

Lecture 04

Quick recap:

Linear regression: $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_k x_{ki} + u_i$, $\text{Var}(u_i) = \sigma_i^2$

Parametric models:

$$\begin{aligned}\sigma_i^2 &= \alpha_1 + \alpha_2 z_{2i} + \alpha_3 z_{3i} + \dots + \alpha_p z_{pi} & \text{(a)} \\ \sigma_i &= \alpha_1 + \alpha_2 z_{2i} + \alpha_3 z_{3i} + \dots + \alpha_p z_{pi} & \text{(b)} \\ \ln \sigma_i^2 &= \alpha_1 + \alpha_2 z_{2i} + \alpha_3 z_{3i} + \dots + \alpha_p z_{pi} & \text{(c)}\end{aligned}$$

where \hat{u}_i^2 , $|\hat{u}_i|$ and $\ln \hat{u}_i^2$ are used as estimator of σ_i^2 , σ_i and $\ln \sigma_i^2$.

Steps:

1. Regress y_i on the constant term and x_{2i}, \dots, x_{ki} : that produces all $\hat{\beta}_{OLS}$;
2. Save residuals: $\hat{u}_i = y_i - \hat{\beta}_1 - \hat{\beta}_2 x_{2i} - \dots - \hat{\beta}_k x_{ki}$;
3. Apply parametric models:
 - (a) square \hat{u}_i , obtain \hat{u}_i^2 , and regress on constant term and z_{2i}, \dots, z_{pi} ;
 - (b) find $|\hat{u}_i|$ and regress on constant term and z_{2i}, \dots, z_{pi} ;
 - (c) find $\ln \hat{u}_i^2$ and regress on constant term and z_{2i}, \dots, z_{pi} ;
4. Compute LM statistics: $LM = N \cdot R^2$;

"So, [it is] a Lagrange Multiplier because we're comparing an unconstrained to a constrained model and looking at the difference between the maximum under the unconstrained and the maximum under the constrained. If that distance is big, we reject; if that difference is small, we don't reject. [...] Lagrange Multipliers are ways of opposing constraints on maximization problems. [...] And we're looking at the distance between the two solutions. If the constraints are valid, it shouldn't matter if they're imposed, they should be in the data anyway: the two maximums should be very close together. If the restrictions are invalid, you're imposing something false on the data, it can't find the maximum anymore, and it will be very far away from the maximum (because you imposed a false assumption on it)." (Mark Thoma)

5. $LM \sim \chi^2(p-1)$ d.f. .

Since

$$\begin{aligned}H_0 &: \alpha_2 = \alpha_3 = \dots = \alpha_p = 0 \\ H_1 &: \exists \alpha_j \neq 0 \quad (j = 2, 3, \dots, p),\end{aligned}$$

if the null hypothesis is true, then:

- (a) $\sigma_i^2 = \alpha_1$;
- (b) $\sigma_i = \alpha_1$;
- (c) $\ln \sigma_i^2 = \alpha_1$.

Then a χ^2 -test should be performed with $p-1$ degrees of freedom (due to the number of restriction, from 2 to p) and the given level of significance.

If the $LM > \chi_{crit}^2$, H_0 should be rejected, which means that there is heteroskedasticity.

Otherwise, we fail to reject H_0 and we assume there to be homoskedasticity.

Moving on to a test based upon the model of variance $\sigma_i^2 = \alpha x_j^2 \dots$

Goldfeld-Quandt test

- Sort the data in ascending order of magnitude of the explanatory variable(s) and divide it into three parts;
- Ignore the middle part (which might represent from one sixth to one third of the data);
- Perform regressions for the lowest- x and highest- x data points, separately;
- Estimate the variances for each of those two parts and divide the higher variance by the lowest one:
 - Not necessarily the lowest variance will stem from the lowest x values!
- As previously seen, each of those two variances follows a χ^2 distribution; thus, the ratio of the two follows an F [distribution](#):

$$\circ \sigma_H^2 \sim \chi^2(d_1), \sigma_L^2 \sim \chi^2(d_2) \Rightarrow \frac{\sigma_H^2}{\sigma_L^2} \sim F(d_1, d_2), \text{ where } d_1 \text{ and } d_2 \text{ are,}$$

respectively, the degrees of freedom in the numerator and in the denominator.

- If there is no heteroskedasticity, that ratio should be very close to one;
- On the other hand, if the ratio is significantly higher than one, then there is heteroskedasticity.

More about this test [here](#).

- If the errors are not normal or if the variances do not follow the expected variance model, the test will perform poorly:
 - LM tests are more robust under those circumstances.
- The test attempts to explain the variance with only one variable:
 - That is, for $\sigma_i^2 = \alpha x_j^2$, j can only assume one value, so that $j \in \{1, 2, \dots, k\}$.

Detailed steps:

1. Identify a variable, say X , that is suspected of being positively or negatively related to the variance;
2. Divide sample into three parts:
 - Define N_1 and N_2 , so that $1 < N_1 < N_2 < N$;
 - It is not required that $N_1 = N - N_2 + 1$, i.e., that the first and the third parts have the same number of observations ;
 - The number of observations between N_1 and N_2 are usually one sixth to one third of all data;
3. Estimate two regressions: one for the first N_1 observations and other for the last $N - N_2 + 1$ ones;

4. Calculate $RSS_1 = \sum_{i=1}^{N_1} \hat{u}_i^2$ and $RSS_2 = \sum_{i=N_2}^N \hat{u}_i^2$;

5. Compute $F = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}$, with $\hat{\sigma}_1^2 = \frac{RSS_1}{N_1 - k}$ and $\hat{\sigma}_2^2 = \frac{RSS_2}{(N - N_2 + 1) - k}$:

◦ $F \sim F(N - N_2 + 1 - k, N_1 - k)$.