

Variation and Identification Issues in STATA

Carlos Noton *

January 24, 2012

```
clear
capture log close
clear matrix
set mem 90M
capture program drop _all
cd "C:CHOOSE THE RIGHT PATH"
log using "EXERCISE", replace
set obs 500
set seed 12345
* Random Terms of the Regression
generate u=10*rnormal(0,1)
hist(u)
sum u, detail
graph save "errors", replace
* Constant of the Regression
generate X1=1
generate v=10*rnormal(0,1)
*****
* VARIATION ISSUES!!! *
*****
* Regressor with Little Variation
generate X2a=100+ .03*v
hist(X2a)
graph save "X2a", replace
* Regressor with Large Variation
generate X2b=100+ 3*v
hist(X2b)
graph save "X2b", replace
sum X2a X2b
```

*C.Noton@Warwick.ac.uk

```

* Creating the respective Outcomes Y (playing nature)
* Real parameters are intercept=8 and slope=3
generate Ya =8*X1+3*X2a+u
generate Yb =8*X1+3*X2b+u
sum X2a Ya X2b Yb
* Do these scatter plot look familiar?
tway (scatter Ya X2a, title(Little Variation!)) (lfit Ya X2a)
graph save "Little", replace
tway (scatter Yb X2b, title(Large Variation!)) (lfit Yb X2b)
graph save "Large", replace

* Formal Estimations with little and large variation
* Estimated parameters differ!!
reg Ya X1 X2a
reg Ya X1 X2a, noco
reg Yb X1 X2b, noco

*****
* IDENTIFICATION ISSUES!!! *
*****
generate w=10*rnormal(0,1)

generate Z1 = 100 + 30*v
generate Z2a = -5*Z1 + .1*w
generate Z2b = -5*(100 +30*w)
sum Z1 Z2a Z2b
* Real parameters are intercept=0 and slope Z1=5 , slope Z2=3
generate Yc = 5*Z1 +3*Z2a+u
generate Yd = 5*Z1 +3*Z2b+u

tway (scatter Z1 Z2a, title(Bad Identification)) (lfit Z1 Z2a)
graph save "BadId", replace
tway (scatter Z1 Z2b, title(Good Identification)) (lfit Z1 Z2b)
graph save "GoodId", replace

* Formal Estimations with different degree of Identification
* Estimated parameters differ!!
reg Yc Z1 Z2a
reg Yd Z1 Z2b
** Better Identification gives us better estimates!
log close

```