University of Essex
2015-16

Department of Government
TA: Stavros Poupakis

GV903: Advanced Research Methods

Lab 5: Finite and Large Sample Properties of Least Squares

1. Generate $x_i \sim N(10, 9)$ (that is mean 10 and variance 9) and $\epsilon \sim N(0, 1)$, then generate $y_i = 2 + 3x_i + \epsilon_i$.

First run a regression in Stata. Then "put" x and y in Mata, using `putmata x y`. Now we will enter mata write a few lines of code and then end the Mata session. That is

```
mata

cons = J(rows(x),1,1)
X = (cons, x)

b_hat = invsym(X'*X) * (X'*y)
b_hat

end
```

Run the above code all together and see what you get. You just calculated the OLS estimator on your own! Without using any regression function, just linear algebra.

2. A friend comes to you excited and tells you that he heard that apart from OLS there is also another estimator, called Least Absolute Deviations (LAD). And that this could work too to find the estimators of the parameters! In fact he argues that this is a "better" estimator than OLS, and is also available in Stata through `qreg` **Think what he could mean by "best". Remember BLUE from lecture?**

Justify yourself which one is better. Your are researchers and you will probably face problems of using different estimators and that are brand new and you should be able to just on your own which one to use. **Run a Monte Carlo experiment to decide and discuss why.**

3. Multivariate distributions
Generate two variables drawing from a 2-dimensional multivariate normal distribution with the following parameters:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 4 & 2 \\ 2 & 9 \end{pmatrix}\right]$$

Note this is a vector of means and a variance-covariance matrix. We will do that using the following code:

```
matrix mu = (0, 2)
matrix VC = (4, 2 \ 2, 9)

set obs 1000
drawnorm x1 x2 , means(mu) cov(VC)
```

Instead of using the variance-covariance matrix you could use the Correlation matrix and the standard deviations of each variable. Just remember that

$$Corr(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)}\sqrt{Var(y)}}$$

and of course that $Corr(x, x) = 1$. Re-do now the previous step by modifying it with this

```
matrix COR = (1, 0.33 \ 0.33, 1)
drawnorm x1 x2 , means(mu) corr(COR) sds(2, 3)
```

4. Now create $y_i = 2 + 3x_i + \epsilon_i$ but draw x_i and ϵ_i from a multivariate normal as

$$\begin{pmatrix} x_i \\ \epsilon_i \end{pmatrix} \sim N\left[\begin{pmatrix} 10 \\ 0 \end{pmatrix}, \begin{pmatrix} 36 & 14.4 \\ 14.4 & 16 \end{pmatrix}\right]$$

Run a MC experiment to see what happens in this case. Is the OLS estimator unbiased? Which assumption is violated?