Endogeneity Issues in STATA

Carlos Noton *

February 1, 2012

```
clear
capture log close
clear matrix
set mem 90M
capture program drop _all
cd "Choose your proper directory here!"
log using "EXERCISE2", replace
set obs 500
* Just to have the entire class can have the same results fix seed
set seed 666
* Random Terms of the Regression
generate u=10*rnormal(0,1)
* Constant of the Regression
generate X1=1
* Regressor with Large Variation
generate v2=10*rnormal(0,1)
generate v3=10*rnormal(0,1)

generate X2=100+ 50*v2
generate X3=100+ v3

* Creating the respective Outcomes Y (playing nature)
* Real parameters are intercept=8 and slope=3
generate Y =8*X1+2*X2-5*X3+u
sum Y X1 X2

* Do these scatter plot look familiar?
twoway (scatter Y X2, title(Y is outcome - X2 is control)) (lfit Y X2)
graph save "regYX", replace
```

*C.Noton@Warwick.ac.uk

```
twoway (scatter X2 Y, title(X2 is outcome- Y is control)) (lfit X2 Y)
graph save "regXY", replace

**********************
* What is endogenous? *
* Estimated parameters differ!!
**********************
reg Y  X2
reg X2 Y

reg Y  X2 X3
reg X2 Y  X3

****************************
* EXOGENEITY ISSUE *
*******************************
* We start with the instrument Z *
generate z1=10*rnormal(0,1)
generate Z= 100+z1
* Let's create some endogeneity *
generate w=10*rnormal(0,1)
* Problematic Variable E
generate E= 4*Z+20*w

*True Process
generate Y1 = -2+3*E + 50*w
* Regression Ignoring the problem
reg Y1 E, robust
* Using Predicted Values
reg E Z
predict E_hat
reg Y1 E_hat, robust
* IV Regression Solving the problem
ivreg Y1 (E=Z), first robust
*************************************************************
* Same with more Exogenous Variables
generate Y2 = -2+3*E +5*X2 + 50*w
* Regression Ignoring the problem
reg Y2 E X2, robust
* Using Predicted Values without X2 (which is not the best!)
reg E Z, robust
```

```
predict E_hat1
reg Y2 E_hat1 X2, robust

* Using Predicted Values with X2 (which is the best!)
reg E Z X2
predict E_hat2
reg Y2 E_hat2 X2

* IV Regression Solving the problem
ivreg Y2 (E=Z) X2, first robust

log close
```