Note 3 to Computer class: The regression model and the conditional expectation from a system of variables

Ragnar Nymoen

September 13, 2011

The have been questions about more background to the derivations at the start of the slide set to CC 2.

Here are some details.

We start with the normal distribution, but take care to note note that this is only for convenience: it is then "easy" to follow step-by-step how the regression model can be derived from the statistical system.

At the end, I note some of the extensions to other joint distributions than the normal.

To read more about this: See Appendix 2.A in Biorn (2003), Appendix B8 and B9 in Greene (2012) (which contains a full derivation). Bårdsen and Nymoen (2011) chapter 4.4.6 and appendix 4.A contain an exposition in Norwegian.

The PDF (probability density function) of $\{x_i, y_i\}$ is

$$\begin{aligned} f(x_i, y_i) &= \frac{1}{\sigma_x \sigma_y 2\pi \sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)}\left\{\frac{(x_i-\mu_x)^2}{\sigma_x^2}\right.\right. \\ &\left.\left. -2\rho \frac{(x_i-\mu_x)}{\sigma_x}\frac{(y_i-\mu_y)}{\sigma_y} + \frac{(y_i-\mu_y)^2}{\sigma_y^2}\right\}\right] \end{aligned} \tag{1}$$

where $-\infty < \mu_x, \mu_y < \infty$, $0 < \sigma_x, \sigma_y < \infty$ and

$$\rho = \frac{E[(x_i-\mu_x)(y_i-\mu_y)]}{\sigma_x \sigma_y} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}, \ -1 < \rho < 1.$$

where σ_{xy} is the covariance of x_i and y_i (denoted $Cov[x_i, y_i]$).

$\sigma_x = \sqrt{\sigma_x^2}$ and $\sigma_y = \sqrt{\sigma_y^2}$, where σ_x^2 and σ_y^2 are the marginal variances of x_i and y_i.

μ_x and μ_y are the marginal means of x_i and y_i.

The marginal PDF of x_i is a normal PDF:

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2}} \exp\left[-\frac{(x_i-\mu_x)^2}{2\sigma_x^2}\right] \tag{2}$$

The conditional PDF of y_i is given by

$$f(y_i \mid x_i) = \frac{f(x_i, y_i)}{f(x_i)} \tag{3}$$

and (1) and (2).

$$f(y_i \mid x_i) = \frac{1}{\sigma_x \sigma_y 2\pi\sqrt{1-\rho^2}} \exp[\frac{-1}{2(1-\rho^2)}\{\frac{(x_i-\mu_x)^2}{\sigma_x^2}$$
$$-2\rho\frac{(x_i-\mu_x)}{\sigma_x}\frac{(y_i-\mu_y)}{\sigma_y} + \frac{(y_i-\mu_y)^2}{\sigma_y^2}\}] \times \left(\frac{1}{\sqrt{2\pi}\sqrt{\sigma_x^2}}\exp[-\frac{(x_i-\mu_x)^2}{2\sigma_x^2}]\right)^{-1}$$

It it not difficult, but a little tedious[1], to show that

$$f(y_i \mid x_i) = A\exp[\frac{-1}{2\sigma_y^2(1-\rho^2)}\{y_i - \mu_y + \rho\frac{\sigma_y}{\sigma_x}\mu_x - \rho\frac{\sigma_y}{\sigma_x}x_i\}^2], \tag{4}$$

where

$$A = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_y^2(1-\rho^2)}}.$$

Next, define the conditional expectation of y_i, $\mathsf{E}[y_i \mid x_i]$ as:

$$E[y_i \mid x_i] = \mu_y - \rho\frac{\sigma_y}{\sigma_x}\mu_x + \rho\frac{\sigma_y}{\sigma_x}x_i \tag{5}$$

and the conditional variance of y_i as

$$Var[y_i \mid x_i] = \sigma_y^2(1-\rho^2) \tag{6}$$

We then have the following important results:

1. The conditional PDF of y_i given by (4) is a normal PDF.
2. If y_i and x_i are correlated, $\rho^2 \neq 0$, the conditional mean of y_i is given by (5) and is a function of x_i. (5) is called the *regression function.*
3. If y_i and x_i are correlated, $\rho^2 \neq 0$, the conditional variance of y_i given by (6) is less than the marginal variance σ_y^2 :

$$Var[y_i \mid x_i] < \sigma_y^2 \text{ iff } \rho^2 \neq 0 \tag{7}$$

"Model form":

We now define the disturbance ε_i by

$$\varepsilon_i = y_t - \mathsf{E}[y_i \mid x_i], \tag{8}$$

which of course is a stochastic variable with a normal distribution, given the initial assumption. We obtain the *regression model* as

$$y_i = \mathsf{E}[y_i \mid x_i] + \varepsilon_i = \beta_1 + \beta_2 x_i + \varepsilon_i \tag{9}$$

where the parameters of the model is linked to the PDF of the stochastic variables in the following way

[1]Bårdsen and Nymoen (2011), Tillegg 4.A goes through all the steps.

$$\beta_1 = \mu_y - \rho_{xy}\frac{\sigma_y}{\sigma_x}\mu_x = \mu_y - \frac{\sigma_{xy}}{\sigma_x^2}\mu_x \tag{10}$$

$$\beta_2 = \rho_{xy}\frac{\sigma_y}{\sigma_{x_i}} = \frac{\sigma_{xy}}{\sigma_x^2} \tag{11}$$

Finally, $\varepsilon_i \sim N(0, \sigma^2)$ with $\sigma^2 = \sigma_y^2(1 - \rho^2)$.

Summary

- Above we have derived the population regression model for the case where the PDF of $\{x_i, y_i\}$ is binormal.
- The results generalize to the multivariate case where $\{y_i$,x_{1i}, x_{2i},...,$x_k\}$ have a multivariate normal PDF.
- It shows that the properties listed on page 92 in Greene's book: A3, A4 and A6 are inherent model properties is this case
- A3 also holds for other joint PDFs.
- The more specific results also holds such as $\beta_2 = \rho_{xy}\frac{\sigma_y}{\sigma_x}$ (only that the moments then refer to that other PDF (i.e. not the normal))
- A4 may not hold for other PDFs, the variance of the disturbance may by non-constant and may also depend on x_i. It is then heteroscedastic. With time series data, we may also have $Cov[\varepsilon_t, \varepsilon_{t-j}] \neq 0$ for $j = 1, 2, ...$ This is called autocorrelation.

Finally: The conditional expectations function need not be linear! This means that one of the most important assumptions/choices that we make in conditional econometric modelling is the specification of the *functional form* of the expectations function.

Transformation of variables, prior to modelling, should be viewed in this context.

References

Biorn, E. (2003). *Økonometriske emner*. Unipub forlag, 2nd edn.

Bårdsen, G. and R. Nymoen (2011). *Innføring i økonometri*. Fagbokforlaget.

Greene, W. (2012). *Econometric Analysis*. Pearson, 7th edn.