Autocorrelation

One of the important assumption of OLS is that there is no autocorrelation, i.e. the random disturbance term is related with its own values.

$$\text{i.e. } cov(u_i, u_j) \neq 0 \quad \forall \ i \neq j$$

$$\Rightarrow \quad E(u_i u_j) \neq 0 \quad \text{as}$$

$$E(u_i) = E(u_j) = 0 \quad \forall \ i \neq j$$

This is known as autocorrelation.

It is a time series problem.

We will consider 1st auto regressive errors [AR(1)] i.e.

$$u_t = \rho u_{t-1} + e_t$$

where e_t are serially uncorrelated with mean zero and variance σ_e^2.

is called an autoregression because it is usual regression model with u_t regressed on u_{t-1}.

It is called 1st order auto-regression because u_t is regressed on its past with only one lag.

Now, we will derive $Var(u_t)$ and the correlation between u_t and lagged values of u_t. Thus u_t depends on e_t, e_{t-1}, e_{t-2} .... Since e_t are serially independent and u_{t-1} depends on e_{t-1}, e_{t-2} and so on but not on e_t we have,

$$E(u_{t-1}, e_t) = 0$$

Since $E(e_t) = 0$, we have $E(u_t) = 0$ for all t.

If we denote $Var(u_t)$ by σ^2 we have,

$$\sigma^2 = Var(u_t) = E(u_t^2)$$

$$= E(\rho u_{t-1} + e_t)^2$$

$$= \rho^2\sigma^2 + \sigma_e^2 \quad \text{since } cov(u_{t-1}, e_t) = 0$$

Thus we have,

$$\sigma^2 = \frac{\sigma_e^2}{1-\rho^2}$$

This gives the variance of u_t in terms of the variance of e_t and the parameter ρ.

Let us now derive the correlations :

Denoting the correlation between u_t and u_{t-s} by ρ_s, we get,

$$E(u_t \cdot u_{t-s}) = \sigma^2 \rho_s$$

But $E(u_t \, u_{t-s})$

$$= \rho \, E(u_{t-1} \, u_{t-s}) + E(e_t \, u_{t-s})$$

Hence, $\rho_s = \rho \cdot \rho_{s-1} + 0$

$\Rightarrow \rho_s = \rho \cdot \rho_{s-1}$

Since, $\rho_0 = 1$ we get by successive substitution.

$$\rho_1 = \rho, \quad \rho_2 = \rho^2, \quad \rho_3 = \rho^3, \ldots$$

Thus, the lag correlations are all powers of ρ and decline geometrically. These expressions can be used to derive the co-variance matrix of the errors and using what is known as generalised least squares (GLS).

Consider the model.

$$Y_t = \alpha + \beta x_t + u_t$$

$$u_t = \rho u_{t-1} + e_t$$

By applying OLS in the above model, we get the OLS estimators are unbiased. But they will not be efficient. Further the tests of significance are apply, which will be based on the wrong co-variance matrix will be wrong.

Here, $Var(u_t) = \sigma^2$

$$cov(u_t\ u_{t-j}) = \rho_j \sigma^2$$

If u_t are $AR(1)$ we have $\rho_j = \rho^j$

The OLS estimator of β is.

$$\hat{\beta} = \frac{\sum x_t Y_t}{\sum x_t^2}$$

Hence, $$\hat{\beta} - \beta = \frac{\sum x_t u_t}{\sum x_t^2}$$

and $$E(\hat{\beta} - \beta) = 0$$

$$Var(\hat{\beta}) = \frac{1}{(\sum x_t^2)^2} Var(\sum x_t u_t)$$

$$= \frac{\sigma^2}{(\Sigma x_t^2)^2}\left[\Sigma x_t^2 + 2\rho \Sigma x_t x_{t-1} + 2\rho^2 \Sigma x_t x_{t-2} + \ldots\right]$$

Since $\operatorname{cov}(u_t, u_{t-j}) = \rho^j \cdot \sigma^2$

Thus we have

$$\operatorname{Var}(\hat{\beta}) = \frac{\sigma^2}{(\Sigma x_t^2)^2}\left[1 + 2\rho \frac{\Sigma x_t x_{t-1}}{\Sigma x_t^2} + 2\rho^2 \frac{\Sigma x_t x_{t-2}}{\Sigma x_t^2} + \ldots\right]$$

If we ignore the autocorrelation problem, we would be computing

$$\operatorname{Var}(\hat{\beta}) = \sigma^2 / \Sigma x_t^2.$$

Thus, we would be ignoring the expression in the bracket of the above expression. To get an idea of the magnitude of this expression, let us assume that the x_t series also follow an AR(1) process with $\operatorname{Var}(x_t) = \sigma_x^2$

and $cov(x_t; x_{t-1}) = r$. Since, we are now assuming x_t to be stochastic we will consider the asymptotic variance of β. The expression in the bracket is now,

$$1 + 2\rho r + 2\rho^2 r^2 + \cdots$$

$$= 1 + \frac{2\rho r}{1-\rho r} = \frac{1+\rho r}{1-\rho r}$$

Thus, $$var(\hat{\beta}) = \frac{\sigma^2}{n\sigma_x^2}\left(\frac{1+\rho r}{1-\rho r}\right)$$

where n is the no. of obs.

We know that, $$Var(u) = \sigma^2 = \frac{\sigma_e^2}{1-\rho^2}$$

or, $$\sigma_e^2 = \sigma^2(1-\rho^2)$$

$$\therefore Var(\hat{\beta}) = \frac{\sigma_e^2}{n\sigma_x^2(1-\rho^2)}\left(\frac{1+\rho r}{1-\rho r}\right)$$

Thus, the consequences of autocorrelation are $\rightarrow$

i) The least squares estimators are unbiased but are not efficient

Sometimes they are considerably less efficient than the procedures that take account of the autocorrelation.

(2) The sampling variances are biased and sometimes likely to be seriously understated. Thus R^2 as well as t and F statistics tend be exaggerated.