Lecture 09

Review on F statistics:

$$\begin{aligned} H_0 &: 4\beta_2 - 2\beta_3 = 6 \\ H_1 &: 4\beta_2 - 2\beta_3 \neq 6 \end{aligned}$$

The model:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

- Rewrite the restriction: $\beta_2 = \dfrac{3}{2} + \dfrac{1}{2}\beta_3$.
- Apply it to the model: $y_i = \beta_1 + \left(\dfrac{3}{2} + \dfrac{1}{2}\beta_3\right) x_{2i} + \beta_3 x_{3i} + u_i$.
- Rewrite the result: $y_i - \dfrac{3}{2} x_{2i} = \beta_1 + \beta_3 \left(\dfrac{1}{2} x_{2i} + x_{3i}\right) + u_i$.
  - It may be useful to rename the variables: $y_i^* = \beta_1 + \beta_3 x_i^* + u_i$.
- Running this regression will produce the residual sum of squares for the restricted model RSS_{R}.
- Running the regression on the original model, one will obtain the residual sum of squares for the unrestricted model RSS_{UR}.
- The test statistic is $F = \dfrac{RSS_{\mathrm{R}} - RSS_{\mathrm{UR}}}{p_{\mathrm{UR}} - p_{\mathrm{R}}} \cdot \dfrac{n - p_{\mathrm{UR}}}{RSS_{\mathrm{UR}}}$.
  - n is the number of observations.
  - p_{R} is the number of parameters in the restricted model (in the example, 2).
  - p_{UR} is the number of parameters in the unrestricted model (3, in the given example).
- The critical value is obtained from an F distribution with $(p_{\mathrm{UR}} - p_{\mathrm{R}}, n - p_{\mathrm{UR}})$ degrees of freedom.
- If $F > F_{\text{critical}}$, reject the null hypothesis.
  - Otherwise, one fails to reject the null hypothesis

Review on heteroskedasticity:

The model:

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i \\ \mathrm{Var}(u_i) &= \sigma_i^2, \quad (i = 1, 2, \ldots, N) \end{aligned}$$

Models for the variance:

$$\begin{array}{lll} (1) & \sigma_i^2 = & \sigma^2 z_i^2 \\ (2) & \sigma_i^2 = & \alpha_1 + \alpha_2 z_{2i} + \ldots + \alpha_p z_{pi} \\ (3) & \sigma_i = & \alpha_1 + \alpha_2 z_{2i} + \ldots + \alpha_p z_{pi} \\ (4) & \ln \sigma_i^2 = & \alpha_1 + \alpha_2 z_{2i} + \ldots + \alpha_p z_{pi} \end{array}$$

- One can think of the σ^2 in (1) as an α : it is simply another parameter to estimate.
- (1) is a special case of (3).

Generally, solving heteroskedasticity involves dividing the original model by the standard deviation σ_i^2.

- Originally, $\mathrm{Var}(u_i) = \sigma_i^2$: this is the source of the problem.
- Once the model is divided by σ_i, then $\mathrm{Var}\left(\frac{u_i}{\sigma_i}\right) = \frac{1}{\sigma_i^2}\mathrm{Var}(u_i) = \frac{\sigma_i^2}{\sigma_i^2} = 1.$
  - More generally, $\mathrm{Var}\left(\frac{u_i}{c \cdot \sigma_i}\right) = \frac{1}{c^2 \cdot \sigma_i^2}\mathrm{Var}(u_i) = \frac{\sigma_i^2}{c^2\sigma_i^2} = \frac{1}{c^2}$, where $c \in \mathbb{R}$ is a constant.
- **The brawn of most methods for correcting heteroskedasticity is to find a model to estimate the variance σ_i by which to divide the original linear model**.
  - Considering the model of the variance (1): dividing the original model by z_i would have the same effect, since that is the variable that defines the behavior of σ_i^2 in the model.
  - Dividing by the original model by $\sigma \cdot z_i$ would have the same effect, given that σ is a constant.
  - In both cases (dividing by z_i or σz_i), one will be in fact dividing the original model by a constant times the standard deviation (or rather an estimate thereof).
    - First case: $\mathrm{Var}\left(\frac{u_i}{z_i}\right) = \frac{1}{z_i^2}\mathrm{Var}(u_i) = \frac{\sigma^2 z_i^2}{z_i^2} = \sigma^2.$
    - Second case: $\mathrm{Var}\left(\frac{u_i}{\sigma z_i}\right) = \frac{1}{\sigma^2 z_i^2}\mathrm{Var}(u_i) = \frac{\sigma^2 z_i^2}{\sigma^2 z_i^2} = 1.$

And why is heteroskedasticity really a problem?

> If the variance varies with the data points, those with larger variances will be given larger weights by OLS. That means that less precise points (those with larger variances to the regression model) will have more importance in determining the parameters. When the whole model is divided by the standard deviation, it is rebalanced, since the observations with larger variances are down-weighted.

Review on the Durbin-Watson test:

Why is the test statistic d between 0 and 4?

- Remembering:
  - The correlation of two variables is given by: $\rho_{XY} = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$, where $\mathrm{Cov}(X,Y)$ is the covariance of X and Y, while σ_X and σ_Y are standard deviations of those variables.
    - A population covariance is given by $\mathrm{Cov}(X,Y) = \frac{1}{N}\sum(x_i - \bar{x})(y_i - \bar{y})$, where N is the population size.
    - A sample covariance is given by $\mathrm{Cov}(X,Y) = \frac{1}{n-1}\sum(x_i - \bar{x})(y_i - \bar{y})$, where n is the sample size.
    - The covariance can also be written as a expected value: $\mathrm{Cov}(X,Y) = \mathrm{E}\big[\big(X - \mathrm{E}[X]\big)\big(Y - \mathrm{E}[Y]\big)\big]$.

The test statistic is given by:

$$d = \frac{\sum_{i=2}^{N}(u_i - u_{i-1})^2}{\sum_{i=1}^{N} u_i^2}$$

- The denominator, when divided by the number of observations, works as an estimate of the variance of u_i.
- If we multiply both the numerator and the denominator by $\frac{1}{N}$ and, we obtain:
  $$d = \frac{\frac{1}{N}\sum_{i=2}^{N}(u_i^2 - 2u_t u_{t-1} + u_{i-1}^2)}{\frac{1}{N}\sum_{i=1}^{N} u_i^2}.$$
  - That can be rewritten as: $d = \frac{\frac{1}{N}\sum_{i=2}^{N} u_i^2 - 2\frac{1}{N}\sum_{i=2}^{N} u_t u_{t-1} + \frac{1}{N}\sum_{i=2}^{N} u_{i-1}^2)}{\frac{1}{N}\sum_{i=1}^{N} u_i^2}$.
    - $\frac{1}{N}\sum_{i=1}^{N} u_i^2$ is an estimate of the variance σ^2.
    - $\frac{1}{N}\sum_{i=2}^{N} u_i^2$ is also an estimate of the variance, but shorter an observation (from u_2 to u_N).
    - $\frac{1}{N}\sum_{i=2}^{N} u_{i-1}^2$, too, is an estimate of the variance, shorter an observation (from u_1 to u_{N-1}).
    - $\frac{1}{N}\sum_{i=2}^{N} u_i u_{i-1}$ is an estimate of the the covariance of u_i and u_{i-1}.
  - Thus, the test statistic can be approximated as $d \approx \frac{\mathrm{Var}(u) - 2\mathrm{Cov}(u_i, u_{i-1}) + \mathrm{Var}(u)}{\mathrm{Var}(u)}$.
    - That means $d \approx 2 - 2\frac{\mathrm{Cov}(u, u_{i-1})}{\mathrm{Var}(u)}$.
    - The correlation between a variable and itself can be written as $\mathrm{Corr}(X, X) = \frac{\mathrm{Cov}(X, X)}{\sigma_X \sigma_X} = \frac{\mathrm{Cov}(X, X)}{\sigma_X^2}$.
    - Thus, $d \approx 2 - 2\mathrm{Corr}(u_i, u_{i-1}) = 2(1 - \rho)$.
  - When there is perfect positive correlation, $\rho = 1 \Rightarrow d = 0$.
  - When there is perfect negative correlation, $\rho = -1 \Rightarrow d = 4$.
  - When there is no correlation, $\rho = 0 \Rightarrow d = 2$.

Review on the Durbin H statistic:

The model:

$$y_t = \beta_1 + \beta_2 x_{2t} + \ldots + \beta_k x_{kt} + u_t$$
$$u_t = \rho u_{t-1} + e_t$$

When one of x_{jt} $(j = 2, \ldots, k)$ is lagged variable, for instance $x_{kt} = y_{t-1}$, bias is introduced.

- From the model, we know that y_{t-1} and u_{t-1} are correlated.

- Since $x_{kt} = y_{t-1}$ and $u_t = \rho u_{t-1} + e_t$, there is correlation between and error term and a RHS variable.
    - According to OLS, $\hat{\beta} = \beta + \dfrac{\sum(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}$.
    - If we multiply both the numerator and the denominator by $\frac{1}{N}$, we get $\hat{\beta} = \beta + \dfrac{\frac{1}{N}\sum(x_i - \bar{x})u_i}{\frac{1}{N}\sum(x_i - \bar{x})^2}$.
    - Taking the expected value, we get: $\mathrm{E}[\hat{\beta}] = \beta + \mathrm{E}\left[\dfrac{\frac{1}{N}\sum(x_i - \bar{x})u_i}{\frac{1}{N}\sum(x_i - \bar{x})^2}\right]$.
    - Then, $\mathrm{E}[\hat{\beta}] = \beta + \dfrac{\mathrm{Cov}(x, u)}{\mathrm{Var}(x)}$.
    - So, whenever $\mathrm{Corr}(x, u) \neq 0$, the estimated parameter will be biased, *id est*, $\mathrm{E}[\hat{\beta}] \neq \beta$.
        - If the parameters are biased, the estimated errors u_t will be biased, as well.
        - If the errors are biased, the Durbin-Watson d statistic, too, will be biased.
- Then, the estimated error becomes biased, and so does the Durbin-Watson d statistic.

The Durbin H statistic is a means of overcoming that bias and it is given by:

$$\begin{aligned} h &= \left(1 - \frac{1}{2}d\right)\sqrt{\frac{T}{1 - T \cdot \widehat{\mathrm{Var}}(\hat{\beta}_j)}} \\ &= \hat{\rho}\sqrt{\frac{T}{1 - T \cdot \widehat{\mathrm{Var}}(\hat{\beta}_j)}} \sim \mathcal{N}(0, 1) \end{aligned}$$

where d is the Durbin-Watson statistic, T is the length of the time series, and $\hat{\beta}$ is the estimated parameter for the x_{jt} that equals the lagged variable.

- **This test only works if** $T \cdot \widehat{\mathrm{Var}}(\hat{\beta}_j) < 1$.
    - Otherwise, do the Breusch-Godfrey LM test.