[Frequently Asked Questions]

Panal Data Analysis

Subject:	Business Economics
Course:	B. A. (Hons.), 5th Semester, Undergraduate
Paper No. & Title:	Paper – 531 Elective Paper Q1 – Advanced Econometrics
Unit No. & Title:	Unit – 5 Panel Data
Lecture No. & Title:	Lecture – 1 Panal Data Analysis

Frequently Asked Questions

Q1. What is meant by Panel data?

A1. We have cross section data pertaining to some fixed periods for one or more variables. Similarly we have time series data pertaining to many periods and for one or more variables. Panel data is a combination of both space as well as time dimensions. Here we have the cross sectional unit (family or firm or state) surveyed over time. We get large data and there can be complexibility also.

Q2. Give some illustrations for panel data

A2. (1) Gender ratio observed in different states of India during some given periods.

(2)Literacy ratio by sex wise for different states of India during some given periods

(3) Productivity measures, wages for different states during some periods.

(4) Income expenditure patterns for different groups of people in the states during some periods.

(5) Data for 5 transport companies in the country giving details about profit and loss patterns during given periods etc.

Q3. What are the advantages for panel data?

A3. (1) They increase sample size reasonably due to large bulk of data.

(2) Panel data present repeated cross sectional observations. This suits to study the dynamics of change.

(3) They help us to study more complicated behavioural models.

(4) Panel data are combinations of time series and cross section data. This gives more information on data, more variability, less

collinearity among variables, more degrees of freedom and more efficiency.

Q4. What are the disadvantages of panel data?

A4. (1) There are several estimation and inference problems.

(2) Since panel data is a mixture of cross sectional data and time series data, there is heteroscadasticity in cross section data and autocorrelation in time series data.

(3) Analysis becomes complex to understand and it also needs sufficient care for implementation.

Q5. Name some panel data regression models.

A5. There are 4 categories of models for study of panel data.

(1) Pooled OLS model (Constant coefficients model)

(2) Fixed effect least squares Dummy variables model (LSDV)

(3) Fixed effect within group model

(4) Random effect model (REM)

Q6. There are 5 companies producing steel in the country. Data are collected for net value added in terms of the gross value of output expenditures for raw materials, fuel consumption and labourforce participation.

How would you construct constant coefficient model for the above when 10 years data are available?

A6. We want to build up the models for these five companies for their net profit.

We define P_{it} = Net profit (Net value added) by i^{th} company in year t

$$(i = 1,2,3,4,5\ , j = 1,2,3 \ldots 10)$$

G_{it} = Gross output in value term for i^{th} industry in year t

R_{it} = Row material expenses for i^{th} company in year t

F_{it} = Fuel consumption expenses for i^{th} company during year t

L_{it} = Laborforce participation for i^{th} company during year t

U_{it} = Disturbance term

Model

$$P_{it} = \beta_1 + \beta_2.G_{it} + \beta_3.R_{it} + \beta_4.F_{it} + \beta_5.L_{it} + U_{it}$$

$$(i = 1,2,3,4,5\ , j = 1,2,3 \ldots 10)$$

Which is pooled OLS model (or constant coefficient model).

Q7. For the above application in Q.6, how would you write fixed effect model?

A7. We modify the above model as fixed effect model by writing β_{1i} in place of β_1 thus

$$P_{it} = \beta_{1i} + \beta_2.G_{it} + \beta_3.R_{it} + \beta_4.F_{it} + \beta_5.L_{it} + U_{it} \quad (i = 1,2,3,4,5\ , j = 1,2,3 \ldots 10)$$

Q8. How would you write LSDV model for the above application in Q.6?

A8. We use 4 dummy variables corresponding to 5 companies and write $\beta_{1i} = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \alpha_5 D_{5i}$

So that LSDV model is

$$P_{it} = (\alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + \alpha_5 D_{5i}) + \beta_2.G_{it} + \beta_3.R_{it} + \beta_4.F_{it} + \beta_5.L_{it} + U_{it}$$

Here $D_{2i} = 1$ for company 2

$= 0$ otherwise

$D_{3i} = 1$ for company 3

$= 0$ otherwise

$D_{4i} = 1$ for company 4

$= 0$ otherwise

$D_{5i} = 1$ for company 5

$= 0$ otherwise

$D_{6i} = 1$ for company 6

$= 0$ otherwise

(D_2, D_3, D_4 and D_5 are dummy variables).

Q9. How would you write REM for application in Q.6 above?

A9. we write becomes $\beta_{1i} = \beta_1 + \epsilon_i$ then the model equation becomes

$$P_{it} = \beta_1 + \beta_2 . G_{it} + \beta_3 . R_{it} + \beta_4 . F_{it} + \beta_5 . L_{it} + W_{it}$$

Where $W_{it} = U_{it} + \epsilon_i$

This is called Random effect model due to the term W_{it}

Q10. What is the thumb rule (working rule) for making a choice between FEM and REM?

A10. (1) If it is assumed that ϵ_i and the $X's$ are uncorrelated, ECM may be appropriate.

(2)If ϵ_i and the $X's$ are correlated, FEM may be appropriate.