Theil Interpretation of Regression

James J. Heckman
University of Chicago

Econ 312, Spring 2023

Theil Weights

- Theil (1950) formula for OLS

$$Y = X\beta + \varepsilon \qquad E(\varepsilon|X) = 0$$

- OLS is weighted average of all pairwise OLS slopes.

$$\hat{\beta}_{\text{OLS}} = \frac{\sum_{1 \le i \le j \le I}(X_j - X_i)(Y_j - Y_i)}{\sum_{1 \le i \le j \le I}(X_j - X_i)^2}$$

- Form pairwise slopes (Theil)

$$b_{ji} = \frac{Y_j - Y_i}{X_j - X_i} 1[X_j \neq X_i]$$

$$\hat{\beta}_{\text{OLS}} = \sum_{1 \le i \le j \le I} b_{ji}\omega_{ji} \qquad \omega_{ji} = \frac{(X_i - X_j)^2}{N\sigma_X^2}$$

- Weights are obviously positive on each b_{ji} if $X_i \neq X_j$
- Yitzhaki, in later work, orders the X and produces a pairwise representation of OLS
- $X_1 < X_2 < \cdots < X_I$ (neglect ties)
- Concomitants $Y_1, \ldots, Y_I$
- These are the weights later used by Imbens and angrist

- Slopes for ordered data

$$b_i = \frac{Y_i - Y_{i-1}}{X_i - X_{i-1}} 1[X_i \neq X_{i-1}]$$

- Substitute into formula for OLS and collect terms on the b_i

$$\hat{\beta}_{\text{OLS}} = \sum_{i=1}^{I} b_i \omega_i \qquad \omega_i = (fracN - iN) \frac{E(X - \bar{X} | X > x_i)}{\sigma_X^2}$$

- $(N - i)/N$ is proportion of X bigger than x_i
- Obviously weights are positive.
- They place more weight on the center of the distribution of the X.