

## Inferences in Multiple Regression Model

- In a classical multiple regression model, we have  $k$  regressors and the inference on individual regressor or a combination of regressors can be done in usual way through  $t$ -test. However, it might be also required to test the joint significance of more than one regressors. One example is;

$$H_0; \beta_2 = 0 \wedge \beta_3 = 0 \wedge \beta_5 = 0 \quad H_0; \beta_2 \neq 0 \vee \beta_3 \neq 0 \vee \beta_5 \neq 0$$

It is obvious that it is impossible to test the hypothesis given above using  $t$ -test. How can we solve the problem? The answer lies in  $F$ -test - It will be shown that the  $t$ -test procedure can be included in this more general test procedure.

- Consider the set of linear hypotheses on the elements of  $\beta$  denoted as

$$\underset{(q \times k)}{R} \underset{(k \times 1)}{\beta} = \underset{(q \times 1)}{r}$$

where  $q$  is the number of linearly independent hypotheses with  $q \leq k$ .

*Examples;*

(i)  $\beta_j = 0 \Rightarrow R = [ 0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0 ]$  and  $r = 0$ ; test of validity of one regressor

(ii)  $\beta_1 + 2\beta_2 - \beta_3 = 1 \Rightarrow R = [ 1 \ 2 \ -1 \ 0 \ \cdots \ 0 ]$  and  $r = 1$ ; test of validity of linear combination of regressors

(iii)  $\beta_2 = 0, \beta_3 = 1, \beta_4 + \beta_5 = 0 \Rightarrow R \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & \cdots & 0 \end{bmatrix}$  and  $r = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ ; joint test

- Now, consider the testing procedure;

$$H_0; R\beta = r \quad H_A; R\beta \neq r$$

The starting point is the distribution of least squares estimator;

$$\hat{\beta} \sim N \left( \beta, \sigma^2 (X'X)^{-1} \right)$$

Then, we have

$$R\hat{\beta} \sim N \left( R\beta, \sigma^2 R (X'X)^{-1} R' \right)$$

It is easy to prove the claim, try it. Therefore,

$$R\hat{\beta} - R\beta \sim N \left( 0, \sigma^2 R (X'X)^{-1} R' \right)$$

Under the null hypothesis,  $R\beta = r$ ;

$$R\hat{\beta} - r \sim N \left( 0, \sigma^2 R (X'X)^{-1} R' \right)$$

Recall the fact that

$$\left( R\hat{\beta} - r \right)' \left[ \sigma^2 R (X'X)^{-1} R' \right]^{-1} \left( R\hat{\beta} - r \right) \sim \chi^2 (q)$$

since  $\rho \left( R (X'X)^{-1} R' \right) = q$ . However, the formula is not readily usable since it is involved in a unknown quantity,  $\sigma^2$ . How can we get rid of  $\sigma^2$  so that we have a computable test statistic? Remember that

$$\frac{(N - k) s^2}{\sigma^2} = \frac{e'e}{\sigma^2} \sim \chi^2 (N - k)$$

Then, consider the following quantity;

$$\begin{aligned}
 & \frac{\left[ (R\hat{\beta} - r)' \left[ \sigma^2 R (X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r) \right] / q}{\left[ \frac{e'e}{\sigma^2} \right] / (N - k)} \\
 &= \frac{\left[ (R\hat{\beta} - r)' \left[ R (X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r) \right] / q}{e'e / (N - k)} \\
 &= \frac{\left[ (R\hat{\beta} - r)' \left[ R (X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r) \right] / q}{s^2} \\
 &= \frac{(R\hat{\beta} - r)' \left[ s^2 R (X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r)}{q}
 \end{aligned}$$

The quantity is a ratio of two  $\chi^2$  random variables with each divided by degrees of freedom. It is exactly the definition of  $F$  distribution provided that the two random variables are independent. You can show they are actually independent as in lecture note. The degrees of freedom for  $F$  distribution is given by  $(q, N - k)$  –  $F$  distribution requires two kinds of degrees of freedom. In sum,

- (i) Compute  $F = \frac{(R\hat{\beta} - r)' \left[ s^2 R (X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r)}{q} \sim F(q, N - k)$
- (ii) Find the critical value for a given significance level from the table.
- (iii) Reject the null hypothesis if  $F$  is larger than the critical value.

One thing to remember in the procedure is that the critical value is determined by

$$F_\alpha(q, N - k)$$

where  $\alpha$  is the given significance level. For example, if  $\alpha = 0.95$  (95%), You have to find a value  $x$  such that

$$P[F(q, N - k) \leq x] = 0.95$$

This is different from  $t$ -test. The critical value in  $t$ -test is defined as, in case  $\alpha = 0.95$  (95%),

$$P[t(N - k) \leq x] = 0.975$$

Why do we have difference? You can find the answer from the formula for  $F$  statistic. The first thing to notice is that  $F$  statistic is always positive since it is the ratio of two positive quantities. See the first line of the definition and note that  $(X'X)^{-1}$  is positive definite. In  $t$ -test, we had both positive and negative test statistic. Using symmetry of  $t$  distribution, we took the absolute value and paid our attention only to positive part. What is going on with  $F$  test is that the formula for  $F$  test itself has already taken care of the procedure so that we have only positive test statistics. Negative part is squared and the probability of the negative part is added to positive part of the probability distribution.

- Now, we want to derive a special form of  $F$  statistic when we have a specific form of null hypothesis;
  - (a)

$$H_0; \beta_j = 0 \quad H_A; \beta_j \neq 0$$

Then,  $R = [ 0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0 ]$  and  $r = 0$ . Therefore,

$$(R\hat{\beta} - r) = [ 0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0 ] \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_{j-1} \\ \hat{\beta}_j \\ \hat{\beta}_{j+1} \\ \vdots \\ \hat{\beta}_k \end{bmatrix} - 0 = \hat{\beta}_j$$

And

$$s^2 R (X'X)^{-1} R' = s^2 \begin{bmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{bmatrix} (X'X)^{-1} \begin{bmatrix} 0 \\ \dots \\ 0 \\ 1 \\ 0 \\ \dots \\ 0 \end{bmatrix} = s^2 (X'X)^{-1}_{jj}$$

where  $(X'X)^{-1}_{jj}$  is the  $(j, j)$  element of the matrix  $(X'X)^{-1}$ . What is  $s^2 (X'X)^{-1}_{jj}$ ? It is the estimate of variance for  $\hat{\beta}_j$  denoted as  $\hat{\sigma}_{\beta_j}^2$ . Finally, we have  $q = 1$ . Then, the test statistic becomes;

$$\begin{aligned} f &= \frac{(R\hat{\beta} - r)' [s^2 R (X'X)^{-1} R']^{-1} (R\hat{\beta} - r)}{q} \\ &= \frac{\hat{\beta}_j [\hat{\sigma}_{\beta_j}^2]^{-1} \hat{\beta}_j}{1} = \frac{\hat{\beta}_j^2}{\hat{\sigma}_{\beta_j}^2} = \left[ \frac{\hat{\beta}_j - 0}{\hat{\sigma}_{\beta_j}} \right]^2 \sim F(1, N - k) \end{aligned}$$

What is the quantity inside the bracket? Yes! It is the square of  $t$  statistic for the hypothesis that

$$H_0; \beta_j = 0 \quad H_A; \beta_j \neq 0$$

Our heuristic argument that  $F$  distribution is a kind of squared distribution is verified in this case.

(b) Validity of subset of regression coefficient

Another common form of hypothesis is given by;

$$H_0; \beta_1 = \beta_2 = \dots = \beta_q = 0$$

where  $q < k$ . The indexation of the parameters is arbitrary in that the order of regressors does not matter in the regression. We have

$$\begin{aligned} (R\hat{\beta} - r) &= \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_q \\ \hat{\beta}_{q+1} \\ \dots \\ \hat{\beta}_k \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ 0 \\ 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_{q-1} \\ \hat{\beta}_q \end{bmatrix} \end{aligned}$$

And,

$$R (X'X)^{-1} R' = [(X'X)^{-1}]_{qq}$$

where  $[(X'X)^{-1}]_{qq}$  is the upper left  $(q \times q)$  block of  $(X'X)^{-1}$ . What is the expression for  $[(X'X)^{-1}]_{qq}$ ? Form the formula for inverse of a partitioned matrix;

$$\begin{aligned} X'X &= \begin{bmatrix} X'_q X_q & X'_q X_{k-q} \\ X'_{k-q} X_q & X'_{k-q} X_{k-q} \end{bmatrix} \\ (X'X)^{-1} &= \begin{bmatrix} \left( X'_q X_q - X'_q X_{k-q} (X'_{k-q} X_{k-q})^{-1} X'_{k-q} X_q \right)^{-1} & \dots \\ \dots & \dots \end{bmatrix} \end{aligned}$$

where  $X_q$  is  $(N \times q)$  matrix corresponding to  $\beta_1, \beta_2, \dots, \beta_q$  and  $X_{k-q}$  is  $(N \times (k-q))$  matrix corresponding to  $\beta_{q+1}, \beta_{q+2}, \dots, \beta_k$ . Then,

$$\begin{aligned} [(X'X)^{-1}]_{qq} &= \left( X'_q X_q - X'_q X_{k-q} (X'_{k-q} X_{k-q})^{-1} X'_{k-q} X_q \right)^{-1} \\ &= \left[ X'_q \left[ I - X_{k-q} (X'_{k-q} X_{k-q})^{-1} X'_{k-q} \right] X_q \right]^{-1} \\ &= [X'_q M_{k-q} X_q]^{-1} \end{aligned}$$

Therefore, the test statistic is given by;

$$\begin{aligned} f &= \frac{\left[ (R\hat{\beta} - r)' \left[ R(X'X)^{-1} R' \right]^{-1} (R\hat{\beta} - r) \right] / q}{e'e / (N - k)} \\ &= \frac{\left\{ \left[ \hat{\beta}_1 \quad \hat{\beta}_2 \quad \dots \quad \hat{\beta}_{q-1} \quad \hat{\beta}_q \right] [X'_q M_{k-q} X_q] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_{q-1} \\ \hat{\beta}_q \end{bmatrix} \right\} / q}{e'e / (N - k)} \\ &= \frac{\hat{\beta}^{q'} [X'_q M_{k-q} X_q] \hat{\beta}^q / q}{e'e / (N - k)} \sim F(q, N - k) \end{aligned}$$

where  $\hat{\beta}^q = \left[ \hat{\beta}_1 \quad \hat{\beta}_2 \quad \dots \quad \hat{\beta}_{q-1} \quad \hat{\beta}_q \right]$ . Now, consider the original regression in partitioned form;

$$y = X_q \hat{\beta}^q + X_{k-q} \hat{\beta}^{k-q} + e$$

where  $\hat{\beta}^{k-q} = \left[ \hat{\beta}_{q+1} \quad \hat{\beta}_{q+2} \quad \dots \quad \hat{\beta}_{k-1} \quad \hat{\beta}_k \right]'$  and  $e$  is again the residual vector. Multiply both sides by  $M_{k-q}$ , we have

$$\begin{aligned} M_{k-q} y &= M_{k-q} X_q \hat{\beta}^q + M_{k-q} X_{k-q} \hat{\beta}^{k-q} + M_{k-q} e \\ &= M_{k-q} X_q \hat{\beta}^q + e \end{aligned}$$

since

$$\begin{aligned} M_{k-q} X_{k-q} &= \left( I - X_{k-q} (X'_{k-q} X_{k-q})^{-1} X'_{k-q} \right) X_{k-q} = 0 \\ M_{k-q} e &= \left( I - X_{k-q} (X'_{k-q} X_{k-q})^{-1} X'_{k-q} \right) e \\ &= e - X_{k-q} (X'_{k-q} X_{k-q})^{-1} X'_{k-q} e = e \end{aligned}$$

Note that  $X'e = \begin{bmatrix} X'_q \\ X'_{k-q} \end{bmatrix} e = \mathbf{0}$  implies that  $X'_{k-q} e = \mathbf{0}$ . Then,

$$\begin{aligned} (M_{k-q} y)' (M_{k-q} y) &= y' M_{k-q} y = \left( M_{k-q} X_q \hat{\beta}^q + e \right)' \left( M_{k-q} X_q \hat{\beta}^q + e \right) \\ &= \hat{\beta}^{q'} X'_q M'_{k-q} M_{k-q} X_q \hat{\beta}^q + \hat{\beta}^{q'} X'_q M'_{k-q} e + e' M_{k-q} X_q \hat{\beta}^q + e'e \\ &= \hat{\beta}^{q'} X'_q M'_{k-q} M_{k-q} X_q \hat{\beta}^q + \hat{\beta}^{q'} X'_q e + e' X_q \hat{\beta}^q + e'e \\ &= \hat{\beta}^{q'} X'_q M_{k-q} X_q \hat{\beta}^q + e'e \end{aligned}$$

since  $M_q$  is symmetric and idempotent and  $M'_{k-q} e = e$ . Then,

$$\begin{aligned} \hat{\beta}^{q'} X'_q M_{k-q} X_q \hat{\beta}^q &= y' M_{k-q} y - e'e \\ &= e^{*'} e^* - e'e \end{aligned}$$

where  $e^*$  is the residual vector from the regression of  $y$  on  $X_{k-q}$  since

$$y' M_{k-q} y = (M_{k-q} y)' (M_{k-q} y) = e^{*'} e^*$$

Note that  $M$  is also called the residual generating matrix. Back to the  $F$  statistic;

$$F = \frac{(e^{*'} e^* - e' e) / q}{e' e / (N - k)} \sim F(q, N - k)$$

The test procedure is involved in running two regressions;

(i) regress  $y$  on  $X$  and get the residual sum of squares,  $e' e$

(ii) regress  $y$  on  $X_{k-q}$ -variables not restricted under the null hypothesis- and get the residual sum of squares,  $e^{*'} e^*$

(iii) Form the  $F$  statistic and compare it with the critical value

(c) Validity of the regression

$$H_0; \beta_2 = \beta_3 = \dots = \beta_k = 0$$

This is a special case of the above presentation. The null hypothesis claims that every slope coefficient is jointly not different from zero. Now, let's apply the result given above;

$$f = \frac{\widehat{\beta}^{*'} [X_*' A X_*] \widehat{\beta}^* / (k - 1)}{e' e / (N - k)} \sim F(k - 1, N - k)$$

where  $\widehat{\beta}^* = \left[ \widehat{\beta}_2 \quad \widehat{\beta}_3 \quad \dots \quad \widehat{\beta}_k \right]'$ ,  $X_* = \left[ X_2 \quad X_3 \quad \dots \quad X_k \right]$  and

$$A = I - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} \mathbf{1}'$$

Can you figure out why it is the case? Basically,  $M_{k-q}$  in this case is nothing but the projection onto space orthogonal to space spanned by the column corresponding to the constant term, which is given by  $A$ . Then, what is the numerator in the  $F$  statistic? Yes, it is the explained sum of squares from the regression.

$$y = \widehat{\beta}_1 \mathbf{1} + X_* \widehat{\beta}^* + e$$

Hence,

$$Ay = AX_* \widehat{\beta}^* + Ae = AX_* \widehat{\beta}^* + e$$

since the mean of  $e$  is always zero - remember that  $A$  transforms a variable in mean deviation form. Then,

$$\begin{aligned} (Ay)' (Ay) &= y' Ay = \left( AX_* \widehat{\beta}^* + e \right)' \left( AX_* \widehat{\beta}^* + e \right) \\ &= \widehat{\beta}^{*'} X_*' A' AX_* \widehat{\beta}^* + \widehat{\beta}^{*'} X_*' A' e + e' AX_* \widehat{\beta}^* + e' e \\ &= \widehat{\beta}^{*'} X_*' AX_* \widehat{\beta}^* + e' e \end{aligned}$$

since  $A$  is symmetric and idempotent and  $A'e = 0$ .  $y' Ay = \sum_{i=1}^N (y_i - \bar{y})^2$  is the definition of the total sum of squares and  $e' e$  is the definition of the residual sum of squares. Hence,

$$\begin{aligned} f &= \frac{ESS / (k - 1)}{RSS / (N - k)} = \frac{(N - k) ESS}{(k - 1) RSS} = \frac{(N - k) \frac{ESS}{TSS}}{(k - 1) \frac{RSS}{TSS}} \\ &= \frac{(N - k) \frac{ESS}{TSS}}{(k - 1) \frac{TSS - ESS}{TSS}} = \frac{(N - k) R^2}{(k - 1) 1 - R^2} \end{aligned}$$

## Partitioned Regression

- Consider the classical multiple regression model

$$y = X\beta + \varepsilon$$

$$E(\varepsilon) = 0, E(\varepsilon\varepsilon') = \sigma^2 I$$

If we partition  $X$  and  $\beta$  conformably to get

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

where  $X_1$  is  $(N \times l)$  matrix and  $X_2$  is  $(N \times (k - l))$  matrix with  $l < k$ .

- Define

$$M_1 = I - X_1(X_1'X_1)^{-1}X_1'$$

and consider the regression

$$M_1y = M_1X_2\beta_2 + \varepsilon$$

- We claim that the least squares estimator for  $\beta_2$  in two regressions are identical. For the second regression, the least square estimator is given by

$$\begin{aligned}\widehat{\beta}_2 &= [(M_1X_2)'(M_1X_2)]^{-1} [(M_1X_2)'(M_1y)] \\ &= (X_2'M_1X_2)^{-1} (X_2'M_1y)\end{aligned}$$

since  $M_1$  is symmetric and idempotent. For the first regression,

$$\widetilde{\beta} = \begin{bmatrix} \widetilde{\beta}_1 \\ \widetilde{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}$$

Therefore, from the formula for the inverse of a partitioned matrix;

$$\begin{aligned}\widetilde{\beta}_2 &= - \left( X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2 \right)^{-1} X_2'X_1(X_1'X_1)^{-1}X_1'y \\ &\quad + \left( X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2 \right)^{-1} X_2'y \\ &= \left( X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2 \right)^{-1} \left( X_2'y - X_2'X_1(X_1'X_1)^{-1}X_1'y \right) \\ &= \left[ X_2' \left( I - X_1(X_1'X_1)^{-1}X_1' \right) X_2 \right]^{-1} \left[ X_2' \left( I - X_1(X_1'X_1)^{-1}X_1' \right) y \right] \\ &= [X_2'M_1X_2]^{-1} [X_2'M_1y] = \widehat{\beta}_2\end{aligned}$$

What about the variance matrix, then? For the first regression,

$$\text{Var}(\widehat{\beta}_2) = \sigma^2 (X_2'M_1X_2)^{-1}$$

For the second regression, it is given by

$$\text{Var}(\widetilde{\beta}_2) = \sigma^2 \left[ (X'X)^{-1} \right]_{22}$$

where  $\left[ (X'X)^{-1} \right]_{22}$  is the lower right block of  $(X'X)^{-1}$  which is given by

$$\left( X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2 \right)^{-1} = (X_2'M_1X_2)^{-1}$$

We have identical results from two different regressions.

- Then, what does the second regression really mean? The dependent variable is  $M_1y$ . This is the residual vector from the regression

$$y = X_1\delta + \varepsilon$$

since

$$e_y = y - X_1\hat{\delta} = y - X_1(X_1'X_1)^{-1}X_1'y = [I - X_1(X_1'X_1)^{-1}X_1']y = M_1y$$

The independent variables are  $M_1X_2$  whose dimension is  $(N \times (k-l))$ . The  $h^{th}$  column of  $M_1X_2$ ,  $h = 1, 2, \dots, (k-l)$ , is the residual vector from the regression

$$X^{l+j} = X_1\gamma + \varepsilon$$

since

$$e_{l+j} = X^{l+j} - X_1(X_1'X_1)^{-1}X_1'X^{l+j} = M_1X^{l+j}$$

where  $X^{l+j}$  is the  $(l+j)^{th}$  column of  $X$ . Therefore, the regression

$$M_1y = M_1\beta_2 + \varepsilon$$

is equivalent to the following procedure;

(i) regress  $y$  on  $X_1$  and get the residual vector  $e_y$

(ii) regress *each* column of  $X_2$  on  $X_1$  and get the residual vectors to form the matrix of the residual vectors - this stage is involved in  $(k-l)$  separate regressions

(iii) regress  $e_y$  on the residuals calculated in (ii). The estimate we get from the exercise is the least squares estimate of  $\beta_2$ .

- It is a good exercise to verify that

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon \tag{1}$$

and

$$M_2y = M_2X_1\beta_1 + \varepsilon \tag{2}$$

gives us identical least squares estimator of  $\beta_1$ .

- Multiplying both sides of (1) by  $M_2$  to obtain (2) –remember  $M_2X_2 = 0$ – has quite an interesting geometric implication.