Lecture 07

Assumptions:

$$\begin{array}{lll}(\mathrm{C.6}) & u_t \perp u_s, & \forall\, t \neq s \\ (\mathrm{C.7}) & u_t \perp x_t, & \forall\, t \in \{1, 2, \ldots, T\}\end{array}$$

Considering $(\mathrm{C.7})$:

- Given the model $y_t = \beta_1 + \beta_2 x_{2,t} + u_t$, then $\hat{\beta}_2^{OLS} = \dfrac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sum(x_t - \bar{x})^2}$.
- Use that $(y_t - \bar{y}) = \beta_2(x_t - \bar{x}) + (u_t - \bar{u})$ - with $\bar{u} = 0$ -, since $\hat{\beta}_1 = \bar{y} - \hat{\beta}_2\bar{x}$ (more here).
- We have $\hat{\beta}_2^{OLS} = \beta_2 \dfrac{\sum(x_t - \bar{x})^2}{\sum(x_t - \bar{x})^2} + \dfrac{\sum(x_t - \bar{x})u_t}{\sum(x_t - \bar{x})^2} = \beta_2 + \dfrac{\sum(x_t - \bar{x})u_t}{\sum(x_t - \bar{x})^2}$
  - Let $a_t := \dfrac{x_t - \bar{x}}{\sum(x_t - \bar{x})^2}$, then $\hat{\beta}_2 = \beta_2 + \sum a_t u_t$.
- To show it is unbiased, we take the expected value: $\mathrm{E}(\hat{\beta}_2) = \beta_2 + \mathrm{E}\big[\sum(a_t u_t)\big]$.
  - If a_t and u_t are uncorrelated, then $\mathrm{E}(\hat{\beta}_2) = \beta_2 + \sum\big[\mathrm{E}(a_t)\mathrm{E}(u_t)\big]$.
  - Since $\mathrm{E}(u_t) = 0$, then $\mathrm{E}(\hat{\beta}_2) = \beta_2$, provided that a_t and u_t are uncorrelated.
  - a_t depends upon **all** $x_t, \quad t = 1, 2, \ldots, T$.

1. If x_t (current) and u_t are correlated, then $\hat{\beta}_2$ is biased ($\mathrm{E}(\hat{\beta}_2) \neq \beta_2$) and inconsistent - $\operatorname*{plim}_{N \to \infty} \hat{\beta}_2 \neq \beta_2$.
   - More on the bias and consistency of an estimator here and here, respectively.
2. If x_t and u_t are uncorrelated, but x_t and u_s are correlated, for $t \neq s$, then $\hat{\beta}_2$ is biased but consistent.

Basic models

$$(1) \quad y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad u_t = \rho u_{t-1} + e_t$$

- In this case, $\hat{\beta}_2$ is unbiased and consistent, but inefficient, because u_t and u_{t-1} are correlated.
  - $(\mathrm{C.6})$ might be an example.
  - More on the efficiency of an estimator here.

$$(2) \quad y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 y_{t-1} + u_t$$

- y_{t-1} is a function of u_{t-1} .
  - y_{t-1} is correlated to u_{t-1}.
  - In other words, the error is correlated with a variable that is not its contemporaneous.
- In this case, $\hat{\beta}_2$ is biased and inefficient but consistent.

$$(3) \quad y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 y_{t-1} + u_t, \quad u_t = \rho u_{t-1} + e_t$$

- The model can be rewritten as: $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 y_{t-1} + \rho u_{t-1} + e_t$.
- Since u_{t-1} is inside y_{t-1}, then u_{t-1} and y_{t-1} are correlated.
  - Now, the errors are dependent upon its contemporaneous *and* other variables.

- Then, $\hat{\beta}_2$ is biased, inefficient and inconsistent.

For the model:

$$(1) \quad y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad u_t = \rho u_{t-1} + e_t$$

- It violates the assumption $(\mathrm{C}.6)$, because the errors are correlated.
- $\hat{\beta}$ is thus unbiased, inconsistent and inefficient.
    - When positive serial correlation exists (*i. e.*, $\rho > 0$), the standard errors are too small.
    - The t statistic $t = \dfrac{\hat{\beta} - \beta^{H_0}}{\sigma_{\hat{\beta}}}$ then blows up.
    - F statistics will get too big, as well.
- Misspecification can also give rise to serial correlation:
    - Suppose a serially correlated variable x_3 (say, GDP) is left out of the model.
        - The effect of that variable would be added to the errors, making them serially correlated, as well.
    - Trying to fit a non-linear relation (say, quadratic) through linear regression may give the impression of serial correlation.
        - If the line is secant to the curve, for instance, the errors would be positive-negative-positive (or negative-positive-negative) and would *seem* serially correlated.

Durbin-Watson test for autocorrelation

This is a test for first order serial correlation **only**.

- For instance, a model $y_t = \beta_1 + \beta_2 x_{2t} + u_t, \quad u_t = \rho u_{t-1} + e_t$ with $|\rho| < 1$.
- This is called a first-order autoregressive process (or $\mathrm{AR}(1)$), because it has one lag.
    - Analogously, if the errors were defined as $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + e_t$, it would be an $\mathrm{AR}(3)$.
    - Since there is no lag on the term e_t, this a 0^{th} order moving average process ($\mathrm{MA}(0)$).
        - An error defined as, say, $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \delta_1 e_{t-1} + \delta_2 e_{t-2}$ would be classified as an autoregressive moving average model $\mathrm{ARMA}(3, 2)$.
        - More generally, an $\mathrm{ARMA}(p, q)$ is given by $u_t = \sum_{i=1}^{p} \rho_i u_{t-i} + \sum_{j=1}^{q} \delta_j e_{t-j}$, for $p, q < t$.
    - For more on autoregressive models, see this link.

Hypotheses:

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

Steps:

1. Estimate $y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t$ by OLS, save $\hat{u}_t$.

2. Compute the test statistic $d = \frac{\sum\limits_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum\limits_{t=1}^{T}\hat{u}_t^2}$, so that $0 < d < 4$.
   - $0 < d < 2$ would suggest a positive serial correlation one-tailed test, while $2 < d < 4$, a negative serial correlation one-tailed test.
3. (a) For the one-tailed test version of the test $H_1 : \rho > 0$ (testing for positive serial correlation), obtain from the table (external link here) the critical lower and upper values d_{L} and d_{U} for the N observations, k variables and significance α.
   - For the positive test, the area of interest will be the interval $[0, 2]$, so that $d_{\mathrm{L}}, d_{\mathrm{U}} \in [0, 2]$ with $d_{\mathrm{L}} < d_{\mathrm{U}}$.
   - If the test statistic $d_{\mathrm{U}} < d_{\text{test}} < 2$, we fail to reject the null hypothesis.
   - If $0 < d_{\text{test}} < d_{\mathrm{L}}$, we reject the null hypothesis.
   - Otherwise, if $d_{\mathrm{L}} < d_{\text{test}} < d_{\mathrm{U}}$, the (positive serial correlation) test is inconclusive.
     - Usually, another test is performed (namely, the LM test).

3. (b) For the other one-tailed version of the test $H_1 : \rho < 0$ (testing for negative serial correlation), obtain d_{L} and d_{U} for the given N, k and α.
   - For the negative test, the area of interest will be the interval $[2, 4]$ so that $(4 - d_{\mathrm{U}}), (4 - d_{\mathrm{L}}) \in [2, 4]$ with $(4 - d_{\mathrm{U}}) < (4 - d_{\mathrm{L}})$.
   - If $2 < d_{\text{test}} < (4 - d_{\mathrm{U}})$ - *id est*, $(4 - d) > d_{\mathrm{U}}$ -, we fail to reject the null hypothesis.
   - If $(4 - d_{\mathrm{L}}) < d_{\text{test}} < 4$ - *id est*, $(4 - d) < d_{\mathrm{L}}$ -, we reject the null hypothesis.
   - Otherwise, if $(4 - d_{\mathrm{U}}) < d_{\text{test}} < (4 - d_{\mathrm{L}})$ - *id est*, $d_{\mathrm{L}} < (4 - d) < d_{\mathrm{U}}$ -, the (negative serial correlation) test is inconclusive.

For more on the Durbin-Watson test, see this link.

Why is d between 0 and 4?

- The correlation can be estimated by $\hat{\rho} = \frac{\sum\limits_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}}{\sum\limits_{t=1}^{T}\hat{u}_t^2}$.
  - Then, $\hat{\rho} = \frac{\frac{1}{(N-k)}\sum\limits_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}}{\frac{1}{(N-k)}\sum\limits_{t=1}^{T}\hat{u}_t^2} \approx \frac{\widehat{\mathrm{Cov}}(\hat{u}_t, \hat{u}_{t-1})}{\widehat{\mathrm{Var}}(\hat{u}_t)}$: note that the sum in the numerator is missing one term.
- The test statistic is given by $d = \frac{\sum\limits_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum\limits_{t=1}^{T}\hat{u}_t^2}$.
  - Expanding the numerator $d = \frac{\sum\limits_{t=2}^{T}(\hat{u}_t^2 - 2\hat{u}_t\hat{u}_{t-1} + \hat{u}_{t-1}^2)}{\sum\limits_{t=1}^{T}\hat{u}_t^2}$.

- Rewriting it: $d = \dfrac{(N-k-1)\displaystyle\sum_{t=2}^{T}\frac{1}{(N-k-1)}(\hat{u}_t^2 - 2\hat{u}_t\hat{u}_{t-1} + \hat{u}_{t-1}^2)}{(N-k)\displaystyle\sum_{t=1}^{T}\frac{1}{(N-k)}\hat{u}_t^2}$.
    - $\displaystyle\sum_{t=1}^{T}\frac{1}{(N-k)}\hat{u}_t^2$ is an estimate of the variance.
    - $\displaystyle\sum_{t=2}^{T}\frac{1}{(N-k-1)}\hat{u}_t^2$ is an estimate of the variance (with one less observation: $\hat{u}_1$).
    - $\displaystyle\sum_{t=2}^{T}\frac{1}{(N-k-1)}\hat{u}_{t-1}^2$ is also an estimate of the variance (short of one observation: $\hat{u}_t$).
    - $\displaystyle\sum_{t=2}^{T}\frac{1}{(N-k-1)}\hat{u}_t\hat{u}_{t-1}$ is an estimate of the covariance of $\hat{u}_t$ and $\hat{u}_{t-1}$.
- The test statistic may be thus estimated $d = \dfrac{N-k-1}{N-k}\left(\dfrac{\widehat{\mathrm{Var}}(\hat{u}_t) - 2\widehat{\mathrm{Cov}}(\hat{u}_t, \hat{u}_{t-1}) + \widehat{\mathrm{Var}}(\hat{u}_t)}{\widehat{\mathrm{Var}}(\hat{u}_t)}\right)$.
    - The larger the number of observations, the closer $\dfrac{N-k-1}{N-k}$ is to 1.
    - $\dfrac{\widehat{\mathrm{Var}}(\hat{u}_t) - 2\widehat{\mathrm{Cov}}(\hat{u}_t, \hat{u}_{t-1}) + \widehat{\mathrm{Var}}(\hat{u}_t)}{\widehat{\mathrm{Var}}(\hat{u}_t)} = 2\left(1 - \dfrac{\widehat{\mathrm{Cov}}(\hat{u}_t, \hat{u}_{t-1})}{\widehat{\mathrm{Var}}(\hat{u}_t)}\right)$.
- Thus, $d \approx 2(1-\hat{\rho})$.
    - Asymptotically, as $t \to \infty$, $d = 2(1-\rho)$.
    - If there is perfect positive correlation ($\rho = 1$), then $d = 0$.
    - If there is no correlation ($\rho = 0$), then $d = 2$.
    - If there is perfect negative correlation ($\rho = -1$), then $d = 4$.

Observation:

$$\hat{\beta}_2^{\mathrm{OLS}} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

If we multiply both the numerator and denominator by $\dfrac{1}{N-k}$:

$$\hat{\beta}_2^{\mathrm{OLS}} = \frac{\frac{1}{N-k}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{N-k}\sum(X_i - \bar{X})^2}$$

then the numerator becomes an estimate of the covariance of X and Y, whereas the denominator turns into an estimate of the variance of X:

$$\hat{\beta}_2^{\mathrm{OLS}} = \frac{\widehat{\mathrm{Cov}}(X,Y)}{\widehat{\mathrm{Var}}(X)}.$$