Econometrics 1

Lecture 17

Pooling Cross Section and Time Series

Need For Pooling Time Series and Cross Section data

- Many economic issues require cause effect analyses of cross-sections of individuals, households or countries over time.

- The major issue that economists like to know remains whether coefficients vary

  across observations in the same time period or

  whether variables have any systematic pattern over time.

Need For Pooling Time Series and Cross Section data

- For instance macroeconomist are interested to know

What makes growth rates differ across countries at a particular year and of the same country over time.

In other words they want to find out whether there are any country specific and time specific effects on economic growth?

Panel studies of growth studies are carried out to know the determinants of growth of an individual country or a group of countries over time.

How much of fluctuations in economic activities is explained by country specific or time specific factors?

- Microeconomic studies aim to investigate whether profits vary systematically by firms and by production periods.

Two Methods of combining Time Series and Cross Section: Panel Model and Pooling Regression

- When confronted with these questions an appropriate econometric method requires using all observations across individuals for each time period under investigation.

- two major methods in econometrics

  1. The paned data method

  2. pooling regression techniques that

- In Panel data models, major emphasis lies on decomposing total variation within a group and between the groups.

$$y_{i,t} = \beta_{0,i} + \beta_{1,i} X_{i,t-1} + e_{i,t} \qquad (1)$$

$$y_{i,t} = \beta_{0,t} + \beta_{1,t} X_{i,t-1} + e_{i,t}$$

Subscript i, t refer to individual and time period respectively.

Total Effect (in deviation form):

$$\beta_{1,OLS} = \frac{\sum_i \sum_t \left(X_{i,t} - \bar{X}\right)\left(Y_{i,t} - \bar{Y}\right)}{\sum_i \sum_t \left(X_{i,t} - \bar{X}\right)^2} = \frac{t_{xy}}{t_{xx}}$$

(2)

$$t_{xy} = \sum_i \sum_t \left(X_{i,t} - \bar{X}\right)\left(Y_{i,t} - \bar{Y}\right)$$

$$= \sum_i \sum_t \left(X_{i,t} - \bar{X}_i + \bar{X}_i - \bar{X}\right)\left(Y_{i,t} - \bar{Y}_i + \bar{Y}_i - \bar{Y}\right)$$

$$= \sum_i \sum_t \left(\left(X_{i,t} - \bar{X}_i\right)\left(Y_{i,t} - \bar{Y}_i\right)\right) + T \sum_i \left(\bar{X}_i - \bar{X}\right)\left(\bar{Y}_i - \bar{Y}\right) =$$

$$W_{xy} + b_{xy}$$

(3)

A Glimpse of Panel Method

Between group effect

$$\hat{\beta}_b = \frac{\sum_t \left(X_{i,t} - \bar{X}\right)\left(Y_{i,t} - \bar{Y}\right)}{\sum_t \left(X_{i,t} - \bar{X}\right)^2} = \frac{b_{xy}}{b_{xx}} \qquad (4)$$

Within group effect

$$\hat{\beta}_W = \frac{\sum_i \left(X_{i,t} - \bar{X}_t\right)\left(Y_{i,t} - \bar{Y}_t\right)}{\sum_i \left(X_{i,t} - \bar{X}_t\right)^2} = \frac{W_{xy}}{W_{xx}} \qquad (5)$$

$$t_{xx}\beta_{OLS} = t_{xy} = W_{xy} + b_{xy}$$

$$= \hat{\beta}_W \frac{W_{xx}}{W_{xx} + b_{xx}} + \hat{\beta}_W \frac{b_{xx}}{W_{xx} + b_{xx}} \qquad (6)$$

Three different methods of pooling observations with time series and cross section dimension

a. A dummy variable method

b. The SURE (systematically unrelated regression equations) method

c. Error component method

Dummy Variable Method

Take a pooled model for growth rate as a function of inflation of i-different countries over t periods.

$$g_{i,t} = \alpha_{0,i} + \alpha_{1,i} \pi_{i,t-1} + e_{i,t} \quad (7)$$

where $g_{i,t}$ is the growth rate in country i, $\pi_{i,t}$ is the inflation rate in country i, and $e_{i,t}$ is the independently and identically distributed random error term.

Error term is normally distributed with mean zero, $e_{i,t} \sim N\left(0, \sigma_i^2\right)$, but the variance may be different from one country to another.

Introducing a dummy variable is the simplest method of isolating individual or time specific effect in a regression model.

$$g_{i,t} = \alpha_{1,1} D_{1,i} + \alpha_{1,2} D_{2,i} + .. + \alpha_{1,m} D_{m,i} + \alpha_{1,i} \pi_{i,t-1} + e_{i,t} \quad (8)$$

Where $D_{1,i} = \begin{cases} 1 = if \quad observation\, i = 1 \\ 0 \quad otherwise \end{cases}$

The individual effects are picked up by the dummy variable $D_{m,i}$.

SURE: An Example from Economic Growth across countries

In an interdependent world the growth rate in one country is affected by the growth rate in another country. In two country case

$$g_{1,t} = \alpha_{0,1} + \alpha_{1,1}\pi_{1,t-1} + e_{1,t} \quad (9)$$

$$g_{2,t} = \alpha_{0,2} + \alpha_{1,2}\pi_{2,t-1} + e_{2,t} \quad (10)$$

Since international economic situation affects the demand for products at home country, any shock occurring in country 2, $e_{2,t}$, affects growth prospects in country 1 $g_{1,t}$.

Similarly negative or positive shocks in country 1, $e_{1,t}$, affect the growth in country 2.

There is **contemporaneous correlation** between error terms. In this sense though growth may depend on internal factors but the external factors influence growth rate of each economy.

Assumption of the Sure Model

- Mean of $e_{1,t}$ is zero for every value of $\pi_{1,t}$, $E\left[e_{1,t}\right] = 0$
- Mean of $e_{2,t}$ is zero for every value of $\pi_{2,t}$, $E\left[e_{2,t}\right] = 0$
- variance of $e_{1,t}$ is constant $\mathrm{var}\left[e_{1,t}\right] = \sigma_1^2$ for every ith observation
- variance of $e_{2,t}$ is constant $\mathrm{var}\left[e_{2,t}\right] = \sigma_2^2$ for every ith observation
- $\mathrm{cov}(e_{1,t}e_{1,t}) = 0 \quad \textit{for all} \quad t \neq s$; this also means there is no autocorrelation
- $\mathrm{cov}(e_{2,t}e_{2,s}) = 0 \quad \textit{for all} \quad t \neq s$; this also means there is no autocorrelation

All of the above assumptions are standard to the OLS assumptions.
The major difference lies on assumption of contemporaneous correlation across the disturbance terms in above two models.

$$\mathrm{cov}(e_{1,t}e_{2,s}) = \sigma_{1,2}^2 \qquad (11)$$

SURE: Stacking Models

Individual regression

$$Y_1 = X_1\beta + \varepsilon_1$$

$$Y_2 = X_2\beta + \varepsilon_2$$

$$Y_m = X_m\beta + \varepsilon_m \qquad (12)$$

There are **m** equations and **T** observations in the SURE system (in growth rate example we have 151 countries and 31 observations). They can be stacked into one large equation system as following.

In matrix notation

$$\begin{bmatrix} Y_1 \\ Y_1 \\ . \\ . \\ Y_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & . & . & 0 \\ 0 & X_2 & 0 & . & 0 \\ 0 & 0 & X_3 & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \\ \\ \beta_m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \\ \\ e_m \end{bmatrix} \qquad (13)$$

Each Y_m and e_m has a dimension of T by 1 and X_m has T by K dimension and each β_m has K by 1 dimension. The covariance matrix of errors has TM by TM dimension.

Variance-Covariance Structure of Errors in a Sure Model

$$ee' = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_m \end{bmatrix} \begin{bmatrix} e_1 & e_2 & . & . & e_5 \end{bmatrix} = \begin{bmatrix} e_1^2 & e_1 e_2 & . & . & e_1 e_m \\ e_2 e_1 & e_2^2 & e_2 e_3 & . & e_2 e_m \\ . & e_3 e_2 & e_3^2 & . & e_3 e_m \\ . & . & . & . & . \\ e_m e_1 & e_m e_2 & . & . & e_m^2 \end{bmatrix} \quad (14)$$

$$E\left[ee'\right] = \begin{bmatrix} \operatorname{var}(e_1) & \operatorname{cov}(e_1 e_2) & \operatorname{cov}(e_1 e_3) & . & \operatorname{cov}(e_1 e_m) \\ \operatorname{cov}(e_2 e_1) & \operatorname{var}(e_2) & \operatorname{cov}(e_2 e_3) & . & \operatorname{cov}(e_2 e_m) \\ \operatorname{cov}(e_3 e_1) & \operatorname{cov}(e_3 e_2) & \operatorname{var}(e_3) & . & \operatorname{cov}(e_3 e_m) \\ . & . & . & . & . \\ \operatorname{cov}(e_m e_1) & \operatorname{cov}(e_m e_2) & \operatorname{cov}(e_m e_3) & . & \operatorname{var}(e_m) \end{bmatrix} \quad (15)$$

$$E\left[ee'\right] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & . & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & . & \sigma_{2m} \\ . & \sigma_{32} & \sigma_{33} & . & \sigma_{3m} \\ . & . & . & . & . \\ \sigma_{m1} & \sigma_{m2} & \sigma_{m3} & . & \sigma_{mm} \end{bmatrix} = V = \sum \otimes I \quad (16)$$

Variance-Covariance Structure of Errors in a Sure Model

Dimension of each of the $\sigma_{i,j}$, like that of the identity matrix I, is T by T, and reflects the variance covariance matrix of the stacked regression.

The Kronnecker product $\sum \otimes I$ is a short way of writing this covariance matrix.

$\sum$ is the variance covariance matrix

$\otimes$ is the symbol for the Kronnecker product

I is Identity Matrix with T×M by T×M dimension.

Sure Estimation

Application of the OLS technique individually on (2) and (3) generates inconsistent results. Sure method aims to correct this problem by estimating both equations simultaneously.

$$V^{-1} = \sum^{-1} \otimes I$$

The SURE method is essentially a generalised least square estimator:

$$\hat{\beta} = \left[X'V^{-1}X\right]^{-1} X'V^{-1}Y = \left[X'\left(\sum^{-1} \otimes I\right)X\right]^{-1} X'\left(\sum^{-1} \otimes I\right)Y$$

(17)

$$\hat{\beta} = \begin{bmatrix} \sigma_{11} X_1' X_1 & \sigma_{12} X_1' X_2 & \sigma_{13} X_1' X_3 & . & \sigma_{1m} X_1' X_m \\ \sigma_{21} X_1' X_2 & \sigma_{22} X_2' X_2 & \sigma_{23} X_2' X_3 & . & \sigma_{2m} X_2' X_m \\ . & \sigma_{32} X_3' X_2 & \sigma_{33} X_3' X_3 & . & \sigma_{3m} X_3' X_m \\ . & . & . & . & . \\ \sigma_{m1} X_1' X_m & \sigma_{m2} X_2' X_m & \sigma_{m3} X_3' X_m & . & \sigma_{mm} X_m' X_m \end{bmatrix}^{-1} \begin{bmatrix} \sum_m \sigma_{1m} X_1' Y_m \\ \sum_m \sigma_{1m} X_2' Y_m \\ . \\ . \\ \sum_m \sigma_{1m} X_m' Y_m \end{bmatrix}$$

(18)

Steps for Sure estimation

- Estimate each equation separately using the least square technique.
- Use the least square residuals from step 1 to estimate the error term.
- Use the estimates from the second step to estimate two equations jointly within a generalised least square framework. If m=2 the variance covariance matrix will be as given below.

GLS Example in SURE

$$\Omega = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix} \quad (19)$$

Using a theorem in matrix algebra W can be decomposed into two parts as

$$P'P = \Omega^{-1} \qquad (20)$$

Use this partition of W to transform the original model as

$$PY = PX\beta + Pe \qquad (21)$$

$$Y^* = X^*\beta + e^* \qquad (22)$$

$$\beta_{GLS} = \left(X^{*\prime}X^*\right)^{-1} X^{*\prime}Y^* =$$

$$\left(X'P'PX\right)X'P'PY \quad =>$$

$$\beta_{GLS} = \left(X^{*\prime}\Omega^{-1}X^*\right)^{-1} X^{*\prime}\Omega^{-1}Y^*$$

The GLS estimates are best, linear and unbiased estimators of the coefficients in the SURE system.

Error component Method

Each cross section unit (country) had its own intercept parameter in the pooled dummy variable model.

The error component method decomposes these errors into a common intercept and the random part.

Thus the model will take the following form:

$$g_{i,t} = \alpha_{0,i} + \alpha_{1,i}\pi_{i,t-1} + e_{i,t}$$

$$\alpha_{0,i} = \overline{\alpha}_1 + \mu_i \quad \text{where i =1.., N} \qquad (23)$$

$\overline{\alpha}_1$ represents the population mean intercept and

μ_i are independent of each error $e_{i,t}$. It has a constant mean and constant variance.

$$E(\mu_i) = 0 \quad \text{and} \quad \text{var}(\mu_i) = \sigma_\mu^2$$

$$g_{i,t} = (\overline{\alpha}_1 + \mu_i) + \alpha_{1,i}\pi_{i,t-1} + e_{i,t}$$

$$g_{i,t} = \overline{\alpha}_1 + \alpha_{1,i}\pi_{i,t-1} + (e_{i,t} + \mu_i) = \overline{\alpha}_1 + \alpha_{1,i}\pi_{i,t-1} + v_{i,t}$$

(24)

Error component Method-2

The error component include overall error $e_{i,t}$ and individual specific error μ_i, $v_{i,t} = \left(e_{i,t} + \mu_i\right)$: common and individual specific errors.

$E\left(v_{it}\right) = 0$ the compound error term $v_{i,t}$ has mean zero

$\operatorname{var}\left(v_{it}\right) = \sigma_{\mu}^{2} + \sigma_{e}^{2}$; $v_{i,t}$ is homoskedastic

$\operatorname{cov}\left(v_{it}, v_{is}\right) = \sigma_{\mu}^{2}$ error from the same country in different periods are correlated

$\operatorname{cov}\left(v_{it}, v_{js}\right) = 0$ for i=j errors from different countries are always uncorrelated.

Like in the SURE method the generalised least square estimator, with transformed method produces the most efficient estimators or error component model.

Application: A Growth and Inflation Across countries over time

Countries included in the study (151)

Albania	Comoros	Jamaica	Peru	Uruguay
Algeria	Congo, Dem. Rep. Of	Japan	Philippines	Vanuatu
Angola	Congo, Republic Of	Jordan	Poland	Venezuela
Antigua And Barbuda	Costa Rica	Kenya	Portugal	Vietnam
Argentina	Cote D Ivoire	Kiribati	Qatar	Zambia
Australia	Cyprus	Korea	Romania	Zimbabwe
Austria	Denmark	Kuwait	Rwanda	
Bahamas, The	Djibouti	Lao People's Dem.Rep	Samoa	
Bahrain	Dominica	Lebanon	Sao Tome & Principe	
Bangladesh	Dominican Republic	Lesotho	Saudi Arabia	
Barbados	Ecuador	Libya	Senegal	
Belgium	Egypt	Luxembourg	Seychelles	
Belize	El Salvador	Madagascar	Sierra Leone	
Benin	Equatorial Guinea	Malawi	Singapore	
Bhutan	Ethiopia	Malaysia	Solomon Islands	
Bolivia	Fiji	Maldives	South Africa	
Botswana	Finland	Mali	Spain	
Brazil	France	Malta	Sri Lanka	
Bulgaria	Gabon	Mauritania	St. Kitts And Nevis	
Burkina Faso	Gambia, The	Mauritius	St. Lucia	
Burundi	Germany	Mexico	St. Vincent & Grens.	
Cameroon	Ghana	Mongolia	Sudan	
Canada	Greece	Morocco	Suriname	
Cape Verde	Grenada	Mozambique	Swaziland	
Central African Rep.	Guatemala	Myanmar	Sweden	
Chad	Guinea	Namibia	Switzerland	
Chile	Guinea-Bissau	Nepal	Syrian Arab Republic	
China P.R.:	Guyana	Netherlands	Taiwan Prov Of China	

Shazam Syntax

```
sample 1 151

par 15000
read y1 y2 y3 y4 y5 y6 y7 y8 y9 y10 y11 y12     y13 y14     y15 y16     y17
y18 y19 y20   y21 y22     y23  y24     y25 y26     y27 y28 y29 y30     y31

read p1p2 p3 p4 p5 p6 p7 p8 p9     p10 p11     p12 p13     p14 p15     p16
p17    p18 p19     p20 p21     p22 p23     p24 p25     p26 p27     p28
p29    p30 p31

system 2 /
ols y8 y1 y2 y3 y4 y5 y6 y7 p2 p3 p4 p5 p6 p7
ols p8 p1 p2 p3 p4 p5 p6 p7 y2 y3 y4 y5 y6 y7 /DN
end

*pooling cross section and time series

matrix
infl=(p1'|p2'|p3'|p4'|p5'|p6'|p7'|p8'|p9'|p10'|p11'|p12'|p13'|p14'|p15'|p16'|p17'|p18'|p19'|p20'|p
21'|p22'|p23'| &
p24'|p25'|p26'|p27'|p28'|p29'|p30'|p31')'

matrix rgg=(y1'|y2'|y3'|y4'|y5'|y6'|y7'|y8'|y9'|y10'|&
y11'|y12'|y13'|y14'|y15'|y16'|y17'|y18'|y19'|y20'| &
y21'|y22'|y23'|y24'|y25'|y26'|y27'|y28'|y29'|y30'|y31')'

sample 1 4681

ols rgg infl
pool rgg infl /NCROSS=151 DN
genr tindex=sum(seas(151))
genr csindex=time(0)-151*(tindex-1)
*print tindex csindex

pool rgg infl /NCROSS=151  COEF=Beta
```

One Result on Pooling Model

```
R-SQUARE =   0.0043     R-SQUARE ADJUSTED =  -0.0021
 VARIANCE OF THE ESTIMATE-SIGMA**2 =   55.671
 STANDARD ERROR OF THE ESTIMATE-SIGMA =   7.4613
 SUM OF SQUARED ERRORS-SSE=  0.25887E+06
 MEAN OF DEPENDENT VARIABLE =   3.7790
 LOG OF THE LIKELIHOOD FUNCTION = -16034.0
 RAW MOMENT R-SQUARE =   0.2079
```

ANALYSIS OF VARIANCE - FROM ZERO

	SS	DF	MS	F
REGRESSION	67964.	31.	2192.4	39.381
ERROR	0.25887E+06	4650.	55.671	P-VALUE
TOTAL	0.32683E+06	4681.	69.821	0.000

VARIABLE NAME	ESTIMATED COEFFICIENT	STANDARD ERROR	T-RATIO 4650 DF	P-VALUE	PARTIAL CORR.	STANDARDIZED COEFFICIENT	ELASTICITY AT MEANS
INFL	-0.87005E-03	0.2204E-03	-3.948	0.000	-0.058	-0.0580	-0.0093
T2	5.4945	0.6072	9.049	0.000	0.132	0.1303	0.0469
T3	5.0154	0.6072	8.260	0.000	0.120	0.1189	0.0428
T4	5.2188	0.6072	8.595	0.000	0.125	0.1237	0.0445
T5	4.8993	0.6072	8.068	0.000	0.118	0.1162	0.0418
T6	2.4485	0.6072	4.032	0.000	0.059	0.0580	0.0209
T7	5.9081	0.6072	9.730	0.000	0.141	0.1401	0.0504
T8	4.3478	0.6072	7.160	0.000	0.104	0.1031	0.0371
T9	4.4664	0.6072	7.356	0.000	0.107	0.1059	0.0381
T10	4.1043	0.6072	6.759	0.000	0.099	0.0973	0.0350
T11	2.8885	0.6072	4.757	0.000	0.070	0.0685	0.0247
T12	2.2897	0.6072	3.771	0.000	0.055	0.0543	0.0195
T13	1.1199	0.6072	1.844	0.065	0.027	0.0266	0.0096
T14	3.1440	0.6072	5.178	0.000	0.076	0.0745	0.0268
T15	3.7295	0.6072	6.142	0.000	0.090	0.0884	0.0318
T16	4.7003	0.6076	7.736	0.000	0.113	0.1114	0.0401
T17	3.4980	0.6072	5.761	0.000	0.084	0.0829	0.0299
T18	3.4993	0.6072	5.763	0.000	0.084	0.0830	0.0299
T19	3.9309	0.6078	6.468	0.000	0.094	0.0932	0.0336
T20	3.5159	0.6076	5.787	0.000	0.085	0.0834	0.0300
T21	3.3853	0.6077	5.571	0.000	0.081	0.0803	0.0289
T22	2.2109	0.6075	3.639	0.000	0.053	0.0524	0.0189
T23	2.9452	0.6073	4.850	0.000	0.071	0.0698	0.0251
T24	2.9669	0.6073	4.886	0.000	0.071	0.0703	0.0253
T25	3.4106	0.6087	5.603	0.000	0.082	0.0809	0.0291
T26	4.2862	0.6072	7.058	0.000	0.103	0.1016	0.0366
T27	4.2761	0.6073	7.042	0.000	0.103	0.1014	0.0365
T28	4.3304	0.6072	7.132	0.000	0.104	0.1027	0.0370
T29	3.0099	0.6072	4.957	0.000	0.073	0.0714	0.0257
T30	3.2318	0.6072	5.322	0.000	0.078	0.0766	0.0276
T31	3.9365	0.6072	6.483	0.000	0.095	0.0933	0.0336