

## ECON 626: Empirical Microeconomics

### Using Selection on Observables to Bound Selection on Unobservables

Department of Economics  
University of Maryland  
Fall 2019

1. Create a dataset with 100,000 observations. Generate  $n1$ ,  $n2$ ,  $n3$  as iid  $N(0, 1)$ . Generate  $W$  so that it is normally distributed with mean zero and variance 10 (i.e. the standard deviation is the square root of 10) based on  $n1$ . Generate  $C$  so that it is normally distributed with mean zero and variance 0.1 based on  $n2$ . Generate  $X$  so that it is  $0.02 * W + 0.02 * C + n3$ . Generate  $Y$  so that it is simply  $W + C$  (and there is no direct effect of  $X$  on  $Y$ ).
2. Regressions.
  - (a) Regress  $Y$  on  $X$  to see the uncontrolled coefficient and  $R^2$ .
  - (b) Regress  $Y$  on  $X$  and  $C$  to see the controlled coefficient and  $R^2$  for the low-variance control.
  - (c) Regress  $Y$  on  $X$  and  $W$  to see the controlled coefficient and  $R^2$  for the high-variance control.
  - (d) How do your answers compare to Oster (2019) Table 1?
3. To see what ratio of unobserved selection to observed selection would be necessary to explain away the entire (spurious) effect of  $X$  on  $Y$ , type:

```
reg Y X W  
psacalc delta X, rmax(1)
```

Then type the same command immediately after regressing  $Y$  on  $X$  and  $C$ . What values do you get? Do they make sense?