

## Seminars 6. Panel data

### Task 1. Firms investment

File: *firms.txt*

We analyse activity of 10 firms during 20 years (1935-1954).

Variables in the file are:

- Firm = Firm ID, 1,...,10
- Year = 1935,...,1954
- I = Investment
- F = Real Value of the Firm
- C = Real Value of the Firm's Capital Stock

1. Create an own model for panel data and estimate it.
2. Make a choice between models with random and fixed effects.
3. Provide economic analysis of the model (less than 1 page).

### Task 2. Spanish Dairy Farm Production

File: *dairy.txt*

Variables in the file are

- FARM = Farm ID, , N = 247
- YEAR = year, 93, 94, ..., 98, , T = 6

Inputs

- COWS, X1 = log of, deviations from means (logs)
- LAND, X2 = same
- LABOR, X3 = same
- FEED, X4 = same
- Translog terms = squares and cross products: X11, X22, X33, X44, X12, X13, X14, X23, X24, X34
- YEAR93,...,YEAR98 = year dummy variables

Output

- MILK = farm output
- YIT = log of MILK production

1. Create two own models for panel data that can explain output variables with input variables. Estimate models.
2. Make a choice between models with random and fixed effects.
3. Provide economic analysis of the model (less than 1 page).

### Task 3. Bank Cost Data, 500 Banks, 5 Years:

File: *banks.txt*

Variables in the file are

- $C_{it}$  = total cost of transformation of financial and physical resources into loans and investments = the sum of the five cost items described below;
- $Y1_{it}$  = installment loans to individuals for personal and household expenses;
- $Y2_{it}$  = real estate loans;
- $Y3_{it}$  = business loans;
- $Y4_{it}$  = federal funds sold and securities purchased under agreements to resell;
- $Y5_{it}$  = other assets;
- $W1_{it}$  = price of labor, average wage per employee;
- $W2_{it}$  = price of capital = expenses on premises and fixed assets divided by the dollar value of premises and fixed assets;
- $W3_{it}$  = price of purchased funds = interest expense on money market deposits plus expense of federal funds purchased and securities sold under agreements to repurchase plus interest expense on demand notes issued by the U.S. Treasury divided by the dollar value of purchased funds;
- $W4_{it}$  = price of interest-bearing deposits in total transaction accounts = interest expense on interest-bearing categories of total transaction accounts;
- $W5_{it}$  = price of interest-bearing deposits in total nontransaction accounts = interest expense on total deposits minus interest expense on money market deposit accounts divided by the dollar value of interest-bearing deposits in total nontransaction accounts;
- $T$  = trend variable,  $t = 1, 2, 3, 4, 5$  for years 1996, 1997, 1998, 1999, 2000

The data in the file are for a translog cost function, linearly homogeneous in the input prices. Specifically,

- $C = \log(\text{Cost}/W5)$ ,  $W1, W2, W3, W4 = \log(W_j/W5)$ ,  $Q1, \dots, Q5 = \log(Y_m)$ , and the squared and cross product terms are  $W_{11}, W_{12}, \dots, Q_{11}, Q_{12}, \dots, W_1Q_1, \dots, W_4Q_5, T, T_2, TW_1, \dots, TW_4, TQ_1, \dots, TQ_5$ .

1. Provide model for panel data that can explain the first variable  $C_{it}$ . Estimate model.
2. Make a choice between models with random and fixed effects.
3. Provide economic analysis of the model (less than 1 page).

#### **Task 4. Dahlberg and Johansson Municipal Expenditure Data**

*File: Municipal.txt*

Variables in the file are

- ID = Identification, 1, ..., 265
  - YEAR = year, 1979, ..., 1987
  - EXPEND = Expenditures
  - REVENUE = Receipts, taxes and Fees
  - GRANTS = Government grants and shared tax revenues
1. Create the model with panel data that explains EXPEND variable against two other factors. Estimate model. Check it for significance.
  2. Make a choice between models with random and fixed effects.
  3. Provide economic analysis of the model (less than 1 page).

### Task 5. World Gasoline Demand Data

*File: Gasoline.txt*

Variables in the file are

- COUNTRY = name of country (Does not appear in the LIMDEP project file)
- YEAR = year, 1960-1978
- LGASPCAR = log of consumption per car
- LINCOME = log of per capita income
- LRPMG = log of real price of gasoline
- LCARPCAP = log of per capita number of cars

1. Create the model with panel data that explains LGASPCAR variable against other factors. Estimate model. Check it for significance.
2. Make a choice between models with random and fixed effects.
3. Provide economic analysis of the model (less than 1 page).

### Task 6. Statewide Capital Productivity Data

*File: product.txt*

Variables in the file are

- STATE = state name
- ST\_ABB = state abbreviation (USA states)
- YR = year, 1970,...,1986
- P\_CAP = public capital
- HWY = highway capital
- WATER = water utility capital
- UTIL = utility capital
- PC = private capital
- GSP = gross state product
- EMP = employment
- UNEMP = unemployment rate

1. Create the models with panel data that explain P\_CAP and GSP against other factors. Estimate models. Check them for significance.
2. Make a choice between models with random and fixed effects.
3. Provide economic analysis of the model (less than 1 page).

### Task 7. Cornwell and Rupert Returns to Schooling Data

*File: returns.txt*

Variables in the file are

- 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
- LWAGE = log of wage

1. Create the model with panel data that explains EXP. Estimate models. Check it for significance.
2. Make a choice between models with random and fixed effects.
3. Test for significance of dummy variables.

### **Task 8. German Health Care Usage Data**

*File: health.txt*

Data downloaded from Journal of Applied Econometrics Archive. This is an unbalanced panel with 7,293 individuals. Note, the variable NUMOBS below tells how many observations there are for each person. This variable is repeated in each row of the data for the person.

Variables in the file are

- ID = person - identification number
- FEMALE = female = 1; male = 0
- YEAR = calendar year of the observation
- AGE = age in years
- HSAT = health satisfaction, coded 0 (low) - 10 (high) Note, this variable has 40 coding errors. Variable NEWHSAT below fixes them.
- HANDDUM = handicapped = 1; otherwise = 0
- HANDPER = degree of handicap in percent (0 - 100)
- HHNINC = household nominal monthly net income in German marks / 10000
- HHKIDS = children under age 16 in the household = 1; otherwise = 0
- EDUC = years of schooling
- MARRIED = married = 1; otherwise = 0
- HAUPTS = highest schooling degree is Hauptschul degree = 1; otherwise = 0
- REALS = highest schooling degree is Realschul degree = 1; otherwise = 0
- FACHHS = highest schooling degree is Polytechnical degree = 1; otherwise = 0
- ABITUR = highest schooling degree is Abitur = 1; otherwise = 0
- UNIV = highest schooling degree is university degree = 1; otherwise = 0
- WORKING = employed = 1; otherwise = 0
- BLUEC = blue collar employee = 1; otherwise = 0
- WHITEC = white collar employee = 1; otherwise = 0
- SELF = self-employed = 1; otherwise = 0

- BEAMT = civil servant = 1; otherwise = 0
  - DOCVIS = number of doctor visits in last three months
  - HOSPVIS = number of hospital visits in last calendar year
  - PUBLIC = insured in public health insurance = 1; otherwise = 0
  - ADDON = insured by add-on insurance = 1; otherwise = 0
  - NUMOBS = number of observations for this person. Repeated in each row of data.
  - NEWHSAT = recoded value of HSAT with coding errors corrected.
1. Create the model with panel data that explains NEWHSAT. Estimate models. Check it for significance.
  2. Make a choice between models with random and fixed effects.
  3. Define the most important factors that influence the dependent variable.
  4. Provide economic analysis of the model (less than 1 page).

### **Task 9. World Health Organization Panel Data on Health Care Attainment**

*File: world\_health.txt*

These data have been used by many researchers to study the Health Care Survey assembled by WHO as part of the Year 2000 World Health Report. To set the proper sample for panel data analysis, use observations for which SMALL = 0. To obtain the balanced panel, then use only observations with GROUPTI = 5.

- COMP = composite measure of health care attainment; LCOMP = logCOMP
- DALE = Disability adjusted life expectancy (other measure); LDALE = logDALE
- YEAR = 1993,...,1997; TIME = 1,2,3,4,5; T93, T94, T95, T96, T97 = year dummy variables
- HEXP = per capita health expenditure; LHEXP = logHEXP; LHEXP2 = log-squaredHEXP
- HC3 = educational attainment; LHC = logHC3; LHC2 = log-squaredHC3; LHEXPHC = logHEXP \* logHC3
- SMALL = indicator for states, provinces, etc. SMALL > 0 implies internal political unit, = 0 implies country observation
- COUNTRY = number assigned to country
- STRATUM = another country indicator
- GROUPTI = number of observations when SMALL = 0. Usually 5, some = 1, one country = 4.
- OECD = dummy variable for OECD country (30 countries)
- GINI = Gini coefficient for income inequality
- GEFF = world bank measure of government effectiveness\*
- VOICE = world bank measure of democratization of the political process\*
- TROPICS = dummy variable for tropical location
- POPDEN = population density\*
- PUBTHE = proportion of health expenditure paid by public authorities
- GDPC = normalized per capita GDP; LGDPC = logGDPC; LGDPC2 = log-squaredGDPC

1. Create the model with panel data that explains COMP. Estimate models. Check it for significance.
2. Make a choice between models with random and fixed effects.
3. Define the most important factors that influence the dependent variable.
4. Provide economic analysis of the model (less than 1 page).

### Task 10. Research market

File: *research.txt*

The data consist a panel on individual research productivity, salary, and other characteristics. The raw data consists of observations on individuals; there are 11 variables:

- $i$  – person identifier
- $t_{0i}$  – year of Phd - 1900
- $t$  – current year - 1900
- $x_i$  –gender (=1 if female)
- $d_{it}$  – indicator of employment in research university (1 if person  $i$  is in a research position in year  $t$ , and 0 otherwise.)
- $y_{it}$  – page equivalents in current year
- $r_i$  – rank of Phd granting University
- $Y_{it}$  – discounted\*cumulative page equivalents
- $s_{it}$  – current annual salary in \$1,000/yr

A relatively new aspect of the data is the availability of the variable  $r_i$ , which is the rank of the  $i^{\text{th}}$  individual's PhD granting University at the time of the PhD. While such rankings are notoriously flawed, it is of obvious interest to investigate the extent to which individual productivity depends upon the perceived quality of their PhD program. Note also however that since unobserved individual effects on productivity are likely to be also correlated with these ranks, this variable should probably not be treated as exogenous.

Write a brief research paper (less than 3 pages) in which you use the data described above to answer some or all of the following questions.

1. What is the shape of research productivity over the life cycle?
2. Does “the scientific revolution” of the 1960's appear to make pre-60's PhD's less productive than their post-60's colleagues?
3. How much is a page of research worth as measured by the resulting discounted stream of academic salary? Note that this may depend upon the point in the life-cycle that the page is published, what the accumulated research output is to date, and other factors. Be as explicit as possible here giving examples of “representative individuals” if necessary.
4. Does the gender of the scientist influence productivity or salary and if so, by how much?
5. The distinguished statistician Emmanuel Parzen has argued that economics is “gradually becoming more scientific” because publication is becoming more concentrated in the hands of fewer researchers. Test this hypothesis for phuzics.

6. How does the rank of an individual PhD program influence his/her subsequent research productivity?

Use hints from: <http://www.econ.uiuc.edu/~roger/courses/472/problems/ps4.pdf>