

## ECON 8686 Problem Set #3- Measurement Error

### Part I. Simulation

1) Create a simple measurement error simulation:

```
set obs 100000
gen z=uniform()
gen a=uniform()
gen b=uniform()
gen e1 = invnorm(a)
gen e2 =invnorm(b)
gen y=3*z+e1
gen x=z+e2
```

2) Using the output from the following command:

```
corr z e2, cov
```

Calculate the bias of the coefficient from the regression of y on x (for simplicity, ignore the covariance term).

3) Confirm your answer to question #2 by regressing y on x

4) Now:

```
replace b=.5*b
replace e2=invnorm(b)
replace x=z+e2
```

5) Repeat questions 2 and 3. What happened? Why?

### Part II. Measurement Error and Panel Data

This part of the problem set is based on:

McKinnish, Terra. 2008. "Panel Data Models and Transitory Fluctuations in the Explanatory Variable." *Advances in Econometrics* Vol.21

The data set birth.dta contains data for all 50 states plus D.C. for the years 1972-93. The variables are:

births: birth rate, white women ages 20-24

afdcben: real monthly AFDC Benefit

earn: real earnings per capita

state: state ID code

year

1) Regress the logged birth rate on logged AFDC benefits, logged per capita earnings, and state and year fixed-effects.

2) Estimate a first-differences, 3-year long-differences and 5-year long-differences version of the state fixed-effects model in question #1. Compare and discuss the results.

3) The ideal solution would be to have an instrument, something correlated with the long-term component of AFDC benefits but not the short-term fluctuations. In some literatures, in absence of “external” instruments, researchers will turn to “internal” instruments, usually lagged values of their explanatory variables. Consider the basic differences model:

$$Y_{it} - Y_{it-j} = \beta(X_{it} - X_{it-j}) + (\varepsilon_{it} - \varepsilon_{it-j})$$

If the measurement error is uncorrelated over time, then any value of  $X$  other than  $X_{it}$  and  $X_{it-j}$ , or any function of these values, is a valid instrument for  $X_{it} - X_{it-j}$ .

So, consider the basic first-differences model, and try lagged AFDC benefits,  $\log(\text{afdcben}_{t-2})$ , and the lagged change,  $\log(\text{afdcben}_{t-2}) - \log(\text{afdcben}_{t-3})$ , as an instrument for  $\log(\text{afdcben}_t) - \log(\text{afdcben}_{t-1})$ . Estimate the IV regression, then estimate and evaluate the first stage. Discuss your findings. Explain why we might expect the first-stage F-stat to be so small for these specifications when there is measurement error.

4) BONUS: Gold star and extra credit to anyone that explains why we can't use lagged values of afdc benefits as instruments in the state fixed-effects specification.