

## Problem Set 4 for Econometrics

due on next lecture

1. Consider the simple regression model

$$y = \beta_0 + \beta_1 x + u,$$

and let  $z$  be a binary instrumental variable for  $x$ . Show that the IV estimator of  $\beta_1$  can be written as

$$\hat{\beta}_1 = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0},$$

where  $\bar{y}_0$  and  $\bar{x}_0$  are the sample averages of  $y_i$  and  $x_i$  over the part of the sample with  $z_i = 0$ , and where  $\bar{y}_1$  and  $\bar{x}_1$  are the sample averages of  $y_i$  and  $x_i$  over the part of the sample with  $z = 1$ . This estimator is known as a grouping estimator.

2. Consider the problem of estimating the effect of cigarette smoking on body weight using the following regression,

$$\log(\text{weight}) = \beta_0 + \beta_1 \text{packs} + \beta_2 \text{height} + u,$$

where *packs* is the number of packs smoked by the mother per day. We might worry that *packs* is correlated with other health factors that affect *weight*, so that *packs* and  $u$  might be correlated.

- Argue for or against the use of cigarette price in each county as instrument for *packs*.
- Can you think of other alternatives?