Lecture 9: Hypothesis Tests and Confidence Intervals in Multiple Regression

Zheng Tian

Outline

1. Hypothesis Tests and Confidence Intervals For a Single Coefficient
2. Tests of Joint Hypotheses
3. Confidence Sets for multiple coefficients
4. Warm-up Exercises
5. Model Specification for Multiple Regression

1. Hypothesis Tests and Confidence Intervals For a Single Coefficient
2. Tests of Joint Hypotheses
3. Confidence Sets for multiple coefficients
4. Warm-up Exercises
5. Model Specification for Multiple Regression

The basic multiple regression model

Consider the following model

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

- $\mathbf{Y}, \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k,$ and $\mathbf{u}$ are $n \times 1$ vectors of the dependent variable, regressors, and errors
- $\beta_0, \beta_1, \beta_2, \ldots,$ and β_k are parameters.
- $\boldsymbol{\iota}$ is the $n \times 1$ vector of 1s.

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

- The homoskedasticity-only covariance matrix if u_i is homoskedastic

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

- The heteroskedasticity-robust covariance matrix if u_i is heteroskedastic

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

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

The multivariate normal distribution of $\hat{\beta}$

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

The estimator of σ_u^2

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

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

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

The estimator of $\mathbf{\Sigma}$

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

where $\mathbf{X}_i$ is the vector of the i$^{\text{th}}$ observation of $(k+1)$ regressors. The heteroskedasticity-consistent (robust) covariance matrix estimator is

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

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

We can get the standard error of $\hat{\beta}_j$ as the square root of the j^{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: $SE(\hat{\beta}_j) = \left(\left[\widehat{\mathrm{Var}_{\mathrm{h}}(\hat{\boldsymbol{\beta}})} \right]_{(j,j)} \right)^{\frac{1}{2}}$

The t-statistic

We can perform a two-sided hypothesis test as

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

- We still use the t-statistic, computed as $t = (\hat{\beta}_j - \beta_{j,0})/SE(\hat{\beta}_j)$, where $SE(\hat{\beta}_j)$ is the standard error of $\hat{\beta}_j$.
- Under the null hypothesis, we have, in large samples, $t \overset{a}{\sim} N(0,1)$. Therefore, the p-value can still be computed as $2\Phi(-|t^{act}|)$.
- The null hypothesis is 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.)

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

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

Application to test scores and the student-teacher ratio

The estimated model can be written as follows

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

- 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.

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

- 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 some 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.

The unrestricted model

Consider the following multiple regression model

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

We call Equation (9) as the full model or the unrestricted model because β_0 to β_k can take any value without restrictions.

Joint hypothesis: a case of two zero restrictions

- Question: Are the coefficients on the first two regressors zero?
- Joint hypotheses

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

- This is a joint hypothesis because $\beta_1 = 0$ 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.

The restricted model with two zero restrictions

- If the null hypothesis is true, we have

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

  We call Equation (10) as the restricted model because we impose two restrictions $\beta_1 = 0$ and $\beta_2 = 0$.

- To test these two restrictions jointly means that we need to use a single statistic to test these restrictions simultaneously. That statistic is F-statistic.

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

- Let us test the null hypothesis above using t-statistics for β_1 and β_2 separately. That is, t_1 is the t-statistic for $\beta_1 = 0$ and t_2 is the t-statistic for $\beta_2 = 0$.
- 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.

What's the problem with the one-at-a-time procedure

How can we reject the null hypothesis with this procedure?

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 $|t_2| \leq 1.96$.

What is the probability of committing Type I error?

Assume t_1 and t_2 to be independent. Then,

$$\Pr(|t_1| \leq 1.96 \ \& \ |t_2| \leq 1.96) = \Pr(|t_1| \leq 1.96)\Pr(|t_2| \leq 1.96) = 0.95^2 = 90.25\%$$

So the probability of rejecting the null when it is true is $1 - 90.25\% = 9.75\%$. We may reject the null hypothesis with a higher probability than what we have pre-specified with the significant level.

Joint hypothesis involving one coefficient for each restriction

q restrictions

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

The restricted model

Suppose that we are testing the q zero hypotheses, that is, q restrictions, $\beta_1 = \beta_2 = \cdots = \beta_q = 0$. 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 linear hypotheses

Joint hypotheses include linear hypotheses like the followings

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

2.
$$H_0: \beta_1 + \beta_2 = 1 \text{ vs. } H_1: \beta_1 + \beta_2 \neq 1$$

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

A general form of joint hypotheses

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.

Examples of $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$

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

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

The general form of the F-statistic

To test the null hypothesis

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

we compute the F-statistic

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

- $\hat{\beta}$ is the estimated coefficients by OLS and $\widehat{\mathrm{Var}(\hat{\beta})}$ is the estimated covariance matrix.
- For homoskedastic errors, we can compute $\widehat{\mathrm{Var}(\hat{\beta})}$ as in Equation (6)
- For heteroskedastic errors, we can compute $\widehat{\mathrm{Var}_{\mathrm{h}}(\hat{\beta})}$ as in Equation (8)

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

The F distribution

If the least square assumptions hold, under the null hypothesis, the F-statistic is asymptotically distributed as F distribution with degree of freedom (q, ∞). That is,

$$F \overset{a}{\sim} F(q, \infty)$$

The critical value and the p-value of F test.

The 5% critical value of the F test using F-statistic is c_α such that

$$\mathrm{Pr}(F < c_\alpha) = 0.95$$

And the p-value of F test can be computed as

$$p\text{-value} = \mathrm{Pr}(F > F^{act})$$

The F-statistic when $q = 2$

The F-statistic for testing the null hypothesis of $H_0 : \beta_1 = 0, \beta_2 = 0$ can be proved to take the following form,

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

- 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 standard normal distribution asymptotically. Then $t_1^2 + t_2^2$ has a chi-squared distribution with 2 degrees of freedom.
- It follows that $F = \frac{1}{2}(t_1^2 + t_2^2)$ has asymptotically distributed as $F(2, \infty)$.
- The discussion about F-statistic in Equation (14) will become complicated when $\hat{\rho}_{t_1,t_2} \neq 0$.

The homoskedasticity-only F-statistic

- Assume that $\mathrm{Var}(u_i|\mathbf{X}_i) = \sigma_u^2$ for $i = 1, \ldots, n$, i.e., u_i has a constant conditional variance for all i.
- Test hypotheses

$$\begin{aligned} H_0 &: \text{the restricted model with q restrictions } \mathbf{R}\beta = \mathbf{r} \\ H_1 &: \text{the unrestricted model } \mathbf{Y} = \beta_0 + \beta_1 \mathbf{X}_1 + \cdots + \beta_k \mathbf{X}_k + \mathbf{u} \end{aligned}$$

- The homoskedasticity-only F-statistic can be computed as

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

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

The homoskedasticity-only F-statistic (cont'd)

- Dividing the numerator and the denominator in F-statistic by TSS, we get 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 know that

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

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

Understanding the homoskedasticity-only F-statistic

- $RSSR \geq USSR$ and $R^2_{unrestrict} \geq R^2_{restrict}$ are always true. Why?
    - The unrestricted model have more regressors than the restricted model.
    - *SSR* will decrease whenever an additional regressor is included in the model and the coefficient on the new regressor is not zero.
    - In other words, R^2 in the unrestricted model will increase when a new regressor is added with a nonzero coefficient.

Understanding the homoskedasticity-only F-statistic (cont'd)

- Suppose that the null hypothesis is true. That is, the true model is really the restricted one. What would this imply?
    - The explanatory power of the additional regressors in the unrestricted model should be very small.
    - 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.
    - That means F should not be a large positive number under the null hypothesis.
    - 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.

Transformation of joint hypothesis testing to single hypothesis testing

- The goal: to convert a joint hypothesis to a single hypothesis.
- Consider the following model

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

- The null hypothesis is $H_0 : \beta_1 = \beta_2$.
- How can we convert H_0 to a single hypothesis test?

Transformation of joint hypothesis testing to single hypothesis testing (cont'd)

- Rewrite the model

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

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

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

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

Application of the California school districts

- The estimated model

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

- The joint hypothesis

$$H_0:\ \beta_1 = 0,\ \text{ and } \beta_2 = 0 \text{ v.s. } H_1:\ \beta_1 \neq 0 \text{ or } \beta_2 \neq 0$$

- The heteroskedasticity-robust F statistic is 5.43. The critical value of the $F_{2,\infty}$ distribution at the 5% significance level is 3.00. So we reject the null hypothesis.

Application to test scores and the student-teacher ratio (cont'd)

- The homoskedasticity-only F statistic.

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

- Now we have $R^2_{\text{unrestricted}} = 0.4366$, $R^2_{\text{restricted}} = 0.4149$, $q = 2$, $n = 420$, and $k = 3$.
- 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.

Check the critical value using the F table

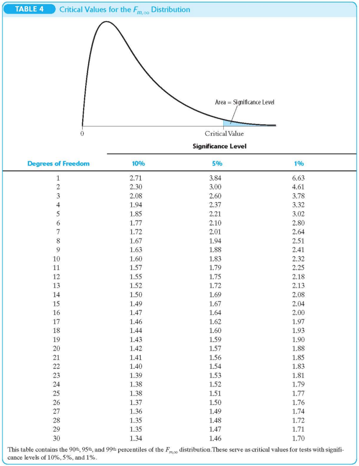
Area = Significance Level
0
Critical Value

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

	Significance Level		
Degrees of Freedom	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%.

Figure: The table of critical values of the $F(q, \infty)$ distribution

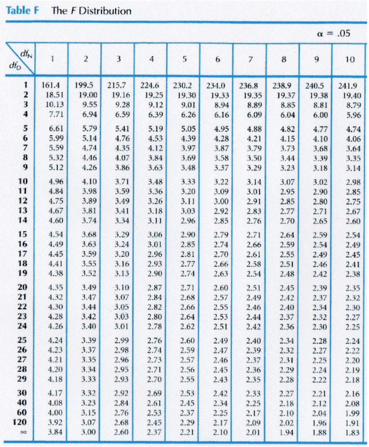

Table F The F Distribution

$\alpha = .05$

df_N \ df_D	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Figure: The table of critical values at the 5% level of the $F(m, n)$ distribution

1. Hypothesis Tests and Confidence Intervals For a Single Coefficient

2. Tests of Joint Hypotheses

3. Confidence Sets for multiple coefficients

4. Warm-up Exercises

5. Model Specification for Multiple Regression

Confidence set: 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.

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 or homoskedasticity-only F-statistic.
- A 95% confidence set is

$$\{\beta_1, \beta_2 : F_{\beta_1,\beta_2} < c_F\}$$

  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 constructed as the nonrejected values contains the true value 95% of the time.

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

- According to Equation (14), the confidence set for $\beta_1,$ and β_2 is

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

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

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

The confidence set: an illustration

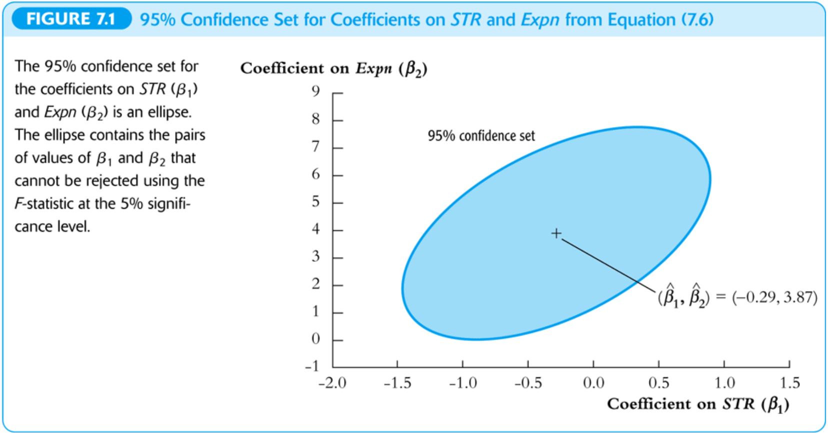

FIGURE 7.1 95% Confidence Set for Coefficients on *STR* and *Expn* from Equation (7.6)

The 95% confidence set for the coefficients on *STR* (β_1) and *Expn* (β_2) is an ellipse. The ellipse contains the pairs of values of β_1 and β_2 that cannot be rejected using the *F*-statistic at the 5% significance level.

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

Question 1

The following linear hypothesis can be tested using the F-test with the exception of

A) $\beta_2 = 1$ and $\beta_3 = \beta_4/\beta_5$

B) $\beta_2 = 0$

C) $\beta_1 + \beta_2 = 1$ and $\beta_3 = -2\beta_4$

D) $\beta_0 = \beta_1$ and $\beta_1 = 0$

Question 1

The following linear hypothesis can be tested using the F-test with the exception of

A) $\beta_2 = 1$ and $\beta_3 = \beta_4/\beta_5$

B) $\beta_2 = 0$

C) $\beta_1 + \beta_2 = 1$ and $\beta_3 = -2\beta_4$

D) $\beta_0 = \beta_1$ and $\beta_1 = 0$

Answer: A

Question 2

To test joint linear hypotheses in the multiple regression model, you need to

A) compare the sums of squared residuals from the restricted and unrestricted model.

B) use the heteroskedasticity-robust F-statistic.

C) use several t-statistics and perform tests using the standard normal distribution.

D) compare the adjusted R^2 for the model which imposes the restrictions, and the unrestricted model.

Question 2

To test joint linear hypotheses in the multiple regression model, you need to

A) compare the sums of squared residuals from the restricted and unrestricted model.

B) use the heteroskedasticity-robust F-statistic.

C) use several t-statistics and perform tests using the standard normal distribution.

D) compare the adjusted R^2 for the model which imposes the restrictions, and the unrestricted model.

Answer: B

Question 3

Let $R^2_{\text{unrestricted}}$ and $R^2_{\text{restricted}}$ be 0.4366 and 0.4149 respectively. The difference between the unrestricted and the restricted model is that you have imposed two restrictions. There are 420 observations and 3 regressors including the intercept. The homoskedasticity-only F-statistic in this case is

A) 4.61

B) 8.01

C) 10.34

D) 7.71

Question 3

Let $R^2_{\text{unrestricted}}$ and $R^2_{\text{restricted}}$ be 0.4366 and 0.4149 respectively. The difference between the unrestricted and the restricted model is that you have imposed two restrictions. There are 420 observations and 3 regressors including the intercept. The homoskedasticity-only F-statistic in this case is

A) 4.61
B) 8.01
C) 10.34
D) 7.71

Answer: B

Question 4

The formulation $\boldsymbol{R}\beta = \boldsymbol{r}$ to test a hypotheses

A) allows for restrictions involving both multiple regression coefficients and single regression coefficients.

B) is F-distributed in large samples.

C) allows only for restrictions involving multiple regression coefficients.

D) allows for testing linear as well as nonlinear hypotheses.

Question 4

The formulation $\boldsymbol{R}\beta = \boldsymbol{r}$ to test a hypotheses

A) allows for restrictions involving both multiple regression coefficients and single regression coefficients.

B) is F-distributed in large samples.

C) allows only for restrictions involving multiple regression coefficients.

D) allows for testing linear as well as nonlinear hypotheses.

Answer: A

1 Hypothesis Tests and Confidence Intervals For a Single Coefficient

2 Tests of Joint Hypotheses

3 Confidence Sets for multiple coefficients

4 Warm-up Exercises

5 Model Specification for Multiple Regression

Omitted variable bias in multiple regression

- We should always be alert to omitted variable bias in the OLS estimation.
- 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 problem of the assumption of $E(u|X) = 0$ and control variables

- The assumption of $E(u|X) = 0$ is essential to ensures unbiasedness and consistency. However, it is too strong to be completely realized in practice.
- What we can do is that we divide all regressors into two groups:
  1. One group consists of regressors whose causal effects on Y are our research interest so that we want unbiased estimates of these coefficients.
  2. 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.
- W 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.

Control variables: definition

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.

The role of control variables

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

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

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?

What makes an effective control variable?

- 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.

Conditional mean independence: definition

- Conditional mean independence: given the control variable, the mean 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:

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

- Conditional mean independence substitute the first least square assumption requiring $E(u_i|X_i, W_i) = 0$.

Implications of conditional mean independence

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,

- β_1 has a causal interpretation.
- $\hat{\beta}_1$ is unbiased.
- The coefficient on the control variable, $\hat{\beta}_2$ is in general biased.

β_1 has a causal interpretation

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

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

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

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

$\hat{\beta}_1$ is unbiased and $\hat{\beta}_2$ is biased (cont'd)

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 $\delta_2 = \beta_2 + \gamma_2$.

What can we conclude from the new model?

- 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.

Model specification in theory and in practice

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

Step 1: Set up a base specification

A core or 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:

1. identifying the variable of interest.
2. thinking of the omitted causal effects that could result in omitted variable bias.
3. including those omitted causal effects if you can. If you can't, include variables correlated with them that serve as control variables.

Model specification in theory and in practice (cont'd)

Step 2: Set up alternative specifications

Also specify a range of plausible alternative model specifications, which include additional candidate variables.

1. 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.
2. If the estimates of the coefficients of interest change substantially across specifications, this often provides evidence that the original specification had omitted variable bias.

Model specification in theory and in practice (cont'd)

Step 3: Use test statistics to judge a model specification

1. 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.
2. 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

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

Analysis of the test score data set

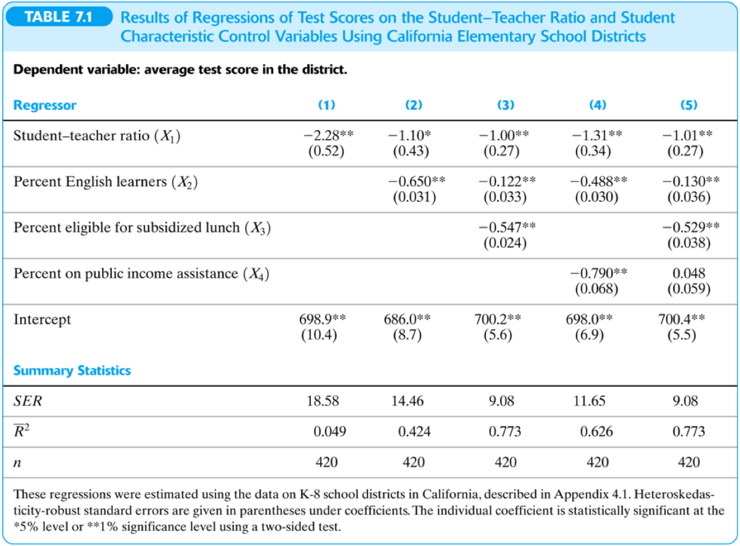

TABLE 7.1 Results of Regressions of Test Scores on the Student–Teacher Ratio and Student Characteristic Control Variables Using California Elementary School Districts

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.