Lecture 9: Hypothesis Tests and Confidence Intervals in Multiple Regression

Zheng Tian

1 Introduction

1.1 Overview

This lecture presents the methods for testing the hypotheses concerning the coefficients in a multiple regression model. Besides the t-statistic that we have learned in Lecture 6, we introduce a new test statistic, the F-statistic, which is used to test the joint hypotheses that involve two or more coefficients. We will also learn some basic ideas of assessing model specification.

1.2 Learning goals

- Know how to test a hypothesis for a single coefficient using the t-statistic.
- Know how to test a joint hypotheses for more than one coefficients using the F-statistic.
- Understand the underlying ideas of the F-statistic, especially when using the homoskedasticity-only F-statistic.

1.3 Reading materials

- Chapter 7 and Section 18.3 in *Introduction to Econometrics* by Stock and Watson.

2 Hypothesis Tests and Confidence Intervals For a Single Coefficient

We consider the general multiple regression model as follows

$$\mathbf{Y} = \beta_0\boldsymbol{\iota} + \beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2 + \cdots + \beta_k\mathbf{X}_k + \mathbf{u} \tag{1}$$

where $\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k,$ and $\mathbf{u}$ are $n \times 1$ vectors of the dependent variable, regressors, and errors. $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)'$ is the $(k+1) \times 1$ vector of coefficients. And $\boldsymbol{\iota}$ is the $n \times 1$ vector of 1s.

2.1 Standard errors for the OLS estimators

A review on $\mathrm{Var}(\hat{\boldsymbol{\beta}}|X)$

Recall that in the last lecture, we concluded that the the covariance matrix of the OLS estimators $\hat{\boldsymbol{\beta}}$ can take the following forms:

- The homoskedasticity-only covariance matrix if u_i is homoskedastic

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1} \tag{2}$$

- The heteroskedasticity-robust covariance matrix if u_i is heteroskedastic

$$\mathrm{Var_h}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\boldsymbol{\Sigma}(\mathbf{X}'\mathbf{X})^{-1} \tag{3}$$

  where $\boldsymbol{\Sigma} = \mathbf{X}'\boldsymbol{\Omega}\mathbf{X}$, and $\boldsymbol{\Omega} = \mathrm{Var}(\mathbf{u}|\mathbf{X})$.

Also, we know that if the least squares assumptions hold, $\hat{\boldsymbol{\beta}}$ has an asymptotic multivariate normal distribution as

$$\hat{\boldsymbol{\beta}} \xrightarrow{d} N(\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}) \tag{4}$$

where $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \mathrm{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X})$ for which use Equation (2) for the homoskedastic case and Equation (3) for the heteroskedastic case.

The estimator of $\text{Var}(\hat{\boldsymbol{\beta}}|X)$

In practice, σ_u^2 and $\boldsymbol{\Sigma}$ are unknown so that we need to estimate them using their sample counterparts.

- The estimator of σ_u^2 is

$$s_u^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i^2 \tag{5}$$

  Thus, the estimator of the homoskedasticity-only covariance matrix is

$$\widehat{\text{Var}(\hat{\boldsymbol{\beta}})} = s_u^2(\mathbf{X}'\mathbf{X})^{-1} \tag{6}$$

- The estimator of $\boldsymbol{\Sigma}$ is $\widehat{\boldsymbol{\Sigma}}$ given by

$$\widehat{\boldsymbol{\Sigma}} = \frac{n}{n-k-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i'\hat{u}_i^2 \tag{7}$$

  observation of $(k+1)$ regressors, including the constant term.

  Therefore, the heteroskedasticity-consistent (robust) covariance matrix estimator is

$$\widehat{\text{Var}_{\text{h}}(\hat{\boldsymbol{\beta}})} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\widehat{\boldsymbol{\Sigma}}(\mathbf{X}'\mathbf{X})^{-1} \tag{8}$$

- The estimator of $SE(\hat{\beta}_j)$

  Finally, we can get the standard error of $\hat{\beta}_j$ as the square root of the j^{th} diagonal element of $\widehat{\text{Var}(\hat{\boldsymbol{\beta}})}$ for homoskedasticity and $\widehat{\text{Var}_{\text{h}}(\hat{\boldsymbol{\beta}})}$ for heteroskedasticity. That is,

  - Homoskedasticity-only standard error: $SE(\hat{\beta}_j) = \left(\left[\widehat{\text{Var}(\hat{\boldsymbol{\beta}})}\right]_{(j,j)}\right)^{\frac{1}{2}}$
  - Heteroskedasticity-robust standard error: $SE(\hat{\beta}_j) = \left(\left[\widehat{\text{Var}_{\text{h}}(\hat{\boldsymbol{\beta}})}\right]_{(j,j)}\right)^{\frac{1}{2}}$

2.2 The t-statistic

With $SE(\hat{\beta}_j)$ at hand, we can test if a single coefficient β_j takes on a specific value, $\beta_{j,0}$. A two-sided hypothesis test suffices, that is,

$$H_0 : \beta_j = \beta_{j,0} \text{ vs. } H_1 : \beta_j \neq \beta_{j,0}$$

The basic ideas of hypothesis testing for a single coefficient in multiple regression are the same as in single regression. In this two-sided test, we still use the t-statistic computed as

$$t = \frac{\hat{\beta}_j - \beta_{j,0}}{SE(\hat{\beta}_j)}$$

Since $\hat{\boldsymbol{\beta}}$ has an asymptotic multivariate normal distribution, $\hat{\beta}_j$ has an asymptotic normal distribution. Under the null hypothesis that the true value of β_j is $\beta_{j,0}$, the t-statistic has a asymptotic standard normal distribution in large samples. Therefore the p-value can still be computed as

$$\text{p-value} = 2\Phi(-|t^{act}|)$$

The null hypothesis can be rejected at the 5% significant level when the p-value is less than 0.05, or equivalently, if $|t^{act}| > 1.96$. (Replace the critical value with 1.64 at the 10% level and 2.58 at the 1% level.)

2.3 Confidence intervals for a single coefficient

The confidence intervals for a single coefficient can be constructed as before using the t-statistic.

Given large samples, a 95% two-sided confidence interval for the coefficient β_j is

$$\left[\hat{\beta}_j - 1.96SE(\hat{\beta}_j),\ \hat{\beta}_j + 1.96SE(\hat{\beta}_j)\right]$$

2.4 Application to test scores and the student-teacher ratio

The regression with two explanatory variables, *STR* and *PctEL*

The regression of test has three estimated coefficients, the intercept, the coefficient on *STR* and the coefficient on *PctEl*. The estimated model can be written in the following format with the standard errors of the three coefficients reporeed in parentheses them.

$$\widehat{TestScore} = \underset{(8.7)}{686.0} - \underset{(0.43)}{1.10} \times STR - \underset{(0.031)}{0.650} \times PctEl$$

- We test $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$. The t-statistic for this test can be computed as $t = (-1.10-0)/0.43 = -2.54 < -1.96$, and the p-value is $2\Phi(-2.54) = 0.011 < 0.05$. Based on either the t-statistic or the p-value, we can reject the null hypothesis at the 5% level.

- The confidence interval that contains the true value of β_1 with a 95% probability can be computed as $-1.10 \pm 1.96 \times 0.43 = (-1.95, -0.26)$.

Adding expenditure per pupil to the equation

Now we add a new explanatory variable in the regression, *Expn*, that is the expenditure per pupil in the district in thousands of dollars. Note expenditure includes not only the spending on new computers, maintenance, and other hardware but also the salaries paid to teachers. So keep in mind that *Expn* and *STR* may be correlated. The new OLS regression line is

$$\widehat{TestScore} = \underset{(15.5)}{649.6} - \underset{(0.48)}{0.29} \times STR + \underset{(1.59)}{3.87} \times Expn - \underset{(0.032)}{0.656} \times PctEl$$

Let's see what's changed regarding *STR* after *Expn* is added.

- The magnitude of the coefficient on *STR* decreases from 1.10 to 0.29 after *Expn* is added.
- The standard error of the coefficient on *STR* increases from 0.43 to 0.48 after *Expn* is added.
- Consequently, in the new model, the t-statistic for the coefficient becomes $t = -0.29/0.48 = -0.60 > -1.96$ so that we cannot reject the zero hypothesis at the 5% level. (neither can we at the 10% level).
- How can we interpret such changes?
  - The decrease in the magnitude of the coefficient reflects that expenditure per pupil is an important factor that carry over most influence of student-teacher ratio on test scores. In other words, holding expenditure per pupil and the percentage of English-learners constant, reducing class sizes by hiring more teachers have only small effect on test scores.
  - The increase in the standard error reflects that *Expn* and *STR* are correlated so that there is imperfect multicollinearity in this model. In fact, the correlation coefficient between the two variables is 0.48, which is relatively high.

3 Tests of joint hypotheses

3.1 The form of joint hypotheses involving more than one coefficients

Rewrite the multiple regression model here

$$\mathbf{Y} = \beta_0\boldsymbol{\iota} + \beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2 + \cdots + \beta_k\mathbf{X}_k + \mathbf{u} \tag{9}$$

Since β_0 to β_k can take any value without restrictions, this model is referred to as the full model or **the unrestricted model**.

Joint hypothesis: an illustration using two zero restrictions

Suppose we want to test whether the coefficients on the first two regressors are zero. Then we can set up a joint hypothesis for these two coefficients like the following

$$H_0 : \beta_1 = 0, \beta_2 = 0, \text{ vs. } H_1 : \text{either } \beta_1 \neq 0 \text{ or } \beta_2 \neq 0 \text{ (or both)}$$

- This is a joint hypothesis because the two restrictions $\beta_1 = 0$ and $\beta_2 = 0$ must hold at the same time. So if either of them is invalid, the null hypothesis is rejected as a whole.
- To test these two restrictions jointly requires that we use a single statistic to test these restrictions simultaneously.
- The null hypothesis of $\beta_1 = 0$, $\beta_2 = 0$ can be considered as two restrictions imposed on Equation (9). If the null hypothesis is true, we have a **restricted model**

$$\mathbf{Y} = \beta_0 + \beta_3\mathbf{X}_3 + \beta_4\mathbf{X}_4 + \cdots + \beta_k\mathbf{X}_k + \mathbf{u} \tag{10}$$

Why not use t-statistic and test individual coefficients one at a time?

What if we test the joint null hypothesis using t-statistics for β_1 and β_2 separately. That is, compute the t-statistics t_1 for $\beta_1 = 0$ and t_2 for $\beta_2 = 0$. We call this "one-at-a-time" testing procedure. For simplicity, we assume t_1 and t_2 are independent.

We can show that the one-at-a-time procedure will commit a type I error with a probability more than 5%.

- A type I error happens when the null hypothesis is rejected when it is true. The probability of committing a type I error is call the size of the test. We want to control the size to be small, so we set the significance level (the prespecified probability of a type I error) at 1%, 5%, or 10%.

- Using the one-at-a-time procedure, at the 5% significance level, we can reject the null hypothesis of $H_0 : \beta_1 = 0$ and $\beta_2 = 0$ when either $|t_1| > 1.96$ or $|t_2| > 1.96$ (or both). In other words, the null is not rejected only when both $|t_1| \leq 1.96$ and $|t_2| \leq 1.96$.
- Because the two t-statistics are assumed to be independent, it implies that

  $$\Pr(|t_1| \leq 1.96 \text{ and } |t_2| \leq 1.96) = \Pr(|t_1| \leq 1.96) \times \Pr(|t_2| \leq 1.96) = 0.95^2 = 90.25\%$$

  So the probability of rejecting the null when it is true is $1 - 90.25\% = 9.75\%$.

More cases of joint hypothesis

- Joint hypothesis involving one coefficient in each restriction

  We can test whether the coefficients take some specific values.

  $$H_0 : \beta_1 = \beta_{1,0},\ \beta_2 = \beta_{2,0},\ \ldots,\ \beta_q = \beta_{q,0} \text{ versus}$$
  $$H_1 : \text{at least one restriction does not hold}$$

  Suppose that we are testing the joint zero hypotheses (i.e., $\beta_1 = \beta_2 = \cdots = \beta_q = 0$). This joint hypothesis imposes q zero restrictions on the unrestricted model (Equation (9)) so that **the restricted model** is

  $$\mathbf{Y} = \beta_0 + \beta_{q+1}\mathbf{X}_{q+1} + \beta_{q+2}\mathbf{X}_{q+2} + \cdots + \beta_k\mathbf{X}_k + \mathbf{u} \tag{11}$$

- Joint hypothesis involving multiple coefficients in each restriction

  Besides testing the hypothesis like $\beta_j = \beta_{j,0}$, we can also test **linear hypotheses** as follows,

  $$H_0 : \beta_1 = \beta_2 \text{ vs. } H_1 : \beta_1 \neq \beta_2$$

  or

  $$H_0 : \beta_1 + \beta_2 = 1 \text{ vs. } H_1 : \beta_1 + \beta_2 \neq 1$$

  or more generally,

  $$H_0 : \beta_1 + \beta_2 = 0,\ 2\beta_2 + 4\beta_3 + \beta_4 = 3 \text{ vs.}$$
  $$H_1 : \text{at least one restriction does not hold}$$

  All the null hypotheses above can be thought of being constructed using a linear function of the coefficients. So we can refer to them as linear hypotheses with regard

to $\boldsymbol{\beta}$.

A general joint hypothesis using matrix notation

We can use a matrix form to represent all linear hypotheses regarding the coefficients in Equation (9) as follows

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \text{ vs. } H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r} \tag{12}$$

where $\mathbf{R}$ is a $q \times (k+1)$ matrix with the **full row rank**, $\boldsymbol{\beta}$ represent the $k+1$ regressors, including the intercept, and $\mathbf{r}$ is a $q \times 1$ vector of real numbers.

For example

- For $H_0 : \beta_1 = 0, \beta_2 = 0$

$$\mathbf{R} = \begin{array}{c} \\ R1 \\ R2 \end{array} \begin{array}{c} \begin{array}{cccccc} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \cdots & \beta_k \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix} \end{array} \text{ and } \mathbf{r} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

- For $H_0 : \beta_1 + \beta_2 = 0$, $2\beta_2 + 4\beta_3 + \beta_4 = 3$, $\beta_1 = 2\beta_3 + 1$

$$\mathbf{R} = \begin{array}{c} \\ R1 \\ R2 \\ R3 \end{array} \begin{array}{c} \begin{array}{ccccccc} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_4 & \cdots & \beta_k \end{array} \\ \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 4 & 1 & \cdots & 0 \\ 0 & 1 & 0 & -2 & 0 & \cdots & 0 \end{pmatrix} \end{array} \text{ and } \mathbf{r} = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}$$

3.2 The F-statistic

We can compute the F-statistic to test all joint hypotheses shown above. Let's first review some properties of F distribution, which is the probability distribution that the F-statistic follows under the null hypothesis.

The general form of the F-statistic for testing the null hypothesis $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$

$$F = \frac{1}{q}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \left[\mathbf{R}\widehat{\text{Var}(\hat{\boldsymbol{\beta}})}\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \tag{13}$$

- $\hat{\boldsymbol{\beta}}$ is the estimated coefficients by OLS and $\widehat{\text{Var}(\hat{\boldsymbol{\beta}})}$ is the estimated covariance matrix.
  - For homoskedastic errors, we can compute $\widehat{\text{Var}(\hat{\boldsymbol{\beta}})}$ as in Equation (6)

- For heteroskedastic errors, we can compute $\widehat{\text{Var}_\text{h}(\hat{\boldsymbol{\beta}})}$ as in Equation (8)
- The F-statistic computed as in Equation (13) is a **heteroskedasticity-robust** F-statistic.

- The F distribution, the critical value, and the p-value

  If the least square assumptions hold, under the null hypothesis, the F-statistic is asymptotically distributed as the $F_{q,\infty}$ distribution. That is, $F \overset{a}{\sim} F(q, \infty)$

  The 5% critical value of the F distribution, c_α, must satisfy $\Pr(F < c_\alpha) = 0.95$. In other words, the p-value of the F test can be computed as $\Pr(F > F^{act})$.

  Note that we are computing the critical value and the p-value using the F distribution as if I were doing a one-sided test. This is because the F-statistic takes only positive values and the F distribution function is defined only in the domain of positive real numbers.

The F-statistic when $q = 2$

When we test the null hypothesis of $H_0 : \beta_1 = 0, \beta_2 = 0$ with the restricted model in Equation (10), the F-statistic for this test is

$$F = \frac{1}{2}\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1,t_2} t_1 t_2}{1 - \hat{\rho}_{t_1,t_2}^2} \tag{14}$$

Equation (14) is mostly for illustration purpose, which shows how to use t_1 and t_2 in a joint hypothesis test.

- For simplicity, suppose t_1 and t_2 are independent so that $\hat{\rho}_{t_1,t_2} = 0$. Then $F = \frac{1}{2}(t_1^2 + t_2^2)$.
- Under the null hypothesis, both t_1 and t_2 have asymptotic standard normal distribution. Then $t_1^2 + t_2^2 \sim \chi^2(2)$.
- It follows that $F = \frac{1}{2}(t_1^2 + t_2^2) \sim F(2, \infty)$.
- The discussion about the F-statistic in Equation (14) will become complicated when $\hat{\rho}_{t_1,t_2} \neq 0$.

3.3 The homoskedasticity-only F-statistic

When the regressor errors are homoskedastic, i.e., assuming that $\text{Var}(u_i|\mathbf{X}_i) = \sigma_u^2$ for $i = 1, \ldots, n$, then we can compute **the homoskedasticity-only F-statistic** that bears

more meaningful implications for the F tests.

Suppose we test the restricted model with q restrictions in Equation (11) versus the unrestricted model in Equation (9). That is,

$$H_0 : \mathbf{Y} = \beta_0 + \beta_{q+1}\mathbf{X}_{q+1} + \beta_{q+2}\mathbf{X}_{q+2} + \cdots + \beta_k\mathbf{X}_k + \mathbf{u} \text{ (the restricted model)}$$
$$H_1 : \mathbf{Y} = \beta_0 + \beta_1\mathbf{X}_1 + \beta_2\mathbf{X}_2 + \cdots + \beta_k\mathbf{X}_k + \mathbf{u} \text{ (the unrestricted model)}$$

Then the homoskedasticity-only F-statistic can be computed as

$$F = \frac{(RSSR - USSR)/q}{USSR/(n-k-1)} \tag{15}$$

where $RSSR$ is the sum of squared residuals of the restricted model and $USSR$ is the sum of squared residuals of the unrestricted model.

Since both restricted and unrestricted models have the same $\mathbf{Y}$, TSS is the same for both models. Therefore, dividing the numerator and the denominator in Equation (15) by TSS, we obtain another expression of the homoskedasticity-only F-statistic in terms of R^2 as

$$F = \frac{(R^2_{unrestrict} - R^2_{restrict})/q}{(1 - R^2_{unrestrict})/(n-k-1)} \tag{16}$$

Suppose that all least square assumptions and the homoskedasticity assumption hold, then we have

$$F \sim F(q, n-k-1)$$

So we can get the p-value and the critical value from the distribution $F(q, n-k-1)$.

We can understand the meaning of the homoskedasticity-only F-statistic by the following reasoning line

1. The unrestricted model have more regressors than the restricted model, on which the coefficients could be non-zero.
2. By the properties of the OLS estimation, SSR will decrease whenever an additional regressor is included in the model and the coefficient on that new regressor is not zero.
3. In other words, given the same sample, R^2 in the unrestricted model will increase when a new regressor is added with a nonzero coefficient.
4. That means $RSSR \geq USSR$ and $R^2_{unrestrict} \geq R^2_{restrict}$ are always true.
5. However, suppose that the null hypothesis is true. That is, the true model is really the restricted one.

6. Then, the explanatory power of the additional regressors in the unrestricted model should be very small.
7. That means that $USSR$ cannot be too much smaller than $RSSR$, or $R^2_{unrestrict}$ cannot be too much larger than $R^2_{restrict}$ if the null hypothesis is true.
8. That means F should not be a large positive number under the null hypothesis.
9. If we compute an F-statistic that is large enough compared with a critical value at some significance level, then we can reject the null hypothesis.

3.4 Transformation of joint hypothesis testing to single hypothesis testing

For some simple joint hypotheses, we can transform the model so that tesing joint hypotheses is converted to testing a single hypothesis. Consider the following model

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \mathbf{u}$$

And the null hypothesis is

$$H_0 : \beta_1 = \beta_2$$

Then we can rewrite the model as

$$\mathbf{Y} = \beta_0 + (\beta_1 - \beta_2)\mathbf{X}_1 + \beta_2(\mathbf{X}_1 + \mathbf{X}_2) + \mathbf{u}$$

Define $\gamma = \beta_1 - \beta_2$ and $\mathbf{W} = \mathbf{X}_1 + \mathbf{X}_2$. Then the original model becomes

$$\mathbf{Y} = \beta_0 + \gamma \mathbf{X}_1 + \beta_2 \mathbf{W} + \mathbf{u}$$

Thus, instead of testing $\beta_1 - \beta_2 = 0$, we test $H_0 : \gamma = 0$ using the t-statistic computed from the transformed model.

3.5 Application to test scores and the student-teacher ratio

We rewrite the estimated regression model of test scores against the student-teacher ratio, expenditures per pupil, and the percentage of English learners below.

$$\widehat{TestScore} = \underset{(15.5)}{649.6} - \underset{(0.48)}{0.29} \times STR + \underset{(1.59)}{3.87} \times Expn - \underset{(0.032)}{0.656} \times PctEl,\ R^2 = 0.4366$$

The null hypothesis is $H_0 : \beta_1 = 0,$ and $\beta_2 = 0$, and the alternative hypothesis is $H_1 : \beta_1 \neq 0$ or $\beta_2 \neq 0$.

- The heteroskedasticity-robust F statistic is 5.43, calculated by the computer program using the heteroskedasticity-consistent covariance matrix. The critical value of the $F_{2,\infty}$ distribution at the 5% significance level is 3.00, and 4.61 at the 1% level. Since $F = 5.43 > 4.61$, we can reject the null hypothesis saying that neither the student-teacher ratio nor expenditures per pupil have an effect on test scores, holding constant the percentage of English learners.

- The homoskedasticity-only F statistic. To compute the homoskedasticity-only F statistic, we need to estimate the restricted model by OLS, which yields

$$\widehat{TestScore} = \underset{(1.0)}{664.7} - \underset{(0.032)}{0.671} \times PctEl,\ R^2 = 0.4149$$

  Now we know that the unrestricted $R^2_{\text{unrestricted}}$ is 0.4366, the restricted $R^2_{\text{restricted}}$ is 0.4149, the number of restrictions $q = 2$, the number of observations $n = 420$, and the number of coefficients in the unrestricted model $k = 3$. Then, the homoskedasticity-only F statistic is computed as

$$F = \frac{(0.4366 - 0.4149)/2}{(1 - 0.4366)/(420 - 3 - 1)} = 8.01$$

  Because 8.01 exceeds the 1% critical value of 4.61 from the $F_{2,\infty}$ distribution, the null hypothesis is rejected.

4 Confidence Sets for multiple coefficients

4.1 Definition

A **95% confidence set** for two or more coefficients is

- a set that contains the true population values of these coefficients in 95% of randomly drawn samples.

- Equivalently, the set of coefficient values that cannot be rejected at the 5% significance level.

4.2 How to construct a confidence set

Suppose that we construct the confidence set for $\beta_1 = \beta_{1,0}, \beta_2 = \beta_{2,0}$.

- Let F_{β_1,β_2} be the heteroskedasticity-robust F-statistic computed according to Equation (13). If the homoskedasticity assumption holds, then F-statistic can be computed based on Equation (15).
- A 95% confidence set $= \{\beta_1, \beta_2 : F_{\beta_1,\beta_2} < c_F\}$, where c_F is the 5% critical value of the $F(2, \infty)$ distribution, which is close to 3 in this case.
- This set has coverage rate 95% because the test on which it is based has the size of 5%.
- Therefore the confidence set which is constructed as the non-rejected values contains the true value 95% of the time.

4.3 The confidence set based on the F-statistic is an ellipse

According to Equation (14), the confidence set for β_1, and β_2 is

$$\left\{\beta_1, \beta_2 : F = \frac{1}{2}\frac{t_1^2 + t_2^2 - 2\hat{\rho}_{t_1,t_2} t_1 t_2}{1 - \hat{\rho}_{t_1,t_2}^2} \leq 3\right\}$$

Plugging the formula of t_1 and t_2, the F-statistic becomes

$$F = \left[\left(\frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}\right)^2 + \left(\frac{\hat{\beta}_2 - \beta_{2,0}}{SE(\hat{\beta}_2)}\right)^2 + 2\hat{\rho}_{t_1,t_2}\left(\frac{\hat{\beta}_1 - \beta_{1,0}}{SE(\hat{\beta}_1)}\right)\left(\frac{\hat{\beta}_2 - \beta_{2,0}}{SE(\hat{\beta}_2)}\right)\right] \leq 3$$

which is an ellipse containing the pairs of values of β_1 and β_2 that cannot be rejected using the F-statistic at the 5% significance level. See Figure 1.

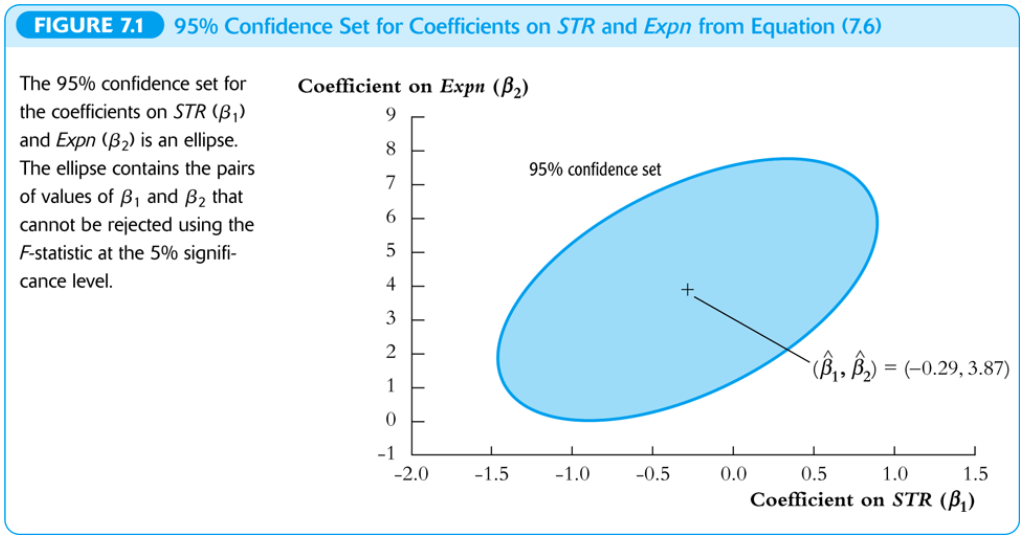

Figure 1: 95% Confidence Set for Coefficients on STR and Expn

5 Model specification for multiple regression

5.1 Omitted variable bias in multiple regression

Omitted variable bias is the bias in the OLS estimator that arises when one or more included regressors are correlated with an omitted variable.

For omitted variable bias to arise, two things must be true:

1. At least one of the included regressors must be correlated with the omitted variable.
2. The omitted variable must be a determinant of the dependent variable, Y.

With omitted variable bias, the least square assumption $E(u|X) = 0$ does not hold any more. The OLS estimator $\hat{\beta}$ is biased however large the sample size is.

5.2 The problem of the assumption of $E(u|X) = 0$ and control variables

The assumption of $E(u|X) = 0$ ensures that the estimated coefficients on all included regressors are unbiased and consistent. However, this assumption is too strong to be completely realized in practice. So we have to make a compromise between the ideal situation and reality.

What we can do is that we divide all regressors into two groups:

- One group consists of regressors whose causal effects on Y are our research interest so that we want unbiased estimates of these coefficients.
- Another group consists of regressors whose causal effects on Y are not our focus. But if we omit them, we would risk making omitted variable bias in the coefficients that we do care.

The regressors in the latter group are called **control variable**. Moreover, we use an assumption that is weaker than the assumption of $E(u|X) = 0$ to ensure that the estimated coefficients on the regressors in the first groups are unbiased, maintaining the causal implication that we want.

5.3 The role of control variables in multiple regression

Definition

A control variable W is a variable that is correlated with, and controls for, an omitted causal factor in the regression of Y on X, but which itself does not necessarily have a causal effect on Y.

A control variable is not the object of interest in the study; rather it is a regressor included to hold constant factors that, if neglected, could lead to the estimated causal effect of interest to suffer from omitted variable bias.

The test score example

$$TestScore = \underset{(5.6)}{700.2} - \underset{(0.27)}{1.00}STR - \underset{(0.033)}{0.122}PctEL - \underset{(0.024)}{0.547}LchPct,\ \bar{R}^2 = 0.773$$

Where $PctEL$ = percent English learners in the school district, $LchPct$ = percent of students receiving a free/subsidized lunch.

- Which variable is the variable of interest? STR
- Which variables are control variables? Do they have causal implications? What do they control for?
  - $PctEL$ probably has a direct causal effect (school is tougher if you are learning English!). But it is also a control variable: immigrant communities tend to be less affluent and often have fewer outside learning opportunities, and $PctEL$ is correlated with those omitted causal variables. $PctEL$ is both a possible causal variable and a control variable.
  - $LchPct$ might have a causal effect (eating lunch helps learning); it is also correlated with and controls for income-related outside learning opportunities. $LchPct$ is both a possible causal variable and a control variable.

What makes an effective control variable?

What variables can we choose to include in regression as effective control variables? The followings are three interchangeable statements about what makes an effective control variable:

- An effective control variable is one which, when included in the regression, makes the error term uncorrelated with the variable of interest.
- Holding constant the control variable(s), the variable of interest is "as if" randomly assigned.
- Among individuals (entities) with the same value of the control variable(s), the variable of interest is uncorrelated with the omitted determinants of Y.

Control variables need not be causal, and their coefficients generally do not have a causal interpretation.

- Does the coefficient on *LchPct* have a causal interpretation? If so, then we should be able to boost test scores (by a lot! Do the math!) by simply eliminating the school lunch program, so that $LchPct = 0$. But it makes nonsense!

5.4 Conditional mean independence

We need a mathematical statement of what makes an effective control variable, which is **conditional mean independence**.

Conditional mean independence says that given the control variable(s), the expectation of u_i doesn't depend on the variable of interest.

Let X_i denote the variable of interest and W_i denote the control variable(s). W is an effective control variable if conditional mean independence holds:

$$E(u_i|X_i, W_i) = E(u_i|W_i)$$

If W is an effective control variable, then we can use conditional mean independence to substitute the first least square assumption requiring $E(u_i|X_i, W_i) = 0$.

Consider the regression model

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u$$

where X is the variable of interest and W is an effective control variable so that conditional mean independence holds. In addition, suppose that the other least square assumptions hold. Then, it has the following tow implications: (1) β_1 has a causal interpretation and (2) $\hat{\beta}_1$ is unbiased and $\hat{\beta}_2$ is biased. Let's see the reasons.

β_1 has a causal interpretation.

It means that controlling for W, the causal effect of X on Y is measured by β_1. That is, we can still interpret $\beta_1 = \Delta Y/\Delta X$, holding other things constant by controlling for W.

The expected change in Y resulting from a change in X, holding W constant, is:

$$\begin{aligned}
&E(Y|X = x + \Delta x, W = w) - E(Y|X = x, W = w) \\
&= \beta_0 + \beta_1(x + \Delta x) + \beta_2 w + E(u|X = x + \Delta x, W = w) \\
&\quad - \beta_0 + \beta_1 x + \beta_2 w + E(u|X = x, W = w) \\
&= \beta_1 \Delta x + [E(u|W = w) - E(u|W = w)] \\
&= \beta_1 \Delta x
\end{aligned}$$

In the second equality, we use conditional mean independence $E(u|X = x + \Delta x, W = w) = E(u|X = x, W = w) = E(u|W = w)$.

$\hat{\beta}_1$ is unbiased and $\hat{\beta}_2$ is biased

For convenience, suppose that $E(u|W) = \gamma_0 + \gamma_1 W$. Thus, under conditional mean independence, we have

$$E(u|X, W) = E(u|W) = \gamma_0 + \gamma_1 W$$

Let $v = u - E(u|W)$ so that

$$E(v|X, W) = E(u|X, W) - E(u|W) = 0$$

Then, it follows that

$$u = E(u|X, W) + v = \gamma_0 + \gamma_1 W + v \ .$$

Then, the original model $Y = \beta_0 + \beta_1 X + \beta_2 W + u$ becomes

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \beta_2 W + \gamma_0 + \gamma_1 W + v \\ &= (\beta_0 + \gamma_0) + \beta_1 X + (\beta_2 + \gamma_1) W + v \\ &= \delta_0 + \beta_1 X + \delta_2 W + v \end{aligned}$$

where $\delta_0 = \beta_0 + \gamma_0$ and $\delta_2 = \beta_2 + \gamma_2$.

For the new model $Y = \delta_0 + \beta_1 X + \delta_2 W + v$, we can conclude as follows.

- The new model satisfy $E(v|X, W) = 0$ so that the OLS estimator of $\delta_0, \beta_1,$ and δ_2 are unbiased.
- The estimated coefficients in the original model are actually $\hat{\beta}_1$ and $\hat{\delta}_2$, which we know that $E(\hat{\beta}_1) = \beta_1$ and $E(\hat{\beta}_2) = \delta_2 \neq \beta_2$ in general.

5.5 Model specification in theory and in practice

In theory, when data are available on the omitted variable, the solution to omitted variable bias is to include the omitted variable in the regression. In practice, however, deciding whether to include a particular variable can be difficult and requires judgment.

The following steps are advocated to set up a regression model:

1. a base set of regressors should be chosen using a combination of expert judgment, economic theory, and knowledge of how data were collected. The regression using

this base set of regressors is referred to as a **base specification**. This step involves the following consideration:

   (a) identifying the variable of interest.

   (b) thinking of the omitted causal effects that could result in omitted variable bias.

   (c) including those omitted causal effects if you can find relevant variables. If you can't, include variables that are correlated with them as control variables. The control variables are effective if the conditional mean independence assumption plausibly holds.

2. Specify a range of plausible **alternative model specifications**, which include additional candidate variables.

   (a) If the estimates of the coefficients of interest are numerically similar across the alternative specifications, then this provides evidence that the estimates from your base specification are reliable.

   (b) If the estimates of the coefficients of interest change substantially across specifications, this often provides evidence that the original specification had omitted variable bias.

3. Use test statistics to judge a model specification

   (a) Use R^2 and $\bar{R}^2$ to see the overall goodness of fit of a model specification. Caution: a high R^2 or $\bar{R}^2$ does not mean that you have eliminated omitted variable bias. Neither does a high R^2 or $\bar{R}^2$ mean that the included variables and the model as a whole are statistically significant.

   (b) Use t-statistic to check the significance of individual coefficients, and use F-statistic to check the overall significance of the model as a whole. That is, use F test for

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

6 Analysis of the test score data set

The complete regression results are formally reported in Table 7.1.

TABLE 7.1 Results of Regressions of Test Scores on the Student–Teacher Ratio and Student Characteristic Control Variables Using California Elementary School Districts

Dependent variable: average test score in the district.

Regressor	(1)	(2)	(3)	(4)	(5)
Student–teacher ratio (X_1)	−2.28** (0.52)	−1.10* (0.43)	−1.00** (0.27)	−1.31** (0.34)	−1.01** (0.27)
Percent English learners (X_2)		−0.650** (0.031)	−0.122** (0.033)	−0.488** (0.030)	−0.130** (0.036)
Percent eligible for subsidized lunch (X_3)			−0.547** (0.024)		−0.529** (0.038)
Percent on public income assistance (X_4)				−0.790** (0.068)	0.048 (0.059)
Intercept	698.9** (10.4)	686.0** (8.7)	700.2** (5.6)	698.0** (6.9)	700.4** (5.5)
Summary Statistics					
SER	18.58	14.46	9.08	11.65	9.08
$\overline{R}^2$	0.049	0.424	0.773	0.626	0.773
n	420	420	420	420	420

These regressions were estimated using the data on K-8 school districts in California, described in Appendix 4.1. Heteroskedasticity-robust standard errors are given in parentheses under coefficients. The individual coefficient is statistically significant at the *5% level or **1% significance level using a two-sided test.