Comment on Lecture 3.5 Slide 7 (Video at 6 minutes and 40 seconds)

The numerical values of AIC and BIC in the lecture were obtained by a software package (EViews version 8). If you compute these values directly from the formulas provided in Lecture 3.2 (Slide 6), then you get different values. Still, the ranking of AIC values remains the same, so that this difference has no effect on the preferred model (with smallest AIC value). The same holds true for BIC.

First two practical conclusions:

(1) You can compute AIC by any package that you like, and you can safely compare the resulting AIC values. In the lecture on Slide 7, the Book-to-market model is preferred over the full model because it has the smallest values of AIC and BIC. This conclusion remains the same if AIC and BIC are computed as on Slide 6 of Lecture 3.2.

(2) **If you correct work of fellow learners, for example in peer reviewing Text Exercise 3, please do not judge the correctness of the solution from the numerical values of AIC and BIC, but only from their ranking. So, if the answer of your fellow learner provides the correct ranking of models (as judged by AIC and BIC), then the answer should be considered as being correct, irrespective of the numerical values obtained for AIC and BIC.**

Second, and only if you are interested in the technical details, we show the cause of the difference between AIC of Lecture 3.2 (Slide 7) and the value computed by EViews (as shown on Slide 6 in Lecture 3.5). To appreciate this explanation, you have to be familiar with maximum likelihood estimation for the normal distribution. If you lack that knowledge, you can safely skip what comes, because it is not at all needed for our MOOC.

The difference can be explained as follows in terms of the likelihood function (L) for n independent random variables $y_i \sim \mathrm{N}(x_i'\beta, \sigma^2)$.

EViews calculates AIC $= (-2\log(\mathrm{L}) + 2k)/n = -(2/n)\log(\mathrm{L}) + 2k/n$, where the final term is equal to that on Slide 7 of Lecture 3.2. We can write the first term log(L) as:

$\log(\mathrm{L}) = -(n/2)\log(2\pi) - (n/2)\log(\sigma^2) - 1/(2\sigma^2)\sum_{i=1}^{n}(y_i - x_i'\beta)^2$.

From the first-order condition for a maximum with respect to σ^2, we get $\widehat{\sigma}^2 = 1/n \sum_{i=1}^{n}(y_i - x_i'\beta)^2$. Substitute this in the expression for log(L) to get:

$\log(\mathrm{L}) = -(n/2)\log(2\pi) - (n/2)\log(\widehat{\sigma}^2) - n/2.$

Because the OLS estimator $s^2 = 1/(n-k)\sum_{i=1}^{n}(y_i - x_i'\beta)^2$ we get $\widehat{\sigma}^2 = ((n-k)/n)s^2$, so

$\log(\mathrm{L}) = -(n/2)\log(2\pi) - n/2 - (n/2)\log((n-k)/n) - (n/2)\log(s^2).$

Substitute this result in the AIC formula of EViews:

$\mathrm{AIC} = -(2/n)\log(\mathrm{L}) + 2k/n = \log(2\pi) + 1 + \log((n-k)/n) + \log(s^2) + 2k/n,$

which is equal to

$\log(s^2) + 2k/n + \log(2\pi) + 1 + \log((n-k)/n).$

As compared to our definition of AIC on Slide 7 of Lecture 3.2, there is an extra term

$\log(2\pi) + 1 + \log((n-k)/n),$

which converges to $\log(2\pi) + 1 \approx 2.838$ if n is large. For large enough sample size, this extra term will therefore not affect the ranking of models, but for small sample size (or many parameters compared to the sample size) the ranking might be affected.

Finally we check the outcomes for the Book-to-market model. The AIC value of the Book-to-market model is -0.486 according to Slide 7 of Lecture 3.5, whereas the formula on Slide 6 of Lecture 3.2 (with $n = 87$, $k = 2$, and $s = 0.1876$) gives $\log(s^2) + 2k/n = -3.3469 + 0.0460 = -3.3009 \approx -3.301$. The difference is $-0.486 - (-3.301) = 2.815$. This difference is indeed equal to $\log(2\pi) + 1 + \log((n-k)/n) = 1.8379 + 1 - 0.0233 = 2.8146 \approx 2.815$.