick a number bigger than 1.96).

Testing Joint Hypotheses

- If the t-statistics are not independent, the situation is worse since we need to account for the correlation between them.
- What now?
- One option would be to modify this "one at a time" method so that its size is equal to its significance level (i.e. pick a number bigger than 1.96).
- This method, called the Bonferroni test, is rarely used in practice (because it has low power) so we will not cover it in this class.

Testing Joint Hypotheses

- If the t-statistics are not independent, the situation is worse since we need to account for the correlation between them.
- What now?
- One option would be to modify this "one at a time" method so that its size is equal to its significance level (i.e. pick a number bigger than 1.96).
- This method, called the Bonferroni test, is rarely used in practice (because it has low power) so we will not cover it in this class.
- Fortunately, there are two other very good options.

Transforming the Regression

(If we can) we might try to transform the regression in such a way that we can just use a single t-test

Example

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \quad (1)$$

Suppose we test

$$H_0 : \beta_1 = \beta_2 \text{ vs. } H_A : \beta_1 \neq \beta_2$$

Transforming the Regression

(If we can) we might try to transform the regression in such a way that we can just use a single t-test

Example

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \qquad (1)$$

Suppose we test

$$H_0 : \beta_1 = \beta_2 \text{ vs. } H_A : \beta_1 \neq \beta_2$$

How? From (1), add and subtract $\beta_2 X_{1i}$ and get

$$Y_i = \beta_0 + (\beta_1 - \beta_2) X_{1i} + \beta_2 (X_{1i} + X_{2i}) + u_i$$

Transforming the Regression

(If we can) we might try to transform the regression in such a way that we can just use a single t-test

Example

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \quad (1)$$

Suppose we test

$$H_0 : \beta_1 = \beta_2 \text{ vs. } H_A : \beta_1 \neq \beta_2$$

How? From (1), add and subtract $\beta_2 X_{1i}$ and get

$$Y_i = \beta_0 + (\beta_1 - \beta_2) X_{1i} + \beta_2 (X_{1i} + X_{2i}) + u_i$$

Letting $\gamma \equiv \beta_1 - \beta_2$, we can write this as

$$Y_i = \beta_0 + \gamma X_{1i} + \beta_2 (X_{1i} + X_{2i}) + u_i \quad (2)$$

Now estimate (2) with OLS and test $\gamma = 0$

Transforming the Regression

Another example. Again let

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \quad (1)$$

But now suppose we test

$$H_0 : \beta_1 + 2\beta_2 = 3 \text{ vs. } H_A : \beta_1 + 2\beta_2 \neq 3$$

From (1), add and subtract $2\beta_2 X_{1i}$ and get

$$Y_i = \beta_0 + (\beta_1 + 2\beta_2) X_{1i} + \beta_2 (X_{2i} - 2X_{1i}) + u_i$$

Letting $\delta \equiv \beta_1 + 2\beta_2$

$$Y_i = \beta_0 + \delta X_{1i} + \beta_2 (X_{2i} - 2X_{1i}) + u_i \quad (3)$$

Now estimate (3) with OLS and test $\delta = 3$

Transforming the Regression

- This type of "clever" test can be done for any null hypothesis that can be written as a unique linear restriction.

Transforming the Regression

- This type of "clever" test can be done for any null hypothesis that can be written as a unique linear restriction.
- But what if

$$H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0$$

or

$$H_0 : \beta_1 = \beta_2 \text{ and } \beta_3 = 2\beta_1$$

Transforming the Regression

- This type of “clever” test can be done for any null hypothesis that can be written as a unique linear restriction.
- But what if

$$H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0$$

  or

$$H_0 : \beta_1 = \beta_2 \text{ and } \beta_3 = 2\beta_1$$

- For these joint restrictions, we need another method: the F-test.

Transforming the Regression

- This type of “clever” test can be done for any null hypothesis that can be written as a unique linear restriction.
- But what if

$$H_0 : \beta_1 = 0 \text{ and } \beta_2 = 0$$

  or

$$H_0 : \beta_1 = \beta_2 \text{ and } \beta_3 = 2\beta_1$$

- For these joint restrictions, we need another method: the F-test.
- Since this new method will work on both the previous (simpler) problems as well as these more complex ones, we rarely uses the “clever” transformation type tests in practice.

F-statistic

- The F-statistic exploits the fact that the t-statistics of individual coefficients are normally distributed.

F-statistic

- The F-statistic exploits the fact that the t-statistics of individual coefficients are normally distributed.
- Suppose we want to test $H_0 : \beta_1 = 0$ and $\beta_2 = 0$

F-statistic

- The F-statistic exploits the fact that the t-statistics of individual coefficients are normally distributed.
- Suppose we want to test $H_0 : \beta_1 = 0$ and $\beta_2 = 0$
- The F-statistic combines the two t-statistics $t_{\widehat{\beta}_1}$ and $t_{\widehat{\beta}_2}$ using the formula

$$F = \frac{1}{2}\left(\frac{t_{\widehat{\beta}_1}^2 + t_{\widehat{\beta}_2}^2 - 2\widehat{\rho}_{t_{\widehat{\beta}_1} t_{\widehat{\beta}_2}} t_{\widehat{\beta}_1} t_{\widehat{\beta}_2}}{1 - \widehat{\rho}^2_{t_{\widehat{\beta}_1} t_{\widehat{\beta}_2}}}\right)$$

where $\widehat{\rho}_{t_{\widehat{\beta}_1} t_{\widehat{\beta}_2}}$ is an estimator of the correlation between the two t-statistics.

F-statistic

- The F-statistic exploits the fact that the t-statistics of individual coefficients are normally distributed.
- Suppose we want to test $H_0 : \beta_1 = 0$ and $\beta_2 = 0$
- The F-statistic combines the two t-statistics $t_{\widehat{\beta}_1}$ and $t_{\widehat{\beta}_2}$ using the formula

$$F = \frac{1}{2}\left(\frac{t^2_{\widehat{\beta}_1} + t^2_{\widehat{\beta}_2} - 2\widehat{\rho}_{t_{\widehat{\beta}_1} t_{\widehat{\beta}_2}} t_{\widehat{\beta}_1} t_{\widehat{\beta}_2}}{1 - \widehat{\rho}^2_{t_{\widehat{\beta}_1} t_{\widehat{\beta}_2}}}\right)$$

  where $\widehat{\rho}_{t_{\widehat{\beta}_1} t_{\widehat{\beta}_2}}$ is an estimator of the correlation between the two t-statistics.
- But how is this new test statistic distributed?

F-statistic

- Consider the case where the t-statistics are uncorrelated, so that

$$F = \frac{1}{2}\left(t_{\widehat{\beta}_1}^2 + t_{\widehat{\beta}_2}^2\right)$$

F-statistic

- Consider the case where the t-statistics are uncorrelated, so that

$$F = \frac{1}{2}\left(t^2_{\widehat{\beta}_1} + t^2_{\widehat{\beta}_2}\right)$$

- Now, under the null hypothesis, $t_{\widehat{\beta}_1}$ and $t_{\widehat{\beta}_2}$ are independent standard normal random variables so that, under the null, F has an $F_{2,\infty}$ distribution.

F-statistic

- Consider the case where the t-statistics are uncorrelated, so that

$$F = \frac{1}{2}\left(t_{\widehat{\beta}_1}^2 + t_{\widehat{\beta}_2}^2\right)$$

- Now, under the null hypothesis, $t_{\widehat{\beta}_1}$ and $t_{\widehat{\beta}_2}$ are independent standard normal random variables so that, under the null, F has an $F_{2,\infty}$ distribution.
- Recall (from earlier slides) that:
  - χ_m^2 is the sum of m squared independent standard normals.
  - $F_{n,m} = \frac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}$ (where the χ^2's are independent).
  - $F_{m,\infty} = \frac{\chi_m^2}{m}$(the average of m squared normals).
- So (in this example) $F = \frac{1}{2}\left(t_{\widehat{\beta}_1}^2 + t_{\widehat{\beta}_2}^2\right) \sim F_{2,\infty}$

F-statistic

- A large value of F will then lead you to reject the null (since a large value of F means that $t^2_{\widehat{\beta}_1}$ or $t^2_{\widehat{\beta}_2}$ (or both) are large).

F-statistic

- A large value of F will then lead you to reject the null (since a large value of F means that $t^2_{\widehat{\beta}_1}$ or $t^2_{\widehat{\beta}_2}$ (or both) are large).
- In general (for this H_0), the above formula adjusts for any correlation between the t-statistics so that under the null, the F-statistic will have a $F_{2,\infty}$ distribution in large samples (regardless of the correlation in the t-statistics).

F-statistic

- A large value of F will then lead you to reject the null (since a large value of F means that $t^2_{\widehat{\beta}_1}$ or $t^2_{\widehat{\beta}_2}$ (or both) are large).
- In general (for this H_0), the above formula adjusts for any correlation between the t-statistics so that under the null, the F-statistic will have a $F_{2,\infty}$ distribution in large samples (regardless of the correlation in the t-statistics).
- The general formula for the F-statistic with q restrictions is quite complicated, so we will not present it here.

F-statistic

- A large value of F will then lead you to reject the null (since a large value of F means that $t^2_{\widehat{\beta}_1}$ or $t^2_{\widehat{\beta}_2}$ (or both) are large).
- In general (for this H_0), the above formula adjusts for any correlation between the t-statistics so that under the null, the F-statistic will have a $F_{2,\infty}$ distribution in large samples (regardless of the correlation in the t-statistics).
- The general formula for the F-statistic with q restrictions is quite complicated, so we will not present it here.
- However, under the null hypothesis, the F-statistic is distributed $F_{q,\infty}$ (in large samples).

F-statistic

- Unlike the transformation approach, the F-statistic can be used to test both complicated hypotheses like

$$H_0 : \beta_1 = \beta_2 \textit{ and } \beta_3 = 2\beta_1$$

  which has 2 restrictions, as well as simple hypotheses like

$$H_0 : \beta_1 = \beta_2 \text{ vs. } H_0 : \beta_1 \neq \beta_2$$

  which has only 1 restriction.
- The F-statistic is also automatically computed by statistical packages (like Stata) with simple commands.

F-statistic

- The F-statistic can be used to test any linear restriction or set of linear restrictions.

F-statistic

- The F-statistic can be used to test any linear restriction or set of linear restrictions.
- Unfortunately, non-linear restrictions (e.g. $H_0 : \beta_1 = \beta_2^2$) require a different methodology that we won't cover here.

F-statistic

- The F-statistic can be used to test any linear restriction or set of linear restrictions.
- Unfortunately, non-linear restrictions (e.g. $H_0 : \beta_1 = \beta_2^2$) require a different methodology that we won't cover here.
- Finally, because the F-statistic is constructed to detect any violation of H_0, it is also valid when the alternative is something like $H_A : \beta_1 > 0$ and/or $\beta_2 > 0$, but it will not be the best test under such (one-sided) alternatives.

F-statistic

- The F-statistic can be used to test any linear restriction or set of linear restrictions.
- Unfortunately, non-linear restrictions (e.g. $H_0 : \beta_1 = \beta_2^2$) require a different methodology that we won't cover here.
- Finally, because the F-statistic is constructed to detect any violation of H_0, it is also valid when the alternative is something like $H_A : \beta_1 > 0$ and/or $\beta_2 > 0$, but it will not be the best test under such (one-sided) alternatives.
  - In such cases, there are alternatives with more power, but they are beyond the scope of this class.

Using the F-statistic in Practice

1. Count the number of "restrictions" under the null (degrees of freedom), call this q

[2] The p-value is calculated as p-value $= \Pr(F_{q,\infty} > F^{act}) = \Pr(\chi^2_q > q \cdot F^{act})$

Using the F-statistic in Practice

1. Count the number of "restrictions" under the null (degrees of freedom), call this q
2. Compute the F-statistic

[2] The p-value is calculated as p-value $= \Pr(F_{q,\infty} > F^{act}) = \Pr(\chi^2_q > q \cdot F^{act})$

Using the F-statistic in Practice

1. Count the number of "restrictions" under the null (degrees of freedom), call this q
2. Compute the F-statistic
3. Check the table for $F_{q,\infty}$ (or get the *p-value* from Stata[2])

[2] The p-value is calculated as $p\text{-value} = \Pr(F_{q,\infty} > F^{act}) = \Pr(\chi^2_q > q \cdot F^{act})$

Using the F-statistic in Practice

1. Count the number of "restrictions" under the null (degrees of freedom), call this q
2. Compute the F-statistic
3. Check the table for $F_{q,\infty}$ (or get the *p-value* from Stata[2])

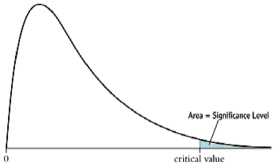
Note: Since

$$F_{q,\infty} = \frac{\chi^2_q}{q} \implies \chi^2_q = qF_{q,\infty}$$

you can also use the tables for the χ^2_q distribution (if you calculate $qF_{q,\infty}$ or are given χ^2_q)

[2] The p-value is calculated as p-value $= \Pr(F_{q,\infty} > F^{act}) = \Pr(\chi^2_q > q \cdot F^{act})$

TABLE 4 Critical Values for the $F_{m,\infty}$ Distribution



Degrees of Freedom	Significance Level 10%	5%	1%
1	2.71	3.84	6.63
2	2.30	3.00	4.61
3	2.08	2.60	3.78
4	1.94	2.37	3.32
5	1.85	2.21	3.02
6	1.77	2.10	2.80
7	1.72	2.01	2.64
8	1.67	1.94	2.51
9	1.63	1.88	2.41
10	1.60	1.83	2.32
11	1.57	1.79	2.25
12	1.55	1.75	2.18
13	1.52	1.72	2.13
14	1.50	1.69	2.08
15	1.49	1.67	2.04
16	1.47	1.64	2.00
17	1.46	1.62	1.97
18	1.44	1.60	1.93
19	1.43	1.59	1.90
20	1.42	1.57	1.88
21	1.41	1.56	1.85
22	1.40	1.54	1.83
23	1.39	1.53	1.81
24	1.38	1.52	1.79
25	1.38	1.51	1.77
26	1.37	1.50	1.76
27	1.36	1.49	1.74
28	1.35	1.48	1.72
29	1.35	1.47	1.71
30	1.34	1.46	1.70

This table contains the 90th, 95th, and 99th percentiles of the $F_{m,\infty}$ distribution. These serve as critical values for tests with significance levels of 10%, 5%, and 1%.

TABLE 3 Critical Values for the χ^2 Distribution

Degrees of Freedom	Significance Level		
	10%	5%	1%
1	2.71	3.84	6.63
2	4.61	5.99	9.21
3	6.25	7.81	11.34
4	7.78	9.49	13.28
5	9.24	11.07	15.09
6	10.64	12.59	16.81
7	12.02	14.07	18.48
8	13.36	15.51	20.09
9	14.68	16.92	21.67
10	15.99	18.31	23.21
11	17.28	19.68	24.72
12	18.55	21.03	26.22
13	19.81	22.36	27.69
14	21.06	23.68	29.14
15	22.31	25.00	30.58
16	23.54	26.30	32.00
17	24.77	27.59	33.41
18	25.99	28.87	34.81
19	27.20	30.14	36.19
20	28.41	31.41	37.57
21	29.62	32.67	38.93
22	30.81	33.92	40.29
23	32.01	35.17	41.64
24	33.20	36.41	42.98
25	34.38	37.65	44.31
26	35.56	38.89	45.64
27	36.74	40.11	46.96
28	37.92	41.34	48.28
29	39.09	42.56	49.59
30	40.26	43.77	50.89

This table contains the 90th, 95th, and 99th percentiles of the χ^2 distribution. These serve as critical values for tests with significance levels of 10%, 5%, and 1%.

Summary: Testing Multiple Restrictions

- Assuming OLS Assumptions 1-4 hold (but not necessarily A5 - homoskedasticity) and a large sample size:

Summary: Testing Multiple Restrictions

- Assuming OLS Assumptions 1-4 hold (but not necessarily A5 - homoskedasticity) and a large sample size:
    - Use the F-statistic (if you have only one restriction you can use a t-statistic).

Summary: Testing Multiple Restrictions

- Assuming OLS Assumptions 1-4 hold (but not necessarily A5 - homoskedasticity) and a large sample size:
  - Use the F-statistic (if you have only one restriction you can use a t-statistic).
  - When you don't have access to a computer, I will give you the value of the F-statistic (or there will be a clever way to figure it out).

Summary: Testing Multiple Restrictions

- Assuming OLS Assumptions 1-4 hold (but not necessarily A5 - homoskedasticity) and a large sample size:
    - Use the F-statistic (if you have only one restriction you can use a t-statistic).
    - When you don't have access to a computer, I will give you the value of the F-statistic (or there will be a clever way to figure it out).
    - If you do have access to a computer, use the proper command to compute the F-statistic and either:

Summary: Testing Multiple Restrictions

- Assuming OLS Assumptions 1-4 hold (but not necessarily A5 - homoskedasticity) and a large sample size:
  - Use the F-statistic (if you have only one restriction you can use a t-statistic).
  - When you don't have access to a computer, I will give you the value of the F-statistic (or there will be a clever way to figure it out).
  - If you do have access to a computer, use the proper command to compute the F-statistic and either:
    - Use the p-value provided and reject the null if the p-value is less than the chosen significance level α.

Summary: Testing Multiple Restrictions

- Assuming OLS Assumptions 1-4 hold (but not necessarily A5 - homoskedasticity) and a large sample size:
    - Use the F-statistic (if you have only one restriction you can use a t-statistic).
    - When you don't have access to a computer, I will give you the value of the F-statistic (or there will be a clever way to figure it out).
    - If you do have access to a computer, use the proper command to compute the F-statistic and either:
        - Use the p-value provided and reject the null if the p-value is less than the chosen significance level α.
        - Use the tables in Wooldridge, choosing the appropriate cell.
- Now, let's look at some examples using the Test Score data...

Examples

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

Examples

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

- Let's start by testing $H_0 : \beta_1 = \beta_2$ and $\beta_3 = 2\beta_1$
- To test this in Stata, you can use the test command

Examples

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

- Let's start by testing $H_0 : \beta_1 = \beta_2$ and $\beta_3 = 2\beta_1$
- To test this in Stata, you can use the test command

```
. test (str=expn_stu) (el_pct=2*str)

 ( 1)  str - expn_stu = 0
 ( 2) - 2 str + el_pct = 0

       F(  2,   416) =  219.65
            Prob > F =    0.0000
```

Examples

- Note that Stata calculates the p-value using the $F_{2,416}$ distribution instead of the $F_{2,\infty}$ distribution.

Examples

- Note that Stata calculates the p-value using the $F_{2,416}$ distribution instead of the $F_{2,\infty}$ distribution.
- Just as with the p-values of the individual coefficients, the results will be essentially the same if the sample is reasonably large.

Examples

- Note that Stata calculates the p-value using the $F_{2,416}$ distribution instead of the $F_{2,\infty}$ distribution.
- Just as with the p-values of the individual coefficients, the results will be essentially the same if the sample is reasonably large.
- However, because $F_{q,\infty} = \frac{\chi^2_q}{q} \Longrightarrow \chi^2_q = qF_{q,\infty}$ you can also calculate the χ^2 statistic and the associated asymptotic p-value using the *dis* (display) command

Examples

- Note that Stata calculates the p-value using the $F_{2,416}$ distribution instead of the $F_{2,\infty}$ distribution.
- Just as with the p-values of the individual coefficients, the results will be essentially the same if the sample is reasonably large.
- However, because $F_{q,\infty} = \frac{\chi^2_q}{q} \Longrightarrow \chi^2_q = qF_{q,\infty}$ you can also calculate the χ^2 statistic and the associated asymptotic p-value using the *dis* (display) command

```
. dis 2*r(F)
439.30201

. dis (1-chi2(2,2*r(F)))
0
```

- Since the p-value $\simeq 0$, we can reject the null at any level of significance.

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001    .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001    .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607   .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008  -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917   679.9641
------------------------------------------------------------------------------
```

- Let's test a more interesting null hypothesis: $H_0 : \beta_1 = 0$ and $\beta_2 = 0$

```
. test  str expn_stu

 ( 1)  str = 0.0
 ( 2)  expn_stu = 0.0

       F(  2,   416) =    5.43
            Prob > F =    0.0047
```

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

- Let's test a more interesting null hypothesis: $H_0 : \beta_1 = 0$ and $\beta_2 = 0$

```
. test  str expn_stu

 ( 1)  str = 0.0
 ( 2)  expn_stu = 0.0

       F(  2,   416) =    5.43
            Prob > F =    0.0047
```

- Again, we can also calculate the χ^2 statistic and the associated *p*-value using the *dis* (display) command

```
. dis 2*r(F)
10.867454

. dis (1-chi2(2,2*r(F)))
.00436679
```

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                Number of obs =     420
                                                      F(  3,   416) =  147.20
                                                      Prob > F      =  0.0000
                                                      R-squared     =  0.4366
                                                      Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

- Let's consider yet another null: $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ and $\beta_3 = 0$

```
. test  str expn_stu  el_pct

 ( 1)  str = 0.0
 ( 2)  expn_stu = 0.0
 ( 3)  el_pct = 0.0

       F(  3,   416) =  147.20
            Prob > F =    0.0000


. dis 3*r(F)
441.61113

. dis (1-chi2(3,3*r(F)))
0
```

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

- Let's consider yet another null: $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ and $\beta_3 = 0$

```
. test  str expn_stu  el_pct

 ( 1)  str = 0.0
 ( 2)  expn_stu = 0.0
 ( 3)  el_pct = 0.0

       F(  3,   416) =  147.20
            Prob > F =    0.0000


. dis 3*r(F)
441.61113

. dis (1-chi2(3,3*r(F)))
0
```

- How about a single restriction? Let's test $H_0 : \beta_1 = 0$

```
. test  str

 ( 1)  str = 0.0

       F(  1,   416) =    0.35
            Prob > F =    0.5528


. dis r(F)
.35295427

. dis (1-chi2(1,r(F)))
.55244553
```

```
. reg  testscr str expn_stu el_pct, robust

Regression with robust standard errors                 Number of obs =     420
                                                       F(  3,   416) =  147.20
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.4366
                                                       Root MSE      =  14.353

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4820728    -0.59   0.553    -1.234001     .661203
    expn_stu |   .0038679   .0015807     2.45   0.015     .0007607    .0069751
      el_pct |  -.6560227   .0317844   -20.64   0.000    -.7185008   -.5935446
       _cons |   649.5779   15.45834    42.02   0.000     619.1917    679.9641
------------------------------------------------------------------------------
```

- Similarly, for $H_0 : \beta_2 = 0$

```
. test  expn_stu

 ( 1)  expn_stu = 0.0

       F(  1,   416) =    5.99
            Prob > F =    0.0148


. dis r(F)
5.9874191

. dis (1-chi2(1,r(F)))
.01440827
```

F-statistic under homoskedasticity

- Just like we constructed H_0 SEs and H_0 t-statistics, we can construct a homoskedasticity-only F-statistic.
- Assume A1-A4 and A5 (homoskedasticity)
- Let the model be given by

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + u_i \quad (1)$$

- This is called the "unrestricted" model since we are estimating the coefficients without imposing any restrictions on the coefficients.
- We can easily calculate SSR_U $\left(= \sum \widehat{u}_i^2\right)$ and $R_U^2 = 1 - \frac{SSR_U}{TSS}$

F-statistic under homoskedasticity

- Suppose we have q linear restrictions
  - For example
$$\beta_1 = 0, \beta_2 = 0, \beta_3 = \beta_4 \implies q = 3$$
  - The model then becomes (in this example)
$$Y_i = \beta_0 + \beta_3 \left(X_{3i} + X_{4i}\right) + ... + \beta_k X_{ki} + u_i \quad (2)$$
    which is the *restricted* model (for this example)
- Since we have incorporated the restriction, we can estimate (2) and calculate SSR_R and $R_R^2 = 1 - \frac{SSR_R}{TSS}$
- We know of course that $SSR_R \geq SSR_U$ and $R_U^2 \geq R_R^2$
- Rejecting the null requires these differences to be "large"

F-statistic under homoskedasticity

- With homoskedasticity, the F-statistic can be written as

$$F = \frac{\left(SSR_R - SSR_U\right)/q}{SSR_U/\left(n-k-1\right)} = \frac{\left(R_U^2 - R_R^2\right)/q}{\left(1-R_U^2\right)/\left(n-k-1\right)}$$

F-statistic under homoskedasticity

- With homoskedasticity, the F-statistic can be written as

$$F = \frac{\left(SSR_R - SSR_U\right)/q}{SSR_U/\left(n-k-1\right)} = \frac{\left(R_U^2 - R_R^2\right)/q}{\left(1-R_U^2\right)/\left(n-k-1\right)}$$

- If $u_i \sim N\left(0, \sigma^2\right)$ it can be shown that

$$\left(SSR_R - SSR_U\right) \sim \chi_q^2 \quad \& \quad SSR_U \sim \chi_{n-k-1}^2$$

F-statistic under homoskedasticity

- With homoskedasticity, the F-statistic can be written as

$$F = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n-k-1)} = \frac{\left(R_U^2 - R_R^2\right)/q}{\left(1 - R_U^2\right)/(n-k-1)}$$

- If $u_i \sim N\left(0, \sigma^2\right)$ it can be shown that

$$(SSR_R - SSR_U) \sim \chi_q^2 \quad \& \quad SSR_U \sim \chi_{n-k-1}^2$$

- So our test is

$$\frac{\chi_q^2/q}{\chi_{n-k-1}^2/(n-k-1))} = F_{q,n-k-1}$$

F-statistic under homoskedasticity

- With homoskedasticity, the F-statistic can be written as

$$F = \frac{(SSR_R - SSR_U)/q}{SSR_U/(n-k-1)} = \frac{\left(R_U^2 - R_R^2\right)/q}{\left(1-R_U^2\right)/(n-k-1)}$$

- If $u_i \sim N\left(0, \sigma^2\right)$ it can be shown that

$$(SSR_R - SSR_U) \sim \chi_q^2 \quad \& \quad SSR_U \sim \chi_{n-k-1}^2$$

- So our test is

$$\frac{\chi_q^2/q}{\chi_{n-k-1}^2/(n-k-1))} = F_{q,n-k-1}$$

- Let's call this the *Rule of Thumb* ("RT") F-statistic.

F-statistic under homoskedasticity

- Using our Test Scores regression

```
. reg testscr str  expn_stu el_pct

      Source |       SS       df       MS              Number of obs =     420
-------------+------------------------------           F(  3,   416) =  107.45
       Model |  66409.8837     3  22136.6279           Prob > F      =  0.0000
    Residual |  85699.7099   416  206.008918           R-squared     =  0.4366
-------------+------------------------------           Adj R-squared =  0.4325
       Total |  152109.594   419  363.030056           Root MSE      =  14.353

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.2863992   .4805232    -0.60   0.551    -1.230955     .658157
    expn_stu |   .0038679   .0014121     2.74   0.006     .0010921    .0066437
      el_pct |  -.6560227   .0391059   -16.78   0.000    -.7328924   -.5791529
       _cons |   649.5779   15.20572    42.72   0.000     619.6883    679.4676
------------------------------------------------------------------------------
```

Now let's construct the "RT" F-statistic for $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

$$F = \frac{\left(R_U^2 - R_R^2\right) / q}{\left(1 - R_U^2\right) / \left(n - k - 1\right)} = \frac{(.4366 - 0)/3}{(1 - .4366)/416} = 107.45$$

- Note that this matches what Stata reports automatically in the upper right corner of the output.

F-statistic under homoskedasticity

- To construct the "RT" F-statistic for $H_0 : \beta_1 = \beta_2 = 0$ we need to run the "restricted" regression explicitly

```
. reg testscr el_pct

      Source |       SS       df       MS              Number of obs =     420
-------------+------------------------------           F(  1,   418) =  296.40
       Model |  63109.5688     1  63109.5688           Prob > F      =  0.0000
    Residual |  89000.0248   418  212.91872            R-squared     =  0.4149
-------------+------------------------------           Adj R-squared =  0.4135
       Total |  152109.594   419  363.030056           Root MSE      =  14.592

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      el_pct |  -.6711562   .0389837   -17.22   0.000    -.7477847   -.5945277
       _cons |   664.7394   .9406415   706.69   0.000     662.8905    666.5884
------------------------------------------------------------------------------
```

F-statistic under homoskedasticity

- To construct the "RT" F-statistic for $H_0 : \beta_1 = \beta_2 = 0$ we need to run the "restricted" regression explicitly

```
. reg testscr el_pct

      Source |       SS       df       MS              Number of obs =     420
-------------+------------------------------           F(  1,   418) =  296.40
       Model |  63109.5688     1  63109.5688           Prob > F      =  0.0000
    Residual |  89000.0248   418   212.91872           R-squared     =  0.4149
-------------+------------------------------           Adj R-squared =  0.4135
       Total |  152109.594   419  363.030056           Root MSE      =  14.592

------------------------------------------------------------------------------
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      el_pct |  -.6711562   .0389837   -17.22   0.000    -.7477847   -.5945277
       _cons |   664.7394   .9406415   706.69   0.000     662.8905    666.5884
------------------------------------------------------------------------------
```

- We can then construct the "RT" F-statistic as

$$F = \frac{\left(R_U^2 - R_R^2\right)/q}{\left(1 - R_U^2\right)/\left(n-k-1\right)} = \frac{(.4366 - .4149)/2}{(1 - .4366)/416} = 8.011$$

F-statistic under homoskedasticity

- Note that this matches what we would find if we ran the test directly in Stata following the original *unrestricted* regression (i.e. after typing: reg testscr str expn_stu el_pct).

F-statistic under homoskedasticity

- Note that this matches what we would find if we ran the test directly in Stata following the original *unrestricted* regression (i.e. after typing: reg testscr str expn_stu el_pct).

```
. test str expn_stu

 ( 1)  str = 0
 ( 2)  expn_stu = 0

       F(  2,   416) =    8.01
            Prob > F =    0.0004
```

F-statistic under homoskedasticity

- Note that this matches what we would find if we ran the test directly in Stata following the original *unrestricted* regression (i.e. after typing: reg testscr str expn_stu el_pct).

```
. test str expn_stu

 ( 1)  str = 0
 ( 2)  expn_stu = 0

       F(  2,   416) =    8.01
            Prob > F =    0.0004
```

- Although it's easy to calculate and intuitively appealing, the Rule of Thumb F-statistic is **not** valid if u_i is not normal or if the errors are not homoskedastic (both strong assumptions).

F-statistic under homoskedasticity

- Note that this matches what we would find if we ran the test directly in Stata following the original *unrestricted* regression (i.e. after typing: reg testscr str expn_stu el_pct).

```
. test str expn_stu

 ( 1)  str = 0
 ( 2)  expn_stu = 0

       F(  2,   416) =    8.01
            Prob > F =    0.0004
```

- Although it's easy to calculate and intuitively appealing, the Rule of Thumb F-statistic is **not** valid if u_i is not normal or if the errors are not homoskedastic (both strong assumptions).
- Therefore, we will not use this particular version of the F-statistic very often.