Threshold Regression

Contents

Motivation 1

Basic Approach 1

The test for the threshold 2

Motivation

Sometimes, the effect of explanatory variable on dependent variable may change when the independent variable is larger than some threshold value. - Big companies often have different investment function from small ones. - Rich people often have different consumption function from poor ones. - The housing demand of people which are borrowing constrained have different purchase behavior from those who are not constrained.

However, oftentimes we have no idea what the threshold value is. Therefore, we naturally want to estimate the threshold as well.

Basic Approach

$$y_i = \begin{cases} x_i\beta_1 + \epsilon_i, x_i < \gamma \\ x_i\beta_2 + \epsilon_i, x_i > \gamma \end{cases} \tag{a}$$

in which both β_1 and γ are parameters that we want to estimate. This is actually quite easy. We can define

$$z_{1i}(\gamma) = 1(x_i < \gamma), z_{2i}(\gamma) = 1(x_i > \gamma)$$

z_{1i} and z_{2i} are functions of γ.This implies that

$$y_i = z_{1i}(\gamma)\beta_1 + z_{2i}(\gamma)\beta_2 + \epsilon_i$$

Therefore, for a given γ, we can estimate the $\beta_1(\gamma)$ and $\beta_2(\gamma)$ and get the $SSR(\gamma)$. By trying differ γ, we can find the γ with smallest SSR.

Similarly, for the panel data, we have the same estimation strategy. Consider

$$y_{it} = \begin{cases} x_{it}\beta_1 + u_i + \epsilon_{it}, x_{it} < \gamma \\ x_{it}\beta_2 + u_i + \epsilon_{it}, x_{it} > \gamma \end{cases}$$

Suppose u_i is correlated with $x_{i:}$, therefore we want to estimate using FE model. define

$$z_{1it}(\gamma) = 1(x_{it} < \gamma), z_{2it}(\gamma) = 1(x_{it} > \gamma)$$

we therefore have

$$y_{it} = z_{1it}(\gamma)\beta_1 + z_{2it}(\gamma)\beta_2 + u_i + \epsilon_{it}$$

Define

$$y_i(\gamma) = \frac{1}{T}\sum_{t=1}^{T} y_{it}(\gamma), z_{1i}(\gamma) = \frac{1}{T}\sum_{t=1}^{T} z_{1it}(\gamma), z_{2i}(\gamma) = \frac{1}{T}\sum_{t=1}^{T} z_{2it}(\gamma)$$

and hence

$$y_{it} - y_i(\gamma) = [z_{1it}(\gamma) - z_{1i}(\gamma)]\,\beta_1 + [z_{2it}(\gamma) - z_{2i}(\gamma)]\,\beta_2 + \epsilon_{it}$$

Then, for a given γ, we can estimate the above model using ols. By trying different γ, we find out the γ that results in the minimum SSR. The above approach can be easily extended to the case where we may want multiple threshold, such as $\gamma_1, \gamma_2, , ,$

The test for the threshold

Of course, we want to the test for whether there is threshold effect. In other words, we want to test the null hypothesis:

$$\beta_1 = \beta_2$$

How to test this? The following logic is important. The regression without threshold effect looks like

$$y_i = x_i\beta + \epsilon_i \tag{b}$$

(b) can be regarded a special version of (a) in the sense (b) imposes a constraint:$\beta_1 = \beta_2$. Therefore, (b) is constrained regression, and naturally, the $SSR_b \geq SSR_a(\gamma)$. If the $\beta_1 = \beta_2$ holds, then SSR_b should not be too larger than SSR_a. Otherwise, we should reject the null hypothesis.