EC402 classes

November 17, 2010

Antoine Goujard

PS5 Question 1, out of sample prediction

Here is a detailed solution to exercise 1 (see also Johnston & Dinardo, p99).
We have 11 observations (1...11) of the model:
$y_i = \beta_1.x_{1i} + \beta_2.x_{2i} + \varepsilon_i = x_i'.\beta + \varepsilon_i$. with ε_i i.i.d $N(0, \sigma^2)$
And we want to predict y_{12} and $E(y_{12})$.

We need to do three steps. 1/ First we need to find the predicted value at x_{12}, $\hat{y}_{12}$ and the precision of our estimates. 2/ Then we need to evaluate the variance and the distribution of $\hat{y}_{12} - y_{12}$ if we are interested in predicting the value of observation 12. 3/ Finally, we need to evaluate the variance and the distribution of $\hat{y}_{12} - E(y_{12})$ if we are interested in predicting the **expected** value of observation 12.

1/ Using the OLS estimator formula, we get:
$\hat{\beta}_{ols} = \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix}$, thus, $\hat{y}_{12} := \hat{\beta}_1.x_{12} + \hat{\beta}_2.x_{12} = (5-2)/3 = 1$
Moreover we know that: $V(\hat{y}_{12}) = V(x_{12}'.\hat{\beta}_{ols}) = x_{12}'.V(\hat{\beta}_{ols}).x_{12}$.
And that we can estimate $V(\hat{\beta}_{ols})$ by $\hat{V}(\hat{\beta}_{ols}) = \frac{RSS}{n-k}.(X'X)^{-1}$.
So we need to evaluate:
$RSS = \hat{\varepsilon}'\hat{\varepsilon} = y'.y - \hat{y}'.\hat{y}$.
$\hat{y}'.\hat{y} = \hat{\beta}_{ols}'.X'X.\hat{\beta}_{ols} = \frac{1}{9}.\begin{pmatrix} 1 & 1 \end{pmatrix}.\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 6/9 = 2/3$
Thus $RSS = 4/3 - 2/3 = 2/3$ and $s^2 = \frac{RSS}{n-k} = \frac{2}{3}.\frac{1}{11-2} = \frac{2}{27}$
This gives us:

$$\hat{V}(\hat{\beta}_{ols}) = s^2.(X'X)^{-1} = \underbrace{\frac{2}{27}.\frac{1}{3}}_{2/81}.\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

So that: $\hat{V}(\hat{y}_{12}) = \frac{2}{81}.\begin{pmatrix} 5 & -2 \end{pmatrix}.\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.\begin{pmatrix} 5 \\ -2 \end{pmatrix} = \frac{156}{81}$

2/ We want a confidence interval for $y_{12} = \beta_1.x_{12} + \beta_2.x_{12} + \varepsilon_{12}$. So we need to evaluate the variance of $\hat{y}_{12} - y_{12} = x_{12}'.(\hat{\beta}_{ols} - \beta) - \varepsilon_{12}$.

$$V(\hat{y}_{12} - y_{12}) = \underbrace{V(\hat{y}_{12})}_{\simeq \frac{156}{81}} + \underbrace{V(y_{12})}_{\simeq \hat{V}(\varepsilon_{12})} + 2.\underbrace{Cov(\hat{y}_{12}, y_{12})}_{=0}$$

We know that the first term estimate is $\frac{156}{81}$, the second term is estimated by $s^2 = \frac{2}{27}$ and the third term is 0 because we have assumed that ε_i is i.i.d, so y_{12} is independent of y_i for $i \leq 11$

and it follows that y_{12} from $\hat{y}_{12}$. Thus,

$\hat{V}(\hat{y}_{12} - y_{12}) = 2$

Moreover, we have:

$\frac{\hat{y}_{12}-y_{12}}{\sqrt{\hat{V}(\hat{y}_{12}-y_{12})}} \sim t_{n-k}$

Because:

- $\frac{\hat{y}_{12}-y_{12}}{\sqrt{V(\hat{y}_{12}-y_{12})}} \sim N(0,1)$ (here A5Normality A3fixed regressors and A4 scalar covariance matrix are satisfied).

- $(n-k).\frac{\hat{V}(\hat{y}_{12}-y_{12})}{V(\hat{y}_{12}-y_{12})} \sim \chi^2_{n-k}$

(To see this point, note that: $\hat{V}(\hat{y}_{12} - y_{12}) = s^2.(1 + x'_{12}.(X'X)^{-1}.x_{12})$ and: $V(\hat{y}_{12} - y_{12}) = \sigma^2.(1 + x'_{12}.(X'X)^{-1}.x_{12})$. This implies: $(n-k).\frac{\hat{V}}{V} = (n-k).\frac{s^2}{\sigma^2}$ so we can apply the results of the lecture notes p30.)

- The numerator and denominator are independent.

(For the fixed regressors case (or conditionnal on X), the random part of the denominator is s^2 computed with observations 1...11. s^2 is independent of $\hat{\beta}$ (see p30) and thus of $\hat{y}_{12}$. Moreover as s^2 do not depend on observation 12, it is independent of y_{12}).

This gives us, $y_{12} \in [\hat{y}_{12} + / - \sqrt{\hat{V}(\hat{y}_{12} - y_{12})}.t^*_{n-k}(10\%)]$

3/ Now if we want a confidence interval for $E(y_{12}) = \beta_1.x_{12} + \beta_2.x_{12}$, we have to evaluate $\hat{V}(\hat{y}_{12} - E(y_{12}))$. This is $\hat{V}(\hat{y}_{12}) = \frac{156}{81}$ by what we did before. By the previous arguments, we can construct the confidence interval using:

$\frac{\hat{y}_{12}-E(y_{12})}{\sqrt{\hat{V}(\hat{y}_{12})}} \sim t_{n-k}$