EC220 classes

November 18, 2008

Antoine Goujard

Why is the usual decomposition $TSS = RSS + ESS$ not valid when we do not include an intercept in the model?

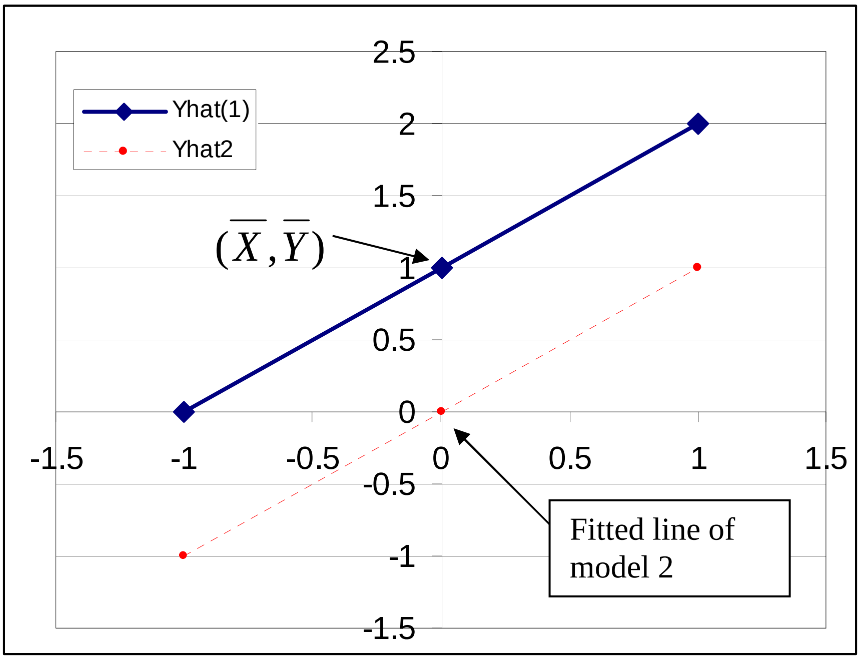
The idea is to look at a simple regression model.
(1) $Y_i = \beta_1 + \beta_2.X_i + u_i$
And to compare it with a model without the intercept β_1.
(2) $Y_i = \alpha_2.X_i + v_i$

Fitting model (1) by OLS, we obtain, $\hat{Y}_{1i}$ the first fitted line, the residuals, $e_{1i} = Y_i - \hat{Y}_{1i}$ and can compute RSS_1 and R_1^2.
Now fitting model (2), we obtain $\hat{Y}_{2i}$ the second fitted line and the residuals, $e_{2i} = Y_i - \hat{Y}_{2i}$.
The R-squarred obtained fitting the first model by OLS can be defined as:
$R_1^2 = \frac{ESS_1}{TSS_1}$ (def. a.) or,
$R_1^2 = 1 - \frac{RSS_1}{TSS_1}$ (def. b.)
From the course and PS3, you should know that in the case of model (1), definitions (a) and (b) are the same. Furthermore, we have the following equalities:
$R_1^2 = r^2_{Y,\hat{Y}_1} = r^2_{Y,X}$
In any case, the equality between definition (a) and definition (b) comes from the fact that **in a model with intercept**: $TSS_1 = ESS_1 + RSS_1$. In the case (1), we know that the equality:
$TSS_1 = ESS_1 + RSS_1$ is valid (see the book, box 1.2 p.62). This is based on the fact that:
$\bar{e}_{1i} = 0$ (which can be proved using the first order conditions of the minimization of the RSS).
By definition of the residuals, this can be rewritten as:
$\bar{\hat{Y}}_{1i} = \bar{Y}$
In other words, the fitted line goes through the average sample observation: $(\bar{X}, \bar{Y})$.
This fact does not hold if we fit model (2) by OLS. In this case, the OLS estimator of α_2 is:
$c_2 = \frac{\sum_{i=1}^{n} X_i.Y_i}{\sum_{i=1}^{n}(X_i)^2}$ (this has been proved in class).
Then the question is: does the second fitted line ($\hat{Y}_{2i}$) go through the average sample observation, $(\bar{X}, \bar{Y})$? The answer is **no in general**.

The best is to look at a simple counter-example. Suppose we have the following data:

Obs	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X}).(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$X_i.Y_i$
1	-1	0	-1	-1	1	1	0
2	0	1	0	0	0	0	0
3	1	2	1	1	1	1	2
$\text{Sum}_{i=1...3}$	0	3	0	0	2	2	2

Then, $c_2 = b_2 = 1$ and $b_1 = \bar{Y} - b_2.\bar{X} = 1 - 0 = 1$. And $c_2.\bar{X} = 0 \neq 1$. Indeed, we have the following graph:



As a consequence, the two definitions of the R^2 are not the same in model (2). From the picture it is clear that $e_{2i} = 1$ for all the three observations and that $RSS_2 = 3$ but,

$ESS_2 = \sum_{i=1}^{n}(\hat{Y}_{2i} - \bar{\hat{Y}}_2)^2 = \sum_{i=1}^{n}(\hat{Y}_{2i} - 0)^2 = \sum_{i=1}^{n}(X_i)^2 = 2$

$TSS_2 = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = 2$

$RSS_2 = 3 \neq TSS_2 - ESS_2 = 2 - 2 = 0$

Moreover, $\bar{e}_2 = 1 \neq 0$.