Serial Correlation in the OLS

Charles Christopher Hyland

Semester 2 2017

Recall that OLS parameter is:

$$\hat{\beta}_1 = \beta_1 + \frac{\sum\limits_{t=1}^{n} x_i \epsilon_i}{SST_x}$$

whereby $d_i = x_i$ since $\bar{x} = 0$ (normally it is $d_i = x_i - \bar{x}$). We simply let $d_i = x_i$ for this question then. Recall that variance of OLS estimator is then:

$$Var(\hat{\beta}_1) = Var(\frac{\sum\limits_{t=1}^{n} x_i \epsilon_i}{SST_x})$$

$$= \frac{1}{SST_x^2} Var(\sum\limits_{t=1}^{n} x_i \epsilon_i)$$

since SST_x is a constant. Now we ignore the fraction in the front (because I'm lazy). Recall we are conditioning on $\mathbf{X}$ so x_i is a constant.

$$= Var(\sum\limits_{t=1}^{n} x_t \epsilon_t | \mathbf{X})$$

$$= \sum\limits_{t=1}^{n-1} \sum\limits_{j=1}^{n-t} x_t x_{t+j} Cov(\epsilon_t, \epsilon_{t+j})$$

$$= \sum\limits_{t=1}^{n-1} x_t^2 Var(\epsilon_t) + 2 \sum\limits_{t=1}^{n-1} \sum\limits_{j=1}^{n-t} x_t x_{t+j} Cov(\epsilon_t, \epsilon_{t+j})$$

Recall that $Cov(\epsilon_t, \epsilon_{t+j}) = E(\epsilon_t, \epsilon_{t+j})$.

$$= \sum\limits_{t=1}^{n-1} x_t^2 Var(\epsilon_t) + 2 \sum\limits_{t=1}^{n-1} \sum\limits_{j=1}^{n-t} x_t x_{t+j} E(\epsilon_t, \epsilon_{t+j})$$

Since ϵ_t is a AR(1) serial correlation, it has the form of a random walk. Recall that for a random walk, $E(\epsilon_t, \epsilon_{t+j}) = \rho^j \sigma^2$.

$$= \sum_{t=1}^{n-1} x_t^2 Var(\epsilon_t) + 2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} x_t x_{t+j} \rho^j \sigma^2$$

$$= \sum_{t=1}^{n-1} x_t^2 Var(\epsilon_t) + 2\sigma^2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} x_t x_{t+j} \rho^j$$

since σ^2 is a constant. Now recall that $SST_x = \sum_{t=1}^{n} x_t^2$.

$$= SST_x Var(\sum_{t=1}^{n} \epsilon_t) + 2\sigma^2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}$$

Now we add in our earlier fraction in the front $\frac{1}{SST_x^2}$

$$= \frac{SST_x Var(\epsilon_t)}{SST_x^2} + \frac{2\sigma^2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}}{SST_x^2}$$

$$= \frac{Var(\epsilon_t)}{SST_x} + \frac{2\sigma^2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}}{SST_x^2}$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_x} + \frac{2\sigma^2 \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} \rho^j x_t x_{t+j}}{SST_x^2}$$

where $var(\epsilon_t = \sigma^2$ since it is still homoskedastic. The first term is the usual variance of an $\hat{\beta}_1$. The second term is the bias. Note that if $\rho > 1$, then the variance of $\hat{\beta}_1$ for an OLS estimate will be underestimating the true variance since it does not take into account of the second term. From this, usual OLS when assuming no serial correlation (but there actually is), will cause the standard errors to be smaller than what it actually should be, leading to test statistics that are too big and then committing too many type 1 errors (falsely rejecting null hypothesis).