

# 14.271 Handout 3 Basic Econometrics and Stata

Andrew Sweeting

October 5, 2001

This handout introduces you to the most basic principles of econometrics and how to use Stata (useful for PS 4).

## 1 Regression

Start with a belief about a data generating process (DGP) which generates our data. For simplicity imagine that the DGP is linear

$$y = X\beta + \epsilon$$

We want to learn about the DGP and observe  $y$  and the set of variables  $X$ . We therefore estimate a model

$$y = X\hat{\beta} + \epsilon$$

The key condition under which  $E(\hat{\beta}) = \beta$  (unbiasedness) when we use OLS ( $\hat{\beta} = (X'X)^{-1}X'y$ ) is that  $E(\epsilon/X)=0$ . We can also derive standard errors (an estimate of our confidence in estimating  $\beta$ ) by using  $\hat{\sigma}^2(X'X)^{-1}$ . If the errors are normal we also know that (even in small samples) that  $\hat{\beta}$  will be distributed normally.

## 2 Instruments

In IO we typically want to estimate demand and supply.

$$q^D = a - bp + \epsilon^D$$

$$q^S = c + dp + \epsilon^S$$

Unless we have perfect competition and constant marginal cost price and the demand error will tend to be correlated. Hence our estimate of  $b$  will be biased. We solve this using instruments i.e. things which will be correlated with movements in price but not with the demand error. A similar principle applies to estimating the supply curve. A key is to identify appropriate instruments:

- exogenous demand shifters (for supply)
- exogenous cost shifters
- product characteristics (e.g. BLP)
- prices in other cities (e.g. Hausman)

Testing instruments

### 3 Conduct Parameter Estimation ( $\theta$ )

This is used to estimate whether an industry appears to be behaving collusively, competitively or somewhere in between, although it has now gone out-of-fashion (Corts Critique, Bresnahan)..

### 4 Stata

- The Stata exercises requires you to estimate basic demand and supply equations. Log your results using "log using h:\porter.log". *here I list the most basic commands – the manual sand on – line help should be used to do anything more complicated.*

#### 4.1 Reading in data

There will be an ASCII file on the website. To read this in use "infile week quantity price lakes coln dm1 dm2 dm3 dm4 seas1 seas2 seas3 seas4 seas5 seas6 seas7

seas8 seas9 seas10 seas11 seas12 seas13 using porter.prn" which will also name the variables for you.

## 4.2 Looking at and generating data

- To produce summary statistics type "sum" or "sum, detail"
- Generate variables using "gen" e.g. "gen logq=log(q)" [there is a shortcut to do this for all variables using the "for" command]
- Save results using "save using h:\portsave.dta, replace" *wherereplacewilloverwriteafileofthesamename*

## 4.3 Regression

- do OLS regression using "regress y x1 x2 x3". This automatically adds a constant.
- do IV regression using "ivreg y x1 (x2 x3=instr1 instr2 instr3) x4 x5". Note that when you interact an endogenous variable with another variable you will need to find another instrument.