

## Computer Exercise 1: Programming the OLS Estimator

1. Given  $y = b_0 + b_1X + u$

and

$$y = \begin{bmatrix} 3 \\ 1 \\ 8 \\ 3 \\ 5 \end{bmatrix} \quad X = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$$

use the stata matrix commands to program the OLS matrix formula to calculate the values of  $b_0$  and  $b_1$

[ check your answers by regressing  $y$  on  $x$  using the data set `check1.dta`, which can be downloaded from the course web site ]

```
. matrix y=(3\1\8\3\5)

.
. matrix list y

y[5,1]
      c1
r1    3
r2    1
r3    8
r4    3
r5    5

. matrix X=(1,2\1,3\1,1\1,2\1,2)

. matrix list X

X[5,2]
      c1  c2
r1    1   2
r2    1   3
r3    1   1
r4    1   2
r5    1   2

. matrix Xy=X'*y

. matrix list Xy

Xy[2,1]
      c1
c1    20
c2    33
```

```

. matrix XpX=X'*X

. matrix list XpX

symmetric XpX[2,2]
      c1  c2
c1    5
c2   10  22

. matrix b=syminv(XpX)*Xy

. matrix list b

b[2,1]
      c1
c1    11
c2   -3.5

.       reg y x

      Source      SS      df      MS      Number of obs      =      5
      Model      24.5      1      24.5      F( 1, 3)           =      21.00
      Residual    3.5      3      1.16666667      Prob > F           =      0.0195
      Total 28      4              7      R-squared           =      0.8750
      Adj R-squared =      0.8333
      Root MSE    =      1.0801

      Coef. Std. Err.      t    P>t    [95% Conf. Interval]
      x      -3.5    .7637626    -4.58  0.020   -5.930634   -1.069366
      _cons   11     1.602082     6.87  0.006    5.90146    16.09854

```

(Note that Stata puts the estimate of the constant last)

## 2. Using the data set *gdp.dta* (which you can download from the course web site)

regress the log of real gdp on a time trend (the coefficient will give you the annual % growth rate of real gdp – Why?)

Now see if you can replicate the OLS coefficient estimates using Stata's matrix algebra commands

[To do this you will have to 1<sup>st</sup> create a column vector of ones which will be the constant term and append this to the time trend variable. To do this type  
`matrix i=J(56,1,1)`

(this creates a 56\*1 column vector of ones )  
and join it to the X matrix containing the data on the time trend ]

```

. u gdp

. mkmat time, mat(X)

. matrix i=J(56,1,1)

. matrix X=(X,i)

. matrix list X

. matrix XpX=X'*X

. matrix list XpX

symmetric XpX[2,2]
      time      c1
time  60116
c1    1596      56

. mkmat lrgdp,mat(y)

. matrix list y

. matrix Xy=X'*y

. matrix b=syminv(XpX)*Xy

. matrix list b

```

```

b[2,1]
      lrgdp
time  .02523257
c1    4.7610699

```

```

. reg lrgdp time

```

Source	SS	df	MS			
Model	9.31466539	1	9.31466539	Number of obs =	56	
Residual	.061749869	54	.001143516	F( 1, 54) =	8145.64	
Total	9.37641526	55	.170480277	Prob > F =	0.0000	
				R-squared =	0.9934	
				Adj R-squared =	0.9933	
				Root MSE =	.03382	

  

lrgdp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
time	.0252326	.0002796	90.25	0.000	.0246721	.0257931
_cons	4.76107	.0091601	519.76	0.000	4.742705	4.779435