LAB1 - Exploring OLS Estimates in TSCS

—First, we need to point Stata to the directory where we've put some useful add-on (ADO) command files. Try first:

```
. help adopath
. adopath + T:\CCB Released Files\Statistics Workshop Sep 2009\ado\plus/
. adopath + "T:\CCB Released Files\Statistics Workshop Sep 2009\ado\plus/"
```

—Let's check if we have proper access to these files now:

```
. help simqi
```

—If both above fail, let's try:

```
sysdir set SITE "T:\CCB Released Files\Statistics Workshop Sep 2009\ado\plus"
```

—Exceptionally useful starting point for exploring Stata's explicitly "xt" (panel/tscs) designed features:

```
. help xt
```

—Also useful are:

```
. help xi
```

—These are data from Garrett & Mitchell: "Globalization, government spending and taxation in the OECD," *European Journal of Political Research* 39(2) (2001): 145–177. Explore variable list.

```
. use "H:\…[PATH]…\garmit_esspanel1.dta", clear
```

—Can also use pull-down menu, or can also download from web like so:

```
. use "http://www-personal.umich.edu/~franzese/ICPSR/garmit_esspanel1.dta"
```

—A good idea to get in the habit of logging your work every session. You can always delete records later if you wish, but you cannot go back to create a log as easily from within a session and not at all once you've closed one.

```
. log using "H:\…[PATH]…\basics1.log", append
```

—This dataset has already been set-up in lots of ways, but let's review that process a bit:

```
. help xtset
. xtset
      panel variable:  cc (strongly balanced)
       time variable:  year, 1961 to 1994
               delta:  1 unit
```

The 18 CC's & Country's are:

1. Australia
2. Austria
3. Belgium
4. Canada
5. Denmark
6. Finland
7. France
8. Germany
9. Ireland
10. Italy
11. Japan
12. Netherlands
13. New Zealand
14. Norway
15. Sweden
16. Switzerland
17. United Kingdom
18. United States

—All right, let's estimate a model already!

```
. reg spend unem growthpc depratio left cdem trade lowwage fdi
      Source |       SS       df       MS              Number of obs =     529
-------------+------------------------------           F(  8,   520) =   91.12
       Model |  33411.5828     8  4176.44785            Prob > F      =  0.0000
    Residual |  23834.1148   520  45.8348362            R-squared     =  0.5837
-------------+------------------------------           Adj R-squared =  0.5772
       Total |  57245.6976   528  108.419882            Root MSE      =  6.7701

------------------------------------------------------------------------------
       spend |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        unem |   .9320547   .0928529    10.04   0.000     .7496417    1.114468
    growthpc |  -.9012817   .1303971    -6.91   0.000    -1.157451   -.6451119
    depratio |  -.4446718   .1235697    -3.60   0.000    -.6874289   -.2019147
        left |    .058016   .0087215     6.65   0.000     .0408823    .0751497
        cdem |   .0287597   .0131838     2.18   0.030     .0028596    .0546597
       trade |   .1293114    .014866     8.70   0.000     .1001066    .1585163
     lowwage |   -.157155   .0478007    -3.29   0.001    -.2510612   -.0632488
         fdi |   .0927398   .2159476     0.43   0.668    -.3314971    .5169768
       _cons |   48.81442   4.943615     9.87   0.000      39.1025    58.52633
------------------------------------------------------------------------------
```

—One low-tech, but sometimes quite useful post-estimation strategy is simple graphical exploration of your estimates, your estimated residuals in particular. For Stata's many built-in post-estimation tools for graphical & statistical exploration, see menu item: Statistics/Linear models and related/Regression diagnostics… Also see:

```
. help regress postestimation
```

And:

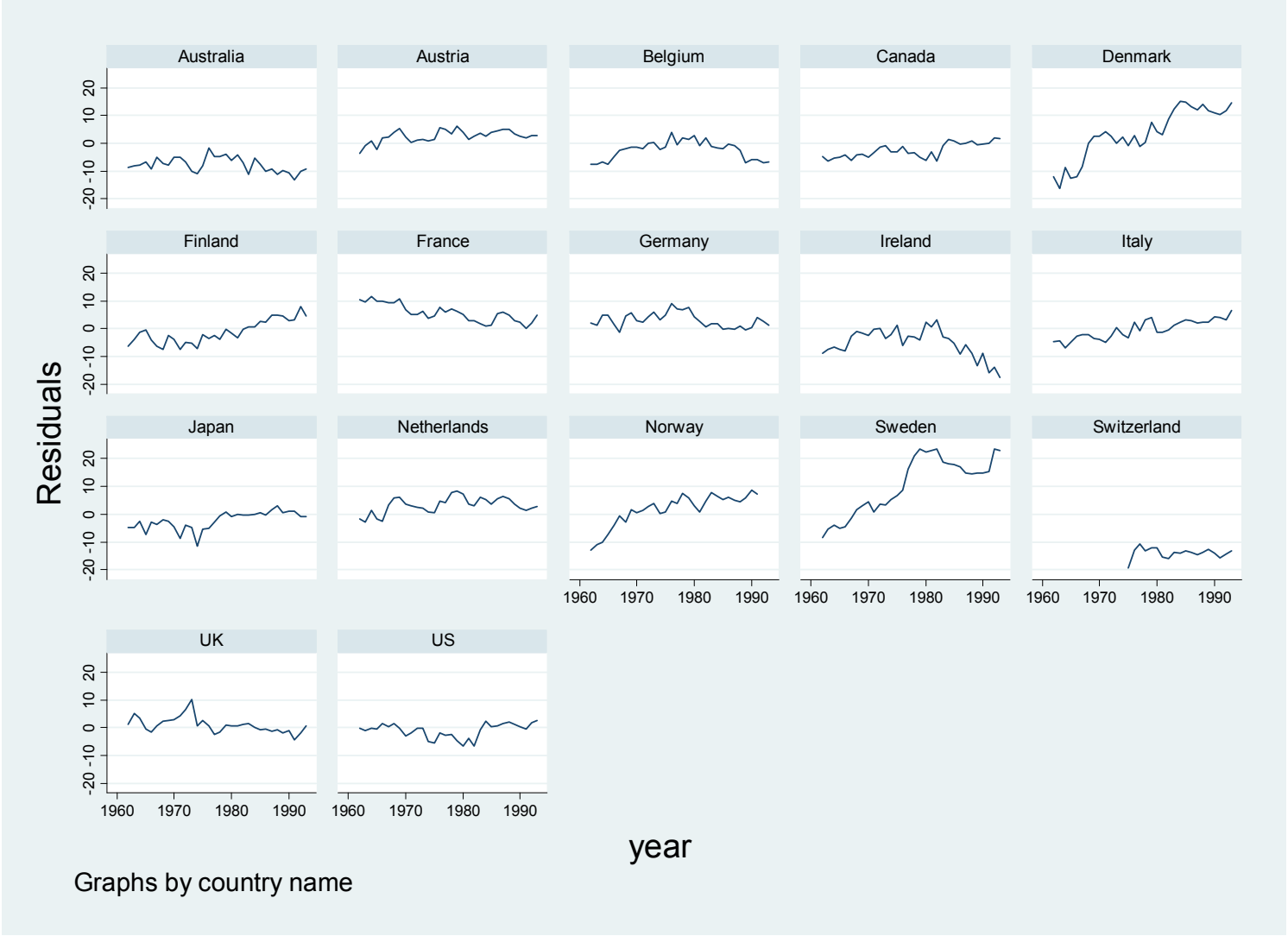
```
. help regress postestimation ts
```

You may also wish to create graphics related to your estimated residuals yourself. Useful for this are to create a variable like the following:

```
. predict e_FullPooledOLS, residuals
(83 missing values generated)
```

A simple line-plot of the residuals by country is now available thus:

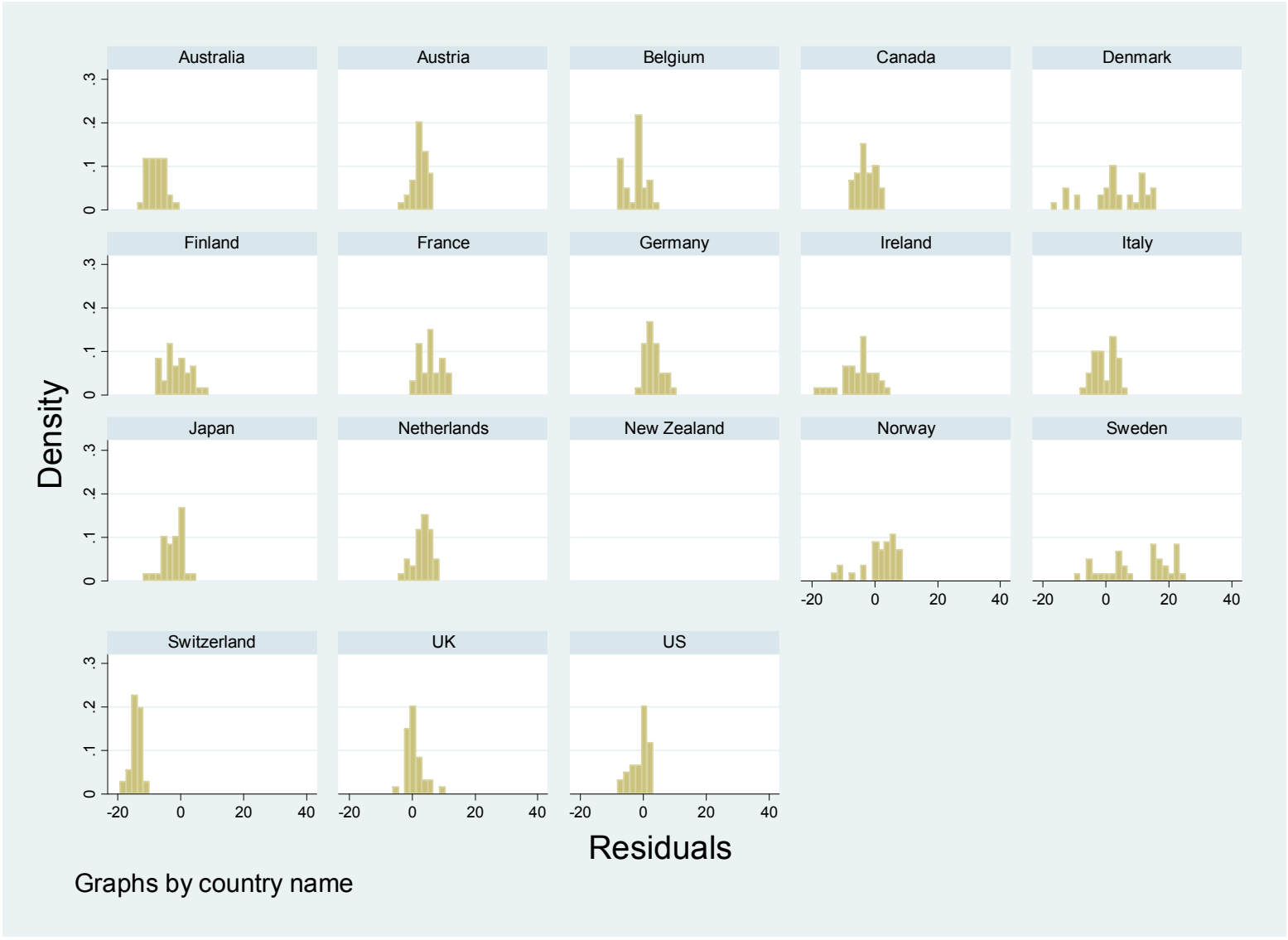
```
. xtline e_FullPooledOLS, i(country) t(year)
```



What patterns should you see in the means, trends, and spreads within and across units and over time? What's wrong with these pictures?

Another useful plot for exploring residuals are histograms.

```
. histogram e_FullPooledOLS, by(cc)
```



What should we see here?

A similarly useful plot would be a box-and-whiskers by country.

Another useful variable to generate for post-estimation graphical and other exploration is this:

```
. gen e2_FullPooledOLS=e_FullPooledOLS^2
(83 missing values generated)
```

What is this an estimate of?

How might you use e and/or e^2 to test for
…insufficiently modeled (orthogonal) country-specific factors? (i.e., the sufficiency of a common intercept?)
…heteroskedasticity that would bias OLS estimates of coefficient-estimate variance-covariance matrices?
…heteroskedasticity by country?
…(time-)serial correlation?

```
. help estat ovtest
```

Claims to be a test for omitted variables. What is it? What does it actually test?

```
. help estat hettest
```

HOMEWORK (?): Conduct some similar post-estimation graphical and statistical explorations with your dataset and bring a log or lab report.

Try this series of estimates:

```
reg spend unem growthpc depratio left cdem trade lowwage fdi
reg spend unem growthpc depratio left cdem trade lowwage fdi , vce(robust)
reg spend unem growthpc depratio left cdem trade lowwage fdi , vce(cluster cc)
```

What are these? What happens to your estimates of **β**? What about of V(**b**)?

Try this series of estimates:

```
xtpcse spend unem growthpc depratio left cdem trade lowwage fdi , independent
xtpcse spend unem growthpc depratio left cdem trade lowwage fdi , hetonly
xtpcse spend unem growthpc depratio left cdem trade lowwage fdi
```

What are these? What happens to your estimates of **β**? What about of V(**b**)?

One last set of estimates:

```
xtgls spend unem growthpc depratio left cdem trade lowwage fdi, p( [i,h,c]) c( [i,a,p])
```

What are these? What happens to your estimates of **β**? What about of V(**b**)?

SOME CRUCIAL REGRESSION INTUITIONS

use "**H:\…[PATH]…**\debt_699.dta", clear

OMITTED VARIABLE BIAS, three true regressors, 2 included, 1 omitted:
True: reg debtx growth ue enop if debtx~=.
OV Biased: reg debtx growth ue if debtx~=.
Omitted on Included: reg enop growth ue if debtx~=.
OVBiasedCoeff=ExclTrue*ExclOnIncl+TrueCoeff (g):
display -1.014774*-.0211177-.4737356
OVBiasedCoeff=ExclTrue*ExclOnIncl+TrueCoeff (u):
display -1.014774*-.0404972+2.019045

OMITTED VARIABLE BIAS, four true regressors, 2 included, 2 omitted:
True: reg debtx growth ue enop cog if debtx~=.
OV Biased: reg debtx growth ue if debtx~=.
Omitted on Included 1:
reg enop growth ue if debtx~=.
Omitted on Included 2:
reg cog growth ue if debtx~=.

$$\begin{bmatrix} \hat{\beta}^{OVB}_{grow} \\ \hat{\beta}^{OVB}_{unem} \end{bmatrix} = \begin{bmatrix} \hat{\beta}^{true}_{grow} \\ \hat{\beta}^{true}_{unem} \end{bmatrix} + \begin{bmatrix} \hat{\beta}^{enop}_{grow} & \hat{\beta}^{cog}_{grow} \\ \hat{\beta}^{enop}_{unem} & \hat{\beta}^{cog}_{unem} \end{bmatrix} \begin{bmatrix} \hat{\beta}^{true}_{enop} \\ \hat{\beta}^{true}_{cog} \end{bmatrix}$$

Reg's @ Om on all Inc

Recall formula to above-right:

$$\hat{\beta}^{OVB}_{grow} = \hat{\beta}^{true}_{grow} + \hat{\beta}^{enop}_{grow}\hat{\beta}^{true}_{enop} + \hat{\beta}^{cog}_{grow}\hat{\beta}^{true}_{cog}$$

display -.5395289-.0211177*-1.03081+.1042465*.6278835

$$\hat{\beta}^{OVB}_{unem} = \hat{\beta}^{true}_{unem} + \hat{\beta}^{enop}_{unem}\hat{\beta}^{true}_{enop} + \hat{\beta}^{cog}_{unem}\hat{\beta}^{true}_{cog}$$

display 2.030808-.0404972*-1.03081-.0197694*.6278835

MEASUREMENT ERROR, bivariate case:

TRUTH: $y = x^{*}\beta + \varepsilon$

ESTIMATE: $y = x\hat{b} + u$, with $x = x^{*} + \omega$, ω = white noise

```
gen xstar=rnormal()
gen omega=rnormal()
gen x=xstar+omega
gen y=xstar+rnormal()
reg y xstar
reg y x
```

MEASUREMENT ERROR, trivariate case:

TRUTH: $y = \alpha x^{*} + z\gamma + \varepsilon$

ESTIMATE: $y = \hat{a}x + \hat{g}z + u$, with $x = x^{*} + \omega$, ω = white noise

Case with z & x uncorrelated:

```
drop y
gen z=2*rnormal()
gen y=xstar+z+rnormal()
reg y xstar z
reg y x z
```

Case with z & x correlated:

```
drop y z
gen z=xstar+2*rnormal()
gen y=xstar+z+rnormal()
reg y xstar z
reg y x z
```

SIMULTANEITY BIAS, 2x2 case:

$$\text{TRUTH:}\quad \begin{aligned} y &= .5x + z + \varepsilon_y \\ x &= .5y + w + \varepsilon_x \end{aligned}$$

```
drop x y z
gen err_y=rnormal()
gen err_x=rnormal()
gen z=rnormal()
gen w=rnormal()
```

$$\text{TRUTH:}\quad \begin{aligned} y &= .5x + z + \varepsilon_y \\ x &= .5y + w + \varepsilon_x \end{aligned} \Rightarrow \begin{aligned} y &= .5(.5y + w + \varepsilon_x) + z + \varepsilon_y = (1-.25)^{-1}\left[.5(w + \varepsilon_x) + z + \varepsilon\right. \\ x &= .5(.5x + z + \varepsilon_y) + w + \varepsilon_x = (1-.25)^{-1}\left[.5(z + \varepsilon_y) + w + \varepsilon\right. \end{aligned}$$

```
gen y=(1/(1-.25))*(.5*w+.5*err_x+z+err_y)
gen x=(1/(1-.25))*(.5*z+.5*err_y+w+err_x)
reg y x z
reg x y w
```

Knowing how to generate some simulated data with certain properties to explore how estimators etc. work a key item in your tool-box.

More Explorations of Typical Heterogeneity in TSCS Data

Generalized (Normal) Linear-Regression Model:

```
. reg debtx growth ue enop cog if debtx~=.
. predict LSresids , res
. reg LSresids us ja ge fr it uk ca au be de fi gr ir ne no po sp sw sz al nz
```

(Lots of) Unit Heterogeneity; Try Fixed Effects:

```
. reg debtx us ja ge fr it uk ca au be de fi gr ir ne no po sp sw sz al nz
growth ue enop cog
. drop LSresids
. predict LSresids, res
. gen LSresids2=LSresids^2
. reg LSresids2 us ja ge fr it uk ca au be de fi gr ir ne no po sp sw sz al nz
```

Still lot of panel heteroskedasticity; return primary model to memory:

```
. reg debtx us ja ge fr it uk ca au be de fi gr ir ne no po sp sw sz al nz
growth ue enop cog
```

Stata's Battery of Postestimation Commands:

```
. help regress postestimation
    dfbeta             DFBETA influence statistics
    estat hettest      tests for heteroskedasticity
    estat imtest       information matrix test
    estat ovtest       Ramsey reg specification-error test for omitted vars
    estat szroeter     Szroeter's rank test for heteroskedasticity
    estat vif          variance inflation factors for the independent variables
    acprplot           augmented component-plus-residual plot
    avplot             added-variable plot
    avplots            all added-variable plots in one image
    cprplot            component-plus-residual plot
    lvr2plot           leverage-versus-squared-residual plot
    rvfplot            residual-versus-fitted plot
    rvpplot            residual-versus-predictor plot
```

Cook-Weisberg:

```
. estat hettest
```

Augmented Cook-Weisberg:

```
. estat hettest, rhs
```

White's General test:

```
. help whitetst [install if needed]
. whitetst

. reg debtx us ja ge fr it uk ca au be de fi gr ir ne no po sp sw sz al nz
growth ue enop cog, vce(robust)

xtset ctry year
xtgls debtx us ja ge fr it uk ca au be de fi gr ir ne no po sp sw sz al nz
growth ue enop cog, p(h)
```

Exploiting Stata's xi for TSCS Post-Estimation Exploration

Compare the following:

```
. reg debtx us ja ge fr it uk ca au be de fi gr ir ne no po sp sw sz al nz
growth ue enop cog

. xi: reg debtx i.ctry growth ue enop cog
```

Now try this:

```
. xi: reg debtx i.ctry*growth i.ctry*ue i.ctry*enop i.ctry*cog
```

It works, but is very messy...
You could then do, e.g. [WARNING: Messy, time-consuming...]:

```
. avplots   or  . avplot growth or  . acprplot growth , low
```