

The F-distribution

- We have seen that a test for all coefficients being significant –aside from the constant is:

$$F = \frac{ESS / k}{RSS / (n - k - 1)} \sim F(k, n - k - 1)$$

- We may therefore compare the resulting F statistic with the 95% or 99% etc. critical value of the appropriate F distribution.
- If the F-statistic exceeds this critical value, we may reject the null hypothesis that $\beta_1 = \beta_2 = \dots = \beta_k = 0$, and conclude that the regression as a whole is significant, that is that the variables X_1, \dots, X_k are jointly significant.
- It is possible for a set of variables to be jointly significant, even if they are all individually insignificant.
- This is particularly likely if there is a very close relationship between the X-variables (multicollinearity);
- Can generalise to test hypothesis that *some of the coefficients* are jointly insignificant.
- Consider $Y_i = \alpha + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + u_i$,
- test the hypothesis that $\beta_1 = \dots = \beta_f = 0$, with $f \leq k$
- – that is, we are testing the hypothesis that X_1, \dots, X_f are jointly insignificant.
- Now the *difference* between the RSS obtained with the original model, and the RSS obtained with the restricted model (with the first f variables removed), divided by σ^2 , has the χ^2 distribution with f degrees of freedom, under the null hypothesis that $\beta_1 = \dots = \beta_f = 0$.

- Let URSS be the RSS obtained with the unrestricted model, and let RRSS be the RSS from the restricted model, that is with $\beta_1 = \dots = \beta_f = 0$. Then under the null hypothesis,

$$F = \frac{(RRSS - URSS) / f}{URSS / (n - k - 1)} \sim F(f, n - k - 1).$$

- (Note that RRSS will be greater than URSS – since we have fewer variables in the restricted model, the residuals will be higher.)
- We may therefore compare this F statistic with the appropriate critical value of the F distribution. If the statistic exceeds the 95% (or 99% or whatever) critical value, then we may reject the null hypothesis, and conclude that the deleted variables, X_1, \dots, X_l , are in fact significant.
- More generally still, the same statistic can be used when testing any 1 linear restrictions on the model such as $\beta_1 = 1$, or $\beta_1 + \beta_2 = 0$, etc.
- URSS is the RRSS from the regression obtained by imposing these restrictions, that is by calculating the least squares estimators based on the requirement that these are true.
- For example, suppose our model is

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i,$$

and we are testing the hypothesis that $\beta_1 = 1$ and $\beta_2 + \beta_3 = 0$,

then the restricted model involves finding estimators for the model

$$Y_i = \alpha + X_{1i} + \beta_2 X_{2i} - \beta_2 X_{3i} + u_i.$$

Hence we choose α and β_2 to minimise the sum of square error.