

MOOC Econometrics: Test Exercise #3

Greg Faletto

Questions

This test exercise is of a theoretic nature. The exercise is based on Exercise 5.2c of 'Econometric Methods with Applications in Business and Economics.' The question of interest is how the decision whether or not to include a group of variables differs based on AIC from that based on the F-test. We will stepwise show that for large samples selection based on AIC corresponds to an F-test with a critical value of approximately 2.

Problem (a)

Consider the usual linear model, where $y = X\beta + \epsilon$. We now compare two regressions, which differ in how many variables are included in the matrix X . In the full (unrestricted) model p_1 regressors are included. In the restricted model only a subset of $p_0 < p_1$ regressors are included. Show that the smallest model is preferred according to the AIC if

$$\frac{s_0^2}{s_1^2} < e^{\frac{2}{n}(p_1 - p_0)}.$$

Solution

AIC:

$$\log(s^2) + \frac{2k}{n}$$

where k is the number of regressors included in the model.

Small model:

$$AIC_0 = \log(s_0^2) + \frac{2p_0}{n}$$

Big model:

$$AIC_1 = \log(s_1^2) + \frac{2p_1}{n}$$

AIC: Choose small model if

$$\begin{aligned} AIC_0 &< AIC_1 \\ \log(s_0^2) + \frac{2p_0}{n} &< \log(s_1^2) + \frac{2p_1}{n} \\ \log(s_0^2) - \log(s_1^2) &< \frac{2p_1}{n} - \frac{2p_0}{n} \\ \log\left(\frac{s_0^2}{s_1^2}\right) &< \frac{2}{n}(p_1 - p_0) \end{aligned}$$

$$\frac{s_0^2}{s_1^2} < e^{\frac{2}{n}(p_1 - p_0)}$$

Problem (b)

Argue that for very large values of n the inequality of (a) is equal to the condition

$$\frac{s_0^2 - s_1^2}{s_1^2} < \frac{2}{n}(p_1 - p_0)$$

Use that $e^x \approx 1 + x$ for small values of x .

Solution

If n is very large,

$$\frac{2}{n}(p_1 - p_0)$$

is small. Therefore, since

$$e^x \approx 1 + x$$

we can approximate that

$$e^{\frac{2}{n}(p_1 - p_0)} \approx 1 + \frac{2}{n}(p_1 - p_0)$$

(if n is very large.) Substituting this expression into the right side of the result from part (a) yields

$$\frac{s_0^2}{s_1^2} < 1 + \frac{2}{n}(p_1 - p_0)$$

$$\frac{s_0^2}{s_1^2} - 1 < \frac{2}{n}(p_1 - p_0)$$

$$\frac{s_0^2}{s_1^2} - \frac{s_1^2}{s_1^2} < \frac{2}{n}(p_1 - p_0)$$

$$\frac{s_0^2 - s_1^2}{s_1^2} < \frac{2}{n}(p_1 - p_0)$$

(for n very large.)

Problem (c)

Show that for very large values of n the condition in (b) is approximately equal to

$$\frac{e_R' e_R - e_U' e_U}{e_U' e_U} < \frac{2}{n}(p_1 - p_0)$$

where e_R is the vector of residuals for the restricted model with p_0 parameters and e_U the vector of residuals for the full unrestricted model with p_1 parameters.

Solution

$$\begin{aligned} s^2 &= \frac{1}{n-k} \sum_{i=1}^n e_i^2 = \frac{1}{n-k} e' e \\ \Rightarrow s_0^2 &= \frac{1}{n-p_0} e_R' e_R, s_1^2 = \frac{1}{n-p_1} e_U' e_U \end{aligned}$$

Substituting these expressions into the result from part (b) yields

$$\frac{\frac{1}{n-p_0} e_R' e_R - \frac{1}{n-p_1} e_U' e_U}{\frac{1}{n-p_1} e_U' e_U} < \frac{2}{n}(p_1 - p_0)$$

For large values of n , $n - p_0 \approx n - p_1 \approx n$. This yields

$$\frac{\frac{1}{n} e_R' e_R - \frac{1}{n} e_U' e_U}{\frac{1}{n} e_U' e_U} < \frac{2}{n}(p_1 - p_0)$$

Finally, multiplying the left side by $\frac{n}{n}$ yields

$$\frac{e_R' e_R - e_U' e_U}{e_U' e_U} < \frac{2}{n}(p_1 - p_0)$$

Problem (d)

Finally, show that the inequality from (c) is approximately equivalent to an F-test with critical value 2, for large sample sizes.

Solution

F-test (formula from Lecture 2.4.2 slides):

$$F = \frac{(e_R' e_R - e_U' e_U)/g}{e_U' e_U/(n - k)}$$

where k is the number of explanatory factors in the unrestricted model, and g is the number of explanatory factors removed from the unrestricted model to create the restricted model.

Under this test, we believe there is significant evidence to suggest that $\beta \neq 0$ (so the unrestricted model is preferred) if $F > F_{critical}$. Therefore a larger model is preferred if $F > F_{critical}$, and we stay with (prefer) a smaller model if $F < F_{critical}$.

Let $F_{critical} = 2$. Then a smaller model is preferred if $F < 2$:

$$\frac{(e_R' e_R - e_U' e_U)/g}{e_U' e_U/(n - k)} < 2$$

In this case, with p_1 factors in the unrestricted model and p_0 in the restricted model, we get

$$\begin{aligned} \frac{(e_R' e_R - e_U' e_U)/(p_1 - p_0)}{e_U' e_U/(n - p_1)} &< 2 \\ \frac{(e_R' e_R - e_U' e_U)}{e_U' e_U} &< \frac{2(p_1 - p_0)}{n - p_1} \end{aligned}$$

If n is very large, $n - p_1 \approx n$. Substituting this in yields

$$\frac{(e_R' e_R - e_U' e_U)}{e_U' e_U} < \frac{2(p_1 - p_0)}{n}$$

which is the same result that was obtained in part (c). Our condition for preferring a restricted model when doing an F-test with $F_{critical} = 2$ (and when n is very large) is approximately the same as our condition for preferring a restricted model when using the AIC (when n is very large).