Lecture 10 Multiple Regression Analysis - Inference Pt. 2

Single linear restriction

Let's begin with our population model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U.$$

Previously, we have learned how to conduct hypotheses testing about a single coefficient, e.g., $\beta_j = b, j = 0, \ldots, k$. However, often we would like to test hypotheses involving more than one coefficient. The simplest case of that would be when we would like to test the equality of any two coefficients, say j and l ($j \neq l$):

$$H_0 : \beta_j = \beta_l.$$

Suppose our alternative hypothesis is two-sided

$$H_1 : \beta_j \neq \beta_l.$$

We can no longer use t statistics for individual coefficients to conduct hypothesis testing. What we can do, is construct a new test statistic of the following form

$$t = \frac{\hat{\beta}_j - \hat{\beta}_l}{\mathrm{se}(\hat{\beta}_j - \hat{\beta}_l)}.$$

Once we compute this test statistic, we can proceed as before: by either comparing the t-statistic with the appropriate critical value, or computing the p-value of the statistic and comparing it with the significance level, or using a confidence interval.

There is one difficulty involved in computing this new test statistic. While we can easily compute the numerator, the standard error in the denominator is not readily available. We would have to use the following property of variances:

$$\mathrm{Var}(\hat{\beta}_j - \hat{\beta}_l) = \mathrm{Var}(\hat{\beta}_j) + \mathrm{Var}(\hat{\beta}_l) - 2\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_l).$$

Then the standard error we are looking for is the squared root of the expression above:

$$\mathrm{se}(\hat{\beta}_j - \hat{\beta}_l) = \mathrm{Var}(\hat{\beta}_j - \hat{\beta}_l)^{1/2}.$$

Notice that we would need to know the covariance between $\hat{\beta}_j$ and $\hat{\beta}_l$, which we can obtain from the full covariance matrix of β.

There is an alternative way to conduct this hypothesis test. Instead of modifying the t statistic, we can instead modify the *model* and use an existing t statistic. How can we achieve that? We know how test simple hypothesis of a kind $H_0 : \theta = 0$. If we can somehow re-write the model so that $\theta = \beta_j - \beta_l$, then testing the hypothesis about θ would be equivalent to testing the hypothesis about the equality of β_j and β_l.

Let's re-write the equality $\theta = \beta_j - \beta_l$ as $\beta_j = \theta + \beta_l$. Then we can re-write our population model as

$$\begin{aligned} Y &= \beta_0 + \beta_j X_j + \beta_l X_l + \ldots + U \\ &= \beta_0 + (\theta + \beta_l) X_j + \beta_l X_l + \ldots + U \\ &= \beta_0 + \theta X_j + \beta_l (X_j + X_l) + \ldots + U \end{aligned}$$

The resulting model is identical to the previous one, except that now we use X_j and $X_j + X_l$ as predictors (plus whatever the remaining predictors are) instead of X_j and X_l. Estimating a model with this modified set of predictors will then allow us to easily test whether θ is statistically significant. This test is exactly equivalent to our original test of whether β_j equals β_l.

Multiple linear restrictions, F test

Exclusion restrictions

Often we would like to test hypotheses about more than one linear restriction. In these cases, we cannot use a clever re-writing of the original model to reduce it to a simple t test. For example, we might be interested in testing whether more than two coefficients, say are *jointly* statistically significant.

Recall our population model. We will call it the *unrestricted* model.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U.$$

Suppose we are interested in testing the following hypothesis:

$$H_0 : \beta_{k-q+1} = 0, \ldots, \beta_k = 0.$$

The hypothesis we make imposes q *exclusion restrictions* on the coefficients. We are interested in whether the last q coefficients are all zero. Taking the last coefficients is without the loss of generality, of course, because the order of variables and coefficients is arbitrary. The alternative hypothesis is that at least one of these coefficients is non-zero.

One possibility would be to conduct q corresponding t tests of significance. However, recall that a single t statistic assumes that only one restriction is placed and that there are no further restrictions on other coefficients. We need to find a way to test for all the restrictions together instead of one by one.

If our null hypothesis is true and the last q coefficients are in fact zero, we can write a *restricted* model as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_{k-q} X_{k-q} + U.$$

The idea behind our joint test is based on the residual sums of squares (SSR) from the unrestricted and restricted models. Let's call them SSR_{ur} and SSR_r, respectively. We know that removing predictors from a model mechanically increases the SSR. Thus the SSR from the restricted model will always be greater than the SSR from the unrestricted model. The question is whether this increase in the SSR is *big enough*.

We can answer this question formally by constructing the following F statistic

$$F \equiv \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n-k-1)}.$$

Since SSR_r is always no smaller than SSR_{ur}, the resulting test statistic will always be non-negative. In essence, the F statistic is measuring the relative increase in the SSR when going from the unrestricted to restricted models. However, we scale the numerator and the denominator by the appropriate degrees of freedom, which is q, the number of imposed restriction, for the numerator and $n-k-1$ for the denominator. You may recall that the denominator is also an unbiased estimator of the variance of the error term.

To make statistical inference, we need to know the sampling distribution of our computed F statistic. The following result holds.

ⓘ Sampling distribution of the F statistic

Under the CLM assumptions and if H_0 holds, the F statistic follows an F distribution with $(q, n-k-1)$ degrees of freedom:

$$F \sim F_{q,n-k-1}$$

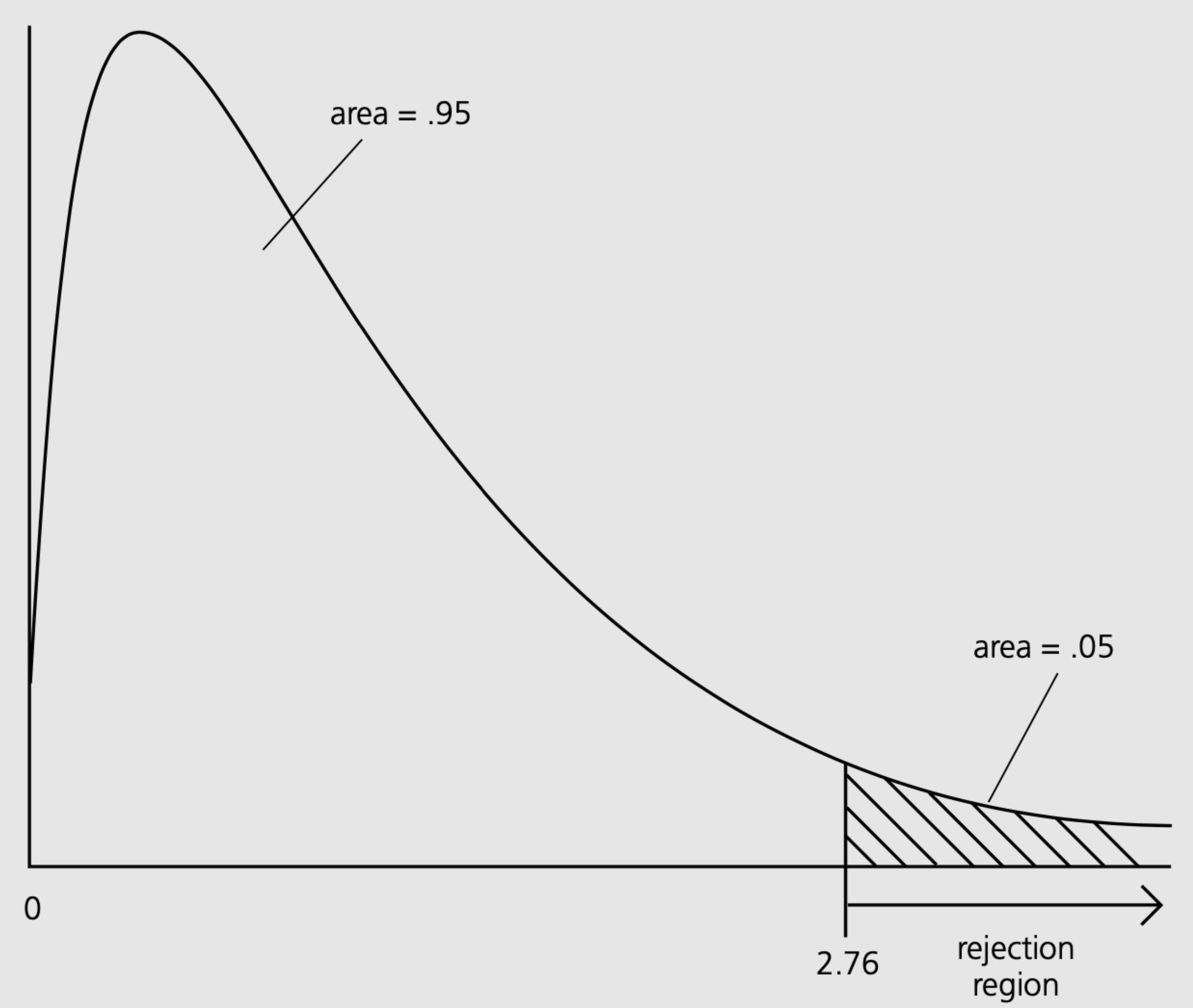
Since the F statistic is non-negative, to reject the null hypothesis, the value of the statistic has to be *large enough*. The general decision rule is

$$\begin{cases} F > c, & \text{reject } \mathrm{H}_0 \\ F \leqslant c, & \text{do not reject } \mathrm{H}_0. \end{cases}$$

The critical value c will be the $1-\alpha$ percentile of the F distribution with $(q, n-k-1)$ df, where α is a chosen significance level.

For example, for a 5% significance level, 3 restrictions and 60 df in the unrestricted model, the critical value is the 95th percentile of the F distribution with (3,60) df (= 2.76). We will reject the null hypothesis wherever the F statistic is above 2.76 and we will not reject the null wherever the F statistic is below 2.76.

The 5% critical value and rejection region in an $F_{3,60}$ distribution.



If we reject the null hypothesis, we would say that $\beta_{k-q+1}, \ldots, \beta_k$ are *jointly statistically significant*. If we do not reject the null, we say that the coefficients are *jointly insignificant*. However, we cannot tell which coefficient exactly is significant and which one is not.

Testing for joint significance can be useful when we are working with categorical predictors. Since we create a bunch of indicator variables for those predictors, we can test whether the variable on the whole is statistically significant by conducting a joint test on the coefficients of those indicator variables.

Relationship between F and t statistics

While we motivated the F statistic as a way to test for multiple linear restrictions, nothing prevents us from setting q to 1 and testing a single restriction. It turns out that the results of an F test in this case will be exactly the same as the result of a corresponding t test. In fact, one can show that the F statistic in this case of a single restriction is exactly the square of the t statistic. In other words, t^2_{n-k-1} has an $F_{1,n-k-1}$ distribution. However, the t statistic is more flexible in this case because we can test a one-sided alternative, too.

R-squared form of F statistic

The F statistic can be re-written using R-squared from the unrestricted and restricted models instead of the SSR. Recall that we can write $\mathrm{SSR} = \mathrm{SST}(1 - R^2)$. Plugging this expression into the formula for the F statistic and doing the substitution for the unrestricted and restricted models, we get

$$\begin{aligned} F &\equiv \frac{(\mathrm{SSR}_r - \mathrm{SSR}_{ur})/q}{\mathrm{SSR}_{ur}/(n-k-1)} \\ &= \frac{(\mathrm{SST}(1-R_r^2) - \mathrm{SST}(1-R_{ur}^2))/q}{\mathrm{SST}(1-R_{ur}^2)/(n-k-1)} \\ &= \frac{(R_{ur}^2 - R_r^2)/q}{(1-R_{ur}^2)/(n-k-1)}. \end{aligned}$$

p-value of F test

Instead of using critical values to conduct an F test, we can also use p-values. The p-value of an F test is defined as

$$p \equiv \mathbb{P}(\mathcal{F} > F),$$

where $\mathcal{F}$ is a random variable following an F distribution with $(q, n-k-1)\ df$.

The interpretation of the p-value for an F test is identical to its interpretation for a t test. It is the probability of observing the value of an F statistic at least as large as we did, given that the null hypothesis is true. To reject the null, the p-value has to be *low enough*.

Our decision rule for a chosen α significance level will be

$$\begin{cases} p \leqslant \alpha, \ \text{reject } H_0, \\ p > \alpha, \ \text{do not reject } H_0. \end{cases}$$

For example, suppose we calculate the p-value of our F test to be 0.024. Then we would reject the null at a 5% significance level and would not reject it at a 1% level.

Overall significance

We can the F test for a hypothesis about the overall significance of our regression. The null hypothesis takes the form

$$H_0 : \beta_1 = 0, \ldots, \beta_k = 0,$$

i.e., all of the coefficients are zero, except for the constant term. If the null hypothesis is true than none of the predictors in our model has any effect on the outcome. Another way to think about this is that under the null, the conditional expectation of the outcome is the same as its *unconditional* expectation:

$$\mathbb{E}[Y \mid X_1, \ldots, X_K] = \mathbb{E}[Y].$$

The alternative hypothesis is here is that at least one coefficient is non-zero.

Imposing all of these k restrictions results in the following restricted model

$$Y = \beta_0 + U,$$

which includes only the constant term. In this model, the R-squared is trivially zero, since there are no predictors. Therefore, we can write the F statistic in the R-squared form as

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)},$$

where R^2 is the R-squared from the unrestricted model.

The F test of the overall significance is a common part of a regression output.

General linear restrictions

While testing for exclusion restrictions, i.e., that a bunch of coefficients is jointly zero, is the most common use of an F test, the test can be applied to more general restrictions. These restrictions are often suggested by a theoretical model. For a concrete example, suppose we have a regression model with three predictors

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U.$$

We hypothesize that all of the regression coefficients are zero except for β_1, which we assume equals b:

$$H_0 : \beta_1 = b, \beta_2 = 0, \beta_3 = 0.$$

To test this hypothesis, we first write down the restricted model

$$Y = \beta_0 + bX_1 + U.$$

We can re-write it as

$$Y - bX_1 = \beta_0 + U.$$

The right-hand side of this expression is the same as in the restricted model in which all the coefficients are zero. The left-hand side, however, can be thought of as a transformed outcome. Instead of using Y, we create a new outcome variable defined as $Y - bX_1$. Then we use this new outcome variable in the restricted model.

Note

We cannot use the R-squared form of an F statistic in this case because the outcome (and hence SST) now differs between the two models. We have to use the original formula with SSR instead.