ECON 626: Empirical Microeconomics

Attenuation Bias

Department of Economics
University of Maryland
Fall 2019

1. Consider a regression of the form $y^* = \beta x^* + \epsilon$ where $x^* \sim \mathcal{U}(0, 2)$ and $\epsilon \sim \mathcal{N}(0, 1)$. You do not observe x^*; instead, you observe $x = x^* + \nu$ where $\nu \sim \mathcal{N}(0, 1/2)$. Assume $x = x^*$, ϵ, and ν are independent.

   (a) Let $\beta = 1$. Write a `.do` file that simulates this data-generating process in a sample of ten thousand observations.

   (b) You are interested in recovering the true coefficient, β. Since you know the data-generating process, you know that $\beta = 1$. How does the true coefficient compare to the coefficient that results from a regression of y^* on the observed x?

   (c) When the independent variable of interest is measured with (mean-zero) error, the OLS coefficient is biased toward zero. Let $\widehat{\beta}_x$ be the OLS coefficient resulting from a regression of y^* on x. Show that your answer to (1b) is consistent with the formula:

   $$\text{plim}\, \widehat{\beta}_x = \beta \left(1 - \frac{s}{1+s} \right)$$

   where $s = \sigma^2_\nu / \sigma^2_{x^*}$, the noise-to-signal ratio. (See Cameron & Trivedi, pp. 903–904, for discussion.)

   (d) Now imagine that you observe x^* but y^* is measured with error — specifically, assume that you only observe $y = y^* + \eta$ where $\eta \sim \mathcal{N}(0, 1/2)$. How does the estimated $\widehat{\beta}_y$ compare to the true β? How does the estimated standard error of $\widehat{\beta}_y$ compare to the estimated standard error from a regression of y^* on x^*? Why is this the case?

2. In statistics, **power** is the probability of rejecting a false null hypothesis. Consider the data-generating process $y^* = \beta x^* + \epsilon$ where $x^* \sim \mathcal{U}(-\sqrt{3}, \sqrt{3})$ and $\epsilon \sim \mathcal{N}(0, \sigma^2_\epsilon)$. Write a program to generate an empirical estimate of the statistical power of a regression of y^* on x^* in a sample of 100 observations. Specifically, write a loop that generates one thousand data sets of 100 observations each using the data-generating process described above; for each data set, record the p-value associated with a test of the hypothesis that $\widehat{\beta} = 0$. For a test size of 0.05, the fraction of observations with $p < 0.05$ provides an empirical estimate of the power of the test.

   (a) Once you've written your loop, use it to identify a value of σ_ϵ that will lead to a power of approximately 0.8 in your 100 observation sample.

   (b) Now modify your loop to compare the results of the ideal regression (described above) to a regression of y^* on $x = x^* + \nu$ where $\nu \sim \mathcal{N}(0, 1)$. How much does attenuation bias reduce statistical power?

(c) Now modify your loop to compare the results of the ideal regression to a regression of $y = y^* + \eta$ on x^* for $\eta \sim \mathcal{N}(0, 1)$. How much measurement error in the dependent variable reduce statistical power?