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

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

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

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

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

- 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

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

- 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$^{\text{th}}$ diagonal element of $\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}})}$ for homoskedasticity and $\widehat{\mathrm{Var}_{\mathrm{h}}(\hat{\boldsymbol{\beta}})}$ for heteroskedasticity. That is,

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

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

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

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.

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

  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,

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

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

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

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 $H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$

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

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

- For heteroskedastic errors, we can compute $\widehat{\mathrm{Var}_{\mathrm{h}}(\hat{\boldsymbol{\beta}})}$ as in Equation (8)

- 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 $\hat{\rho}_{t_1,t_2} \neq 0$.