

Econometric Analysis Using *Stata*

Introduction

Time Series

Panel Data

Stata: Data Analysis and Statistical Software

- Windows Interface
 - Windows
 - Command, Results, ...
 - Pulldown Menu
 - File, Edit, Data, ..., Help
- Operation
 - Interactive Mode
 - Command
 - Menu
 - Batch Mode
 - do files
- Working Directory

Where is the Data?

- Stata Dataset
 - Datasets installed with Stata
 - Datasets from Stata Web site
 - Datasets from courses using Stata
- Text Data
 - Type your own data
 - Import from Excel, Notebook, GAUSS, ...

Stata Dataset

sp500.dta

- Interactive mode
- Data Conversion: Dataset to Text
- Data Analysis
- Graphics

Text Data

gasoline.txt

- Batch Mode: do files
- Data Conversion: Text to Dataset
- Data Analysis
- Graphics

demo1.do

```
* Demo of do file #1
clear
sysuse sp500
describe
summarize
generate volatility=high-low
generate lnvolume=ln(volume)
generate time=_n
graph twoway line close time
more
graph twoway scatter change volatility
correlate change volatility
regress change volatility lnvolume
generate changel=change[_n-1]
regress change changel volatility
lnvolume
```

demo2.do

```
* Demo of do file #2
clear
infile year gasexp pop gasp income pnc puc ppt pd pn ps   ///
using "c:\course09\ec570\data\gasoline.txt"
label data "Greene [2008], Table F2.2: The U.S. Gasoline Market"
label variable year "Year, 1953-2004"
label variable gasexp "Total U.S. gasoline expenditure"
label variable pop "U.S. total population in thousands"
label variable gasp "Price index for gasoline"
label variable income "Per capita disposable income"
label variable pnc "Price index for new cars"
label variable puc "Price index for used cars"
label variable ppt "Price index for public transportation"
label variable pd "Aggregate price index for consumer durables"
label variable pn "Aggregate price index for consumer nondurables"
label variable ps "Aggregate price index for consumer services"
describe
summarize
* save it as a Stata dataset
save "c:\course09\ec570\data\gasoline", replace
/*
generate g=ln(gasexp/pop) // log-per-capita gas consumption
generate y=ln(income/pop) // log-per-capita income
generate pg=ln(gasp) // log price of gas
generate pnew=ln(pnc) // log price of new cars
generate pused=ln(puc) // log price of used cars
regress g y pg pnew pused
*/
```

demo3.do

```
* Demo of do file #3
set more off
clear
//use "c:\course09\ec570\data\gasoline"
use http://www.econ.pdx.edu/faculty/KPL/ec570/data/gasoline
describe
summarize
generate g=ln(gasexp/pop) // log-per-capita gas consumption
generate y=ln(income/pop) // log-per-capita income
generate pg=ln(gasp) // log price of gas
generate pnew=ln(pnc) // log price of new cars
generate pused=ln(puc) // log price of used cars
regress g y pg pnew pused
// dynamic model with lagged variables
generate g1=g[_n-1]
regress g y pg pnew pused g1
// hypothesis testing
test pnew pused
test pnew+pused=0
// prediction
predict e, residual
predict ghat, xb
twoway scatter e ghat
// structural break
```


Must-Know Commands

- System
 - clear
 - exit
 - log
 - set
 - # delimit
 - net
 - search
 - help
- Data Management
 - use
 - Infile, infix
 - list
 - describe
 - keep, drop
 - generate, replace, rename
 - save, outfile

Must-Know Commands

- Data Analysis
 - summarize
 - correlate
 - graph
 - twoway, scatter,...
 - hist
- Statistical Analysis
 - regress
 - predict
 - test
 - dwstat
 - hettest

Econometric Analysis Using *Stata*

Introduction

Time Series

Panel Data

Time Series Analysis Using Stata

- Declare time series data and variables
 - tsset
- Time series operators
 - L.; F.; D.; S.
- Commands with time series options
 - regress ..., if tin(.,.)
 - generate
 - summarize

Example: U.S. GDP Growth

gdp2000.txt

gdp2000.dta

- Time series setup
- Time series operators
- Time series line plot (graphics)
- Time series regression

demo_gdp1.do

```
* Read text data file (.asc) and convert it to Stata dataset file (.dta)
infile year quarter gdp pgdp using "c:\course07\ec572\gdp2000.txt"
describe
summarize
label data "U. S. GDP: 1947.1-2006.3"
label variable gdp "GDP (billion of current dollars)"
label variable pgdp "Implicit GDP price deflator (year 2000 = 100)"
* delete the 1st line of variable names
drop in 1
* create a time series dataset
generate time=yq(year,quarter)
tsset time, quarterly
label variable time "Time"
drop year quarter
describe
summarize
* save it as a Stata dataset, if it has not done yet
save "c:\course07\ec572\gdp2000"
```

demo_gdp2.do

```
* use graph to represent the data
* a graph is worth of thousand words
clear
use c:\course07\ec572\gdp2000.dta
* is this time series data?
tsset
d
su
* generate new variable
gen rgdp=100*gdp/pgdp
gen lrgdp=ln(rgdp)
gen gq=100*D.lrgdp
gen ga=100*(lrgdp-L4.lrgdp)
su
* time series line plots
tsline rgdp, name(rgdp)
tsline gq ga, name(growth)

* time regression
reg lrgdp time
```

Econometric Analysis Using *Stata*

Introduction

Time Series

Panel Data

Panel Data Analysis Using Stata

- Declare panel data and variables
 - xtset
- Panel data analysis: xt commands
 - xtodes
 - xtsum
 - xtdata
 - xtline
- Panel data regression
 - xtreg

Example: Returns to Schooling

- **Cornwell and Rupert Data, 595 Individuals, 7 Years**
- These data were analyzed in Cornwell, C. and Rupert, P., "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variable Estimators," *Journal of Applied Econometrics*, 3, 1988, pp. 149-155. Data Source: Panel Study of Income Dynamics.
- Data file: cornwell&rupert.txt

Example: Returns to Schooling

- LWAGE = log of wage
- EXP = work experience
- WKS = weeks worked
- OCC = occupation, 1 if blue collar,
- IND = 1 if manufacturing industry
- SOUTH = 1 if resides in south
- SMSA = 1 if resides in a city (SMSA)
- MS = 1 if married
- FEM = 1 if female
- UNION = 1 if wage set by union contract
- ED = years of education
- BLK = 1 if individual is black

Example: Returns to Schooling

- The Model

$$\begin{aligned} LWAGE_{it} = & \beta_0 + \beta_1 EXP_{it} + \beta_2 EXP_{it}^2 + \beta_3 WKS_{it} + \beta_4 OCC_{it} \\ & + \beta_5 IND_{it} + \beta_6 SOUTH + \beta_7 SMSA + \beta_8 MS_{it} + \beta_9 UNION_{it} \\ & + \beta_{10} ED_i + \beta_{11} FEM_i + \beta_{12} BLK_i + \varepsilon_{it} \end{aligned}$$

```

/*
** Panel Data (Cornwell and Rupert, 1988)
** Greene [2008], Chap. 9
** Data is stacked in long form, 595 individuals 7 years
*/
clear
set more off
infile exp wks occ ind south smsa ms fem union ed blk lwage ///
using "c:\course09\ec510\data\cornwell&rupert.txt"
drop in 1
describe
summarize
generate person=group(595)
bysort person: generate period=group(7)
* panel data definition
xtset person period
xtdes
xtsum
* one-way tabulation of data
xttab union
xttab ind
xttrans ms
xttab ed // ed is time invariant
* plots of panel data
xtline lwage if person<=10, overlay
generate exp2=exp^2
local x1 exp exp2 wks occ ind south smsa ms union
local x2 ed blk fem
* panel data regression: y=lwage
* x1=[1 exp exp2 wks occ ind south smsa ms union],
* x2=[ed blk fem] (time-invariant regressors)
regress lwage `x1' `x2'
regress lwage `x1' `x2', vce(cluster person)

```