Econometrics: Methods and Applications

Erasmus
University
Rotterdam

Peer-graded Assignment: Test Exercise 3

Goals and skills being used:

- Apply the Akaike Information Criterion (AIC) for model selection.
- Examine large-sample behavior of AIC.
- Link AIC to F-test discussed in multiple regression lectures.

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.

(a) Consider the usual linear model, where $y = X\beta + \varepsilon$. We now compare two regressions, which differ in how many variables are included in the matrix X. In the full (unrestricted) model $p1$ 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

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

Answer

- 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

Akaike information criterion
(Smaller values indicate better model fitting.)

$$\mathrm{AIC}(p) = \log\left(s_p^2\right) + \boxed{\frac{2p}{n}}$$

penalty term associated with the number of variables

p is the number of included regressors and s_p^2 is the maximum likelihood estimator of the error variance in the model with *p* regressors. These criteria involve a **penalty term** for the number of parameters, to account for the fact that the model fit always increases (that is, s_p^2 decreases) if more explanatory variables are included (Source: Econometric Methods, p.279).

We know: $p_0 < p_1$

⇨ $p_1 - p_0 > 0$
⇨ $e^{p_1 - p_0} > 1$

We use the rule:

$$(x^a)^b = x^{a \cdot b}$$

and x = e and a = $\frac{2}{n}$ > 0 and b = $p_1 - p_0 > 0$

⇨ $e^{\frac{2}{n}(p_1 - p_0)} > 1$

⇨ $s_0^2 \quad < \quad s_1^2$

Therefore, the AIC of the full (unrestricted) model *p*1 is higher than the AIC of the restricted model only a subset of $p_0 < p_1$.

⇨ **The unrestricted model has higher values in both terms of the AIC equation than the restricted model, because $s_0^2 \quad < s_1^2$ and $p_0 < p_1$.**

$$\mathrm{AIC}(p) = \log\left(s_p^2\right) + \frac{2p}{n}$$

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

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

Answer

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

and $e^x \approx 1 + x$ for small values of x

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

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

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

➔

(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.

Answer

s = $\sqrt{\frac{e'\ e}{(n-k)}}$ standard error of the regression (p.128 textbook)

as $n \rightarrow \propto$ ➔ $\frac{1}{(n-k)} \rightarrow 0$

s = $\sqrt{e'\ e}$

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

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

The F-test can also be written as

$$F = \frac{(e_R' e_R - e'e)/g}{e'e/(n-k)}, \tag{5.1}$$

where e_R and e are the residuals of the restricted and unrestricted model, respectively. The critical value of 1 corresponds to a size of more than 5 per cent — that is, the TMSP criterion used in this way is more liberal in accepting additional regressors.

(Textbook p. 279-280)

For the linear regression model, the information criteria are related to the F-test (5.1). For large enough sample size n, the comparison of AIC values corresponds to an F-test with critical value 2 and SIC corresponds to an F-test with critical value $\log(n)$ (see Exercise 5.2). For instance, the restricted model is preferred above the unrestricted model by AIC, in the sense that $\text{AIC}(k-g) < \text{AIC}(k)$, if the F-test in (5.1) is smaller than 2.