Heteroskedasticity

- Heteroskedasticity: the conditional variance of the error term is not constant across all observations (violates LR3).
- If other LR assumptions are satisfied, OLS estimators are still linear, unbiased, consistent, and asymptotically normal.
- But usual way of estimating standard errors leads to biased estimates and misleading inference from parametric hypothesis tests.
- Evidence of heteroskedasticity can be detected using graphical methods (e.g., residual e_i vs. a regressor x_i) or statistical tests (e.g., White's test).
- If heteroskedasticity is present, we will rely on White's heteroskedasticity-consistent (HC) standard errors to make inference.

Prediction using the Sample Regression Equation

- You can either conduct: (i) individual and (ii) mean/average prediction.
- Individual prediction aims to predict the value of the dependent variable for a specific observation (or individual); and is random as it depends on the random error ϵ_0:

$$y_0 = \beta_0 + \beta_1 x_{0,1} + \beta_2 x_{0,2} + \ldots + \beta_k x_{0,k} + \epsilon_0$$

- Mean/average prediction aims to predict the average value of the dependent variable for a specific value of the independent variable; and is a constant.

$$E(y_0) = E(Y|x_{0,1}, x_{0,2}, \ldots, x_{0,k}) = \beta_0 + \beta_1 x_{0,1} + \beta_2 x_{0,2} + \ldots + \beta_k x_{0,k}$$

- For point prediction, the predicted value for both individual and mean will be the same:

$$\hat{y}_0 = \hat{E}(Y|x_0) = \hat{\beta}_0 + \hat{\beta}_1 x_{0,1} + \hat{\beta}_2 x_{0,2} + \ldots + \hat{\beta}_k x_{0,k}$$

Prediction using the Sample Regression Equation

- You can either conduct: (i) individual and (ii) mean/average prediction.
- Individual prediction aims to predict the value of the dependent variable for a specific observation (or individual); and is random as it depends on the random error ϵ_0:

$$y_0 = \beta_0 + \beta_1 x_{0,1} + \beta_2 x_{0,2} + \ldots + \beta_k x_{0,k} + \epsilon_0$$

- Mean/average prediction aims to predict the average value of the dependent variable for a specific value of the independent variable; and is a constant.

$$E(y_0) = E(Y|x_{0,1}, x_{0,2}, \ldots, x_{0,k}) = \beta_0 + \beta_1 x_{0,1} + \beta_2 x_{0,2} + \ldots + \beta_k x_{0,k}$$

- For interval prediction, due to different variances, the confidence interval for individual prediction is wider.