Econometrics II:

Class Note - 1 ⓥ

Date - 12/4/2021

The Chow test:

Set up: Given a regression model with two ~~sub~~ distinct subsamples A & B.

For example 1:

$$\text{Wage}_i = \beta_0 + \beta_1 \text{Edu}_i + e_i \quad ; i \in A$$

$$\& \quad \text{Wage}_j = \gamma_0 + \gamma_1 \text{Edu}_j + e_j \quad ; j \in B$$

Example 2: Time series data divided into two subsamples based on dates, e.g.,

$$y_t = \alpha + \beta t + e_t \quad , t = 1\ (1)\ t_0$$

$$y_t = \gamma + \delta t + e_t \quad , t = (t_0+1)\ (1)\ T.$$

Example 3: Cross sectional data divided into two subsamples based on a dummy variable (e.g. Education (high vs low)). or Married vs unmarried

Two choices you have in all of the above scenarios:

(1) Run two regressions, one for each subsample

(2) Run one regression ~~but~~ using the entire (pooled) sample.

Note: Purpose of the Chow test is to determine which of these two options you should choose.

Example: Let A: $i=1,2,\cdots 5$

B: $i=6,\cdots 10$

The data ~~etc~~ could be cross sectional (Male vs female) or it could be time series data (before a financial crisis and after the crisis).

Now you can fit a simple linear regression model for both the groups

$$y_i = \beta_0 + \beta_1 x_i + e_i \; ; \; i = 1(1)10 \text{ (Pooled)}$$

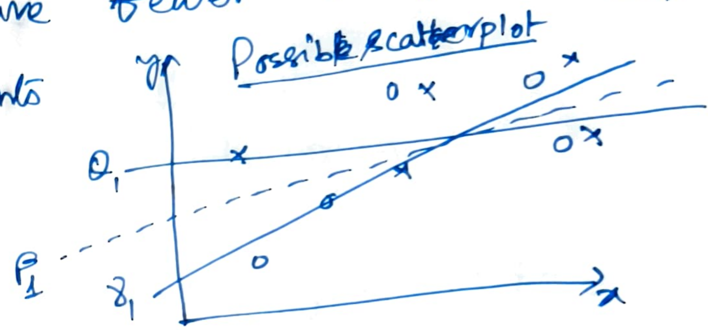
or you can fit two simple regression models for each of these two groups:

$$y_i = \theta_0 + \theta_1 x_i + e_i \quad , \; i = 1(1)5 \; : A$$

$$y_i = \delta_0 + \delta_1 x_i + e_i \quad i = 6(1)10. : B$$

Now advantages with the first model is you have 10 data points and only have two parameters whereas in the second ~~model~~ approach you have 5 data points for each of the two models consisting of two parameters each. ~~Now~~ Note that in statistics

By division of the whole sample, each regression model will have fewer samples (5 sample each) and hence the estimate coefficients will be less efficient.



However, ~~the~~ it is the case that we have two different marginal effects then, by running a pooled regression, we are forcing the marginal effects to be same.

Ho: Pooled model is correct vs Ha: Ho is not correct.

ie. Ho: $\theta_1 = \delta_1$, $\theta_0 = \delta_0$

or Ho: There is no significant improvement in fit from running two regressions

If we reject the null hypothesis using Chow test. then we will go for two seperate regression models for ~~th~~ two groups.

Step 1: Run the pooled regression and ~~estimal~~ calculate RSS_P

Step 2: Run the regressions for the two subsamples and compute RSS_A & RSS_B. (Note that $RSS_P > RSS_A + RSS_B$)

<u>Step 3:</u> Compute

$$RSS_P - RSS_A - RSS_B.$$

If this is large then Ho should be rejected.

<u>Step 4:</u> By doing slight normalization, we can have

$$F = \frac{(RSS_P - RSS_A - RSS_B)/K}{(RSS_A + RSS_B)/(n-2K)}.$$

which follows an F-distribution under Ho with d.f. K & n-2K.

ie $F \overset{H_0}{\sim} F_{K,(n-2K)}$.

· Under Ho, the F-statistic follows an F-distribution with K and (n-2K) degrees of freedom.

~~The Ch~~

<u>Note 1</u>(~~Backg~~ The Chow test was proposed by econometrician ~~Ges~~ Gregory Chow in 1960 is a test of whether the ~~tra~~ true coefficients in two linear reg.s on different data sets are equal. In econometrics, it is most commonly used in time series analysis ~~tt~~ test for the presence of a structural break at a centrain time point.