Lecture 03

Heteroskedasticity

- When errors are not independently and identically distributed (iid).
- In other words, $u_i \sim (0, \sigma_i)$ (errors have different variances).

How might heteroskedasticity arise?

1. Learning:
    - An adaptive behavior might lead to a decrease in variance over time, as more information is acquired.
2. Scale variables:
    - income, wealth, sales etc.
    - For instance, in a simple model of consumption (c_i) as a (linear) function of income (y_i):
        - That is: $c_i = \beta_1 + \beta_2 \cdot y_i + u_i$;
        - The variance of consumption across the lower levels of income are expected to be smaller than the variance for the higher levels.
3. Better data collection techniques:
    - An example: the data for US. macroeconomic variable collected before and after the WWII:
        - The employment of better collection techniques reduced the variance of the data.
4. Outliers:
    - The estimation of parameters can be stated as a minimization problem: $\underset{\beta}{min} \ \sum \hat{u}_i^2$;
    - Thus, the estimation can be heavily influenced by outliers;
    - An alternative might be the use of mean absolute deviation (MAD) estimators: $\sum |\hat{u}_i|$.
5. Model incorrectly specified:
    - Suppose that an important variable is left out of the model:
        - In what should be $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$, the x_{3i} term is left out, being understood as part of the error;
        - The model would be specified as $y_i = \beta_1 + \beta_2 x_{2i} + u_i'$:
            - For instance, if x_{3i} is a scale variable, it will pass on that behavior to the errors.
6. Incorrect data transformations:
    - Applying log or squaring erroneously, for example.
7. Incorrect functional form:
    - Trying to fit a blatant non-linear relationship through a linear model;
    - For example, when trying to fit a quadratic relationship with a linear equation,the errors will tend to increase over time.

<u>Suppose we have heteroskedasticity, what if we do OLS anyway</u>?

- If the data points, although having different variances, are centered around the right value, OLS is still unbiased and consistent.

- Unbiasedness: for a given N (sample size), the distribution of the estimated parameter ($\hat{\beta}$) is centered on the true value (β) ;
- Consistency: as N approaches infinity, the estimated parameter converges to the true value:
    - $\operatorname*{plim}_{N \to \infty} \hat{\beta} = \beta$ (for more on probability limit, see this);
    - Example: $E(\hat{\beta}) = \beta + \dfrac{1}{N}$:
        - Biased for small values of N ;
        - Consistent as $N \to \infty$.
- Unbiasedness implies consistency, but the other way around is not necessarily true.

- $\hat{\beta}_{OLS}$ is not efficient, *i. e.*, $\text{Var}(\hat{\beta})$ is biased and inefficient:
    - $\text{Var}(\hat{\beta}_{OLS}) = \dfrac{\sigma^2}{\sum (x_i - \bar{x})^2}$;
    - However, when there is heteroskedasticity, $\text{Var}(\hat{\beta}) = \dfrac{\sigma_i^2 \sum x_i^2}{\left(\sum x_i^2\right)^2}$;
    - Then, using OLS will produce a wrong value.

- Since $\text{Var}(\hat{\beta})$ is biased, then the statistic $t = \dfrac{\hat{\beta} - \beta^{H_0}}{\sqrt{\text{Var}(\hat{\beta})}}$ is wrong.

Testing for heteroskedasticity

- Only suggestive, but helpful:
    - Graph $\hat{e}_i^2$ against the variable suspected of causing the problem;
    - If there is no heteroskedasticity, the graph will be fairly constant;
    - Otherwise, the graph will be somehow sloped.

1. La Grange Multiplier Tests for Heteroskedasticity (LM tests):
    - Fairly robust: performs pretty well in the presence of specification errors.
    - For $y_i = \beta_1 + \beta_2 x_{2i} + \ldots \beta_k x_{ki} + u_i$, assuming $\text{Var}(u_i) = \sigma_i^2$ (there is heteroskedasticity) ;
    - With N observations:
        - How many parameters do we need to estimate? k ;
        - How many σ_i^2's are there to estimate? N ;
        - Total number of estimations: $N + k \quad \Rightarrow$ unfeasible with only N observations.
    - Thus, suppose that the variances can be modeled as $\sigma_i^2 = \alpha x_i^2$ (α is a parameter):
        - There is only one more parameter to estimate, that is, the number of estimations is then $k + 1$;
        - This model captures well heteroskedasticity stemmed from scale variables;
        - Alternative models for the variance:
            1. Breusch-Pagan: $\sigma_i^2 = \alpha_1 + \alpha_2 z_{2i} + \ldots + \alpha_p z_{pi}$;
            2. Glejser: $\sigma_i = \alpha_1 + \alpha_2 z_{2i} + \ldots + \alpha_p z_{pi}$;
            3. Park: $\ln(\sigma_i^2) = \alpha_1 + \alpha_2 z_{2i} + \ldots + \alpha_p z_{pi}$.
            - The z variables can be the x's (or some of them) or not;
            - There will be a k number of β's and p different α's to estimate: the total amount of estimations will be $k + p < N$.
            - How many restrictions are needed to get the constant variance?

        - For $H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_p = 0 \quad \Rightarrow \quad (p-1)$ restrictions;
        - Perform a χ^2 test with $p-1$ restrictions.
- Breusch-Pagan model: $\sigma_i^2 = \alpha_1 + \alpha_2 z_{2i} + \ldots + \alpha_p z_{pi}$:
    - $\sigma_i^2 = f(z)$;
    - Suitable for when the variances increase (or decrease) linearly;
    - σ_i^2 is estimated by $\hat{u}_i^2$.
- Glejser model : $\sigma_i = \alpha_1 + \alpha_2 z_{2i} + \ldots + \alpha_p z_{pi}$:
    - $\sigma_i^2 = f(z^2)$;
    - Appropriate for when the variances increase (or decrease) at a seemingly quadratic pace;
    - σ_i is estimated by $|\hat{u}_i|$.
- Park model: $\ln(\sigma_i^2) = \alpha_1 + \alpha_2 z_{2i} + \ldots + \alpha_p z_{pi}$:
    - $\sigma_i^2 = \exp\{\alpha_1 + \alpha_2 z_{2i} + \ldots \alpha_p z_{pi}\}$;
    - Fit for when the variances increase (or decrease) in a approximately exponential manner;
    - $\ln(\sigma_i^2)$ is estimated by $\ln(\hat{u}_i^2)$.
- Steps for LM Test (Case 1 - Breusch-Pagan model):
    1. Regress y on $x_1, \ldots, x_k$, get $\hat{\beta}_{OLS}$;
    2. Compute $\hat{u}_i = y - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \ \ldots \ - \hat{\beta}_k x_{ki}$;
    3. Square $\hat{u}_i$, get $\hat{u}_i^2$;
    4. Regress $\hat{u}_i^2$ on $z_{1i}, \ldots, z_{pi}$;
    5. Compute the statistic $\mathrm{LM} = N \cdot R^2$, (where R^2 is the coefficient of determination for the auxiliary regression);
    6. $\mathrm{LM} \sim \chi^2(p-1)$: (perform a χ^2 test with $p-1$ degrees of freedom);
    7. Accept or reject $H_0 : \alpha_2 = \alpha_3 = \ldots = \alpha_p$ (variance is constant).