Homework 6

1. Get the dataset from here (←this is a link). We have two variables: hourly wage (denoted as y) and ASVAB percentile (denoted as x). ASVAB is a standardized test. We use it as a proxy for ability.

(a) Use nonparametric regression to estimate the mean of y conditional on x, i.e., $\hat{E}(y|x)$. For the bandwidth selection, use leave-one-out least squares cross-validation. Plot $\hat{E}(y|x)$ versus x for $0 \leq x \leq 100$.

(b) Comment on the estimated relationship between hourly wage and ASVAB percentile.

(c) (Optional) Redo part (a) and (b) with conditional mean being replaced by conditional median.

Submit your codes and results.

2. Consider a measurement-error model

$$y^* = \beta_0 + \beta_1 x^* + \epsilon$$

$$x = x^* + \eta$$

with $\beta_0 = 1$ and $\beta_1 = 1$. x^*, η and ϵ are independent of each other, and all follow standard normal distribution. Let the sample size be 1000.

(a) If we run regression of y^* on x, calculate the theoretical attenuation bias. Simulate the model 1000 times and calculate the mean of $\hat{\beta}_1$. Is the estimated bias close to the theoretical bias?

(b) Consider a second measurement

$$z = x^* + v$$

with $v \sim N(0, 1)$ being independent of x^*, η and ϵ. Use z as IV for x to estimate $\hat{\beta}_1$. Simulate the model 1000 times and calculate the mean of $\hat{\beta}_1$. Does the IV fix the bias?

Submit your codes and results.